

# Statistical analysis of the wind speed at the top of la Soufrière Volcano

Didier Bernard, Olivier Tossa, Richard Emilion

▶ **To cite this version:**

Didier Bernard, Olivier Tossa, Richard Emilion. Statistical analysis of the wind speed at the top of la Soufrière Volcano. 2007. hal-00165572

**HAL Id: hal-00165572**
**https://hal.science/hal-00165572**

Preprint submitted on 30 Jul 2007

**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.

# Statistical analysis of the wind speed at the top of La Soufrière volcano

Didier BERNARD

*LPAT, Université Antilles-Guyane, B.P. 250 97157 Pointe-à-Pitre Guadeloupe*

Adaté TOSSA

*Université Paris Dauphine, 75016 Paris, France*

Richard EMILION

*MAPMO, Université d'Orléans, 45100 Orléans, France*

**Abstract**

We first analyze some atmospheric parameters measured in 2004 at the top of La Soufrière volcano: wind speed, pressure and temperature. Next, we study the stochastic process of the wind speed modulus which is of interest for prediction of pollution induced by the plume. We fit to the data various models: FARIMA, periodic trend plus multifractional Brownian noise, multifractal continuous cascades. The estimations show that the parameters are quite constant during some time intervals but that they evoluate during some transition phases in response to environment changes. Thus, we finally suggest, for future works, some settings that should be more appropriate: time series with regime changes and stochastic processes in random environment.

*Key words:* atmospheric parameters, La Soufrière, multiplicative cascades, turbulence, wind speed.

*Email addresses:* didier.bernard@univ-ag.fr (Didier BERNARD), otossa@yahoo.fr (Adaté TOSSA), richard.emilion@univ-orleans.fr (Richard EMILION).

## 1 Introduction

Since several years, a strong fumarolic plume is emitted from the south crater of La soufrière volcano, in Guadeloupe. In 2004, in order to characterize gases, OP-FTIR measurements in that volcanic plume were performed, revealing presence of HCl, SO2, SiF4, CH4. Accurate estimations of gas traces were difficult to obtain due to spatial and temporal variability of the emissions, acidity of gases, volcano morphology, and weather conditions. For example, humidity largely contaminated the OP-FTIR spectra while high variations of wind intensity and wind direction disturbed the measurements in the plume. These results and observations can be found in [Bernard et al.(2006)].
As temporal evolution and forecasting of gas traces is a matter of fundamental importance for the assessment of the volcanic risk and for public safety in the volcanic areas, it is of great interest to elaborate models that can predict atmospheric gase concentrations, taking in account the wind process. As a first step to this purpose, we report in the present work a statisticial analysis of measurements on various meteorological parameters, performed near the summit of the volcano.
The paper is organized as follows. In Section 2, we shortly present the measurements context. Section 3 is concerned with elementary but useful statistics such as daily or monthly profiles. In Section 4 we perform a fast Fourier transform analysis and we represent the autocorellation function of the wind modulus process. In Section 5 we put in evidence a periodic trend and a high local variability of the noise process by fitting a multifractional Brownian motion (MFBM). Due to the difficulties of identifying a model for the Hurst function in the MFBM, we turn to, in Section 6, a model which has a deep interpretability in physics: the continuous cascades model with fractional stochastic integral drift and standard noise [Schmitt (2001)], [Schmitt (2003)], [Schmitt (2005)]. We also show why the estimations that we obtain suggest an extension of such a model. Conclusion and perspectives on future works are presented in the last section.

## 2 Material and methods

### *2.1 Instruments*

Since 2000, a meteorological station has been installed by the "Observatoire Volcanologique et Sismologique de Guadeloupe" at the top of the volcano La Soufrière. The station is situated at Piton Sanner, 16°02.70'N, 61°39.76'W, h=1442 m on the the mount "Sanner". Horizontal wind speed and direction were measured by a wind monitor (Campbell scientific 05103-03) Accuracy :

0.3 m $s^{-1}$, 1 to 60 m $s^{-1}$; $\pm$1.0 m $s^{-1}$ 60 to 100 m $s^{-1}$) at $z = 2$m. This type of anemometer was selected because its design offers improved resistance to corrosion from atmospheric pollutants (volcanic plume) and gusts. Temperature (HMP45C), relative humidity (HMP45C Campbell), barometric pressure (PTB 101B Campbell), radiation (pyranometer SP1110 Campbell) and pluviometry were also measured. Signals from the station were recorded by dataloggers (CR10X Campbell scientific). All the equipments are Campbell. A detailed physical description of these systems is given in "http://www.ovsg.univ-ag.fr/meteo/public/SANNER".

### *2.2 Sampling*

The dataset used in this analysis concerns three meteorological parameters, wind speed, pressure, wind direction, supplied by the 'Observatoire volcanologique et Sismologique de Guadeloupe and denoted S, P and D, respectively. The data were discretely sampled every minute but only averaged intervals of ten minutes are recorded. The experiment has provided a time series of 12 month length (from january 1st to december 31th 2004). The mean ten-minute wind speed (in m $s^{-1}$) corresponds to the meteorological mean wind so that variations due to turbulence are filtered. Therefore, measurements made here are well adapted to synoptic and average scale phenomenon representation.

## 3 Elementary statistics

### *3.1 Summary statistics*

The following table is concerned with some elementary yearly statistics of the three variables S, P and D. As it must be expected, the atmospheric pressure seems to be a stationary process with small variation, at the opposite of the wind speed and wind direction processes.

| | Min | Max | Mean | Median | Std. dev. |
|---|---|---|---|---|---|
| Wind velocity | 0 | 27.81 | 10.89 | 11.39 | 5.55 |
| Pressure (mm.Hg) | 849 | 862.2 | 857.5 | 857.5 | 1.676 |
| Direction (degrees) | 21.6 | 342.4 | 118.2 | 107.6 | 36.23 |

Table 1: Summary statistics of (S,P,D)

### *3.2 Time series (S, P, D)*

The figure below represents the time series of the wind speed, the atmospheric pressure and the wind direction, respectively. It can be observed that wind speed is more chaotic than atmospheric pressure while wind direction has some random extreme jumps.

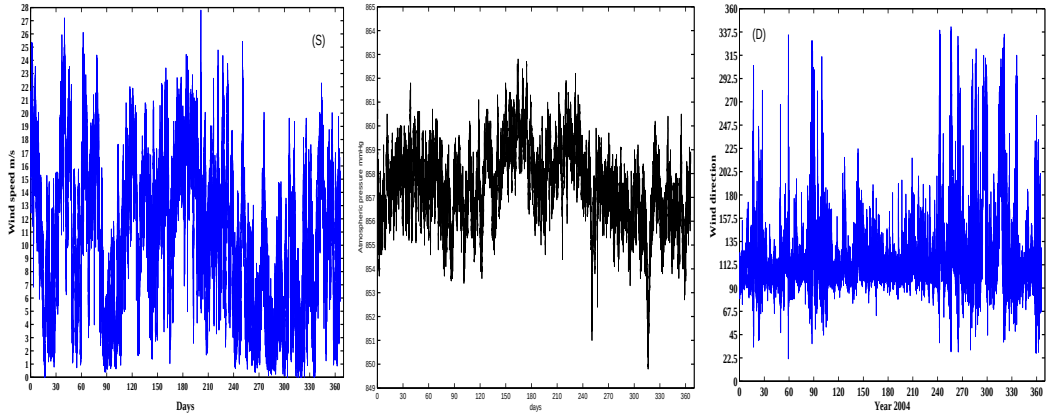



Fig. 1. Time series of (S,P,D), La Soufrière Volcano

### *3.3 Distributions and occupation time*

It seems that the connection between the distribution of the sampled values of a continous stochastic process and its local time is not very popular. First recall that the bar histograms as those plotted in Fig. 2 below are computed as follows. The min-max interval of the data values is splitted into regular bins and the surface of each bar is the frequency of the values belonging to the bin. As the measurements were sampled at regular time intervals, we see that this frequency is also a time frequency. Therefore, if the wind speed and pressure are considered as continuous stochastic processes observed during a time interval $(t, T)$, these bar histograms estimate a probability distribution function (pdf) $L(t, T)$, the occupation time density of these processes, also called local time in the probabilistic literature.

#### *3.3.1 (S,P,D) occupation time densities*

For 2004, the wind intensity varies from 0 to 28 m $s^{-1}$ with a mean speed of 10.89 m $s^{-1}$ and a standard deviation of 5.55 m $s^{-1}$. Its graphical representation is a bimodal asymetric histogram (Fig. 2). This corresponds to the well known property of the wind speed density to generally be a mixture of two Weibull distributions.

Recall that a peak of the density (for example at 4m $s^{-1}$ and at 12m $s^{-1}$ in april 2004) means that the speed is more frequent at the corresponding value during the monitoring interval.
The atmospheric pressure varies from 849 to 862.5 mm Hg. Its representation seems to be quite a normal distribution. It is more difficult to characterize the wind direction which seems to be a gamma distribution.

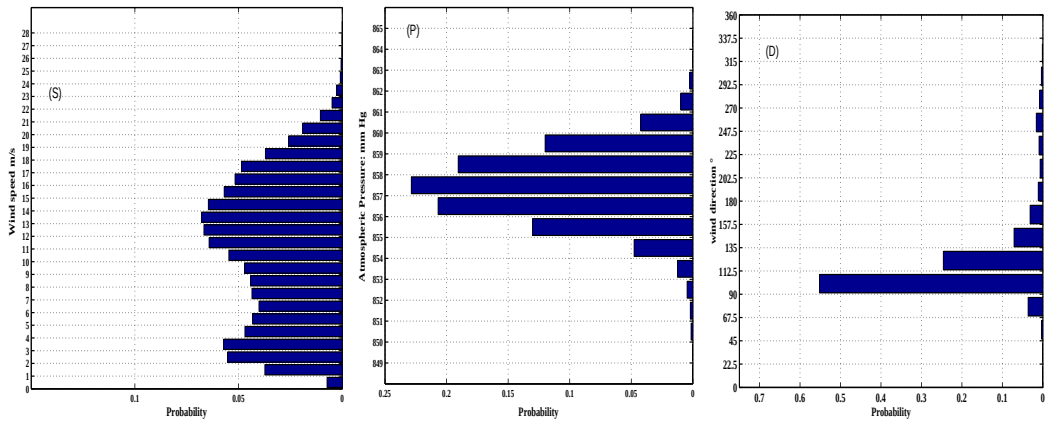



Fig. 2. Densities of (S,P,D)

### *3.3.2 Monthly profiles*

Monthly averages of the wind speed, atmospheric pressure and wind direction are represented in Fig. 3. Monthly profiles wind speed vary between 14.70 m $s^{-1}$ in June to 5.87 m $s^{-1}$ in October. These profiles are very similar to those recorded and calculated for the synoptic flux in a climatic station of Météo-France in Guadeloupe [Brevignon et al.(2006)]. Only the order of magnitude is different: the intensity is multiplied by 1.6 at the summit.

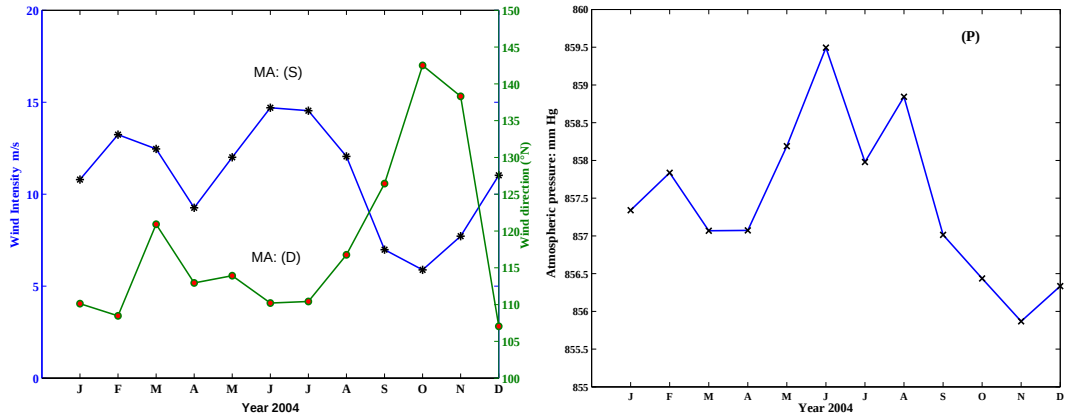



Fig. 3. Monthly average of (S,P,D)

Monthly profiles could be separated in four parts:

- January to february: we observed a constant flux for wind speed and atmospheric pressure
- March-June: the Azores anticyclone grows stronger and stronger while the ITCZ goes up to the north with an increase of the atmospheric pressure accompanied with a strengthening of the wind flux.
- July to October: flux intensity decreases to reach a minima in october. A similar trend is also observed for the atmospheric pressure.
- October- December: a steady increase of fluxes is observed.

The monthly evolution of the average pressure follows the same profile as the intensity of the wind.
Concerning wind direction, it is about 110-115 south until July. From August a progressive rotation towards the south is observed. It reaches its maximum, about 140, in October and then decreases towards the values measured at the beginning of the year.

The following figures show the distributions of S, P and D during two extreme behaviours: June and October 2004.

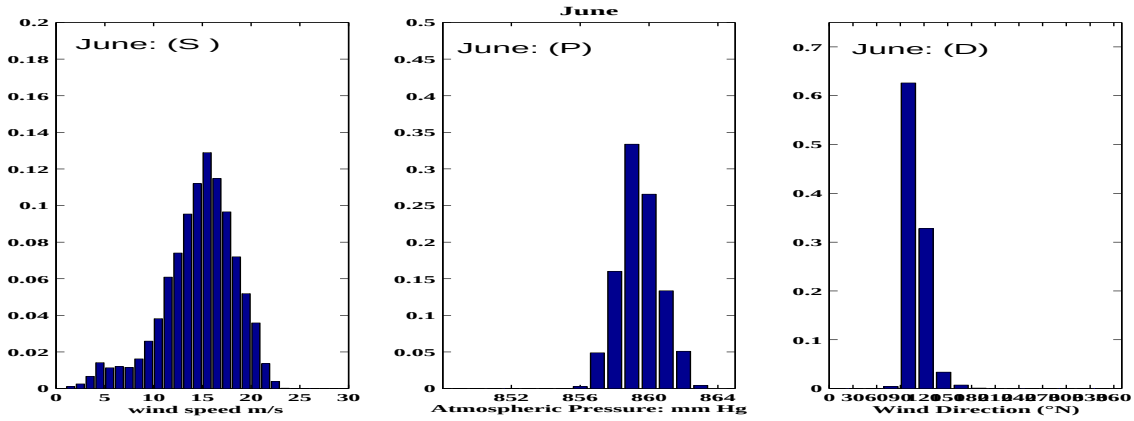



Fig. 4. Densities of (S,P,D) in June 2004

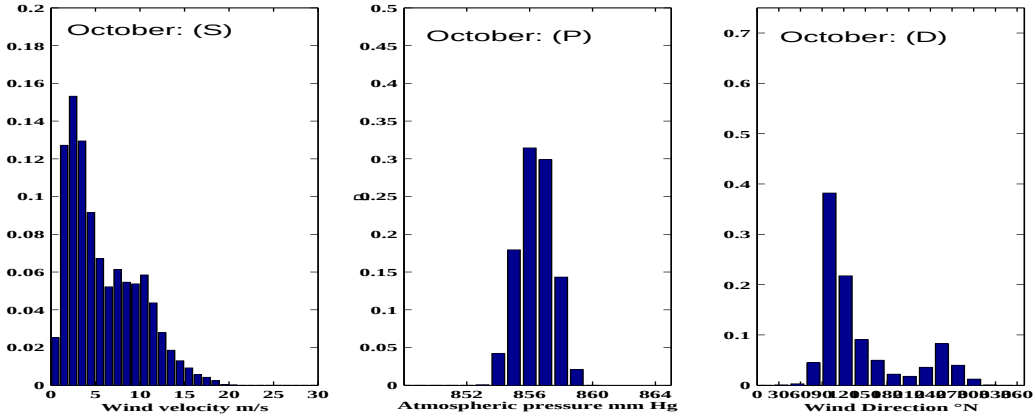



Fig. 5. Densities of (S,P,D) in October 2004

Observe that the probability that the speed is less than 10 m $s^{-1}$ is small while

it is large in October. Concerning the pressure, there is a sliding of the high pressures towards the weak ones with a variation of magnitude around 4 mb. The wind direction is mainly south-east in June and south to south-west in October.

### *3.3.3 Average day profile (V,P)*

Since the sampling is done every 10 min, 6 data are collected per hour and 144 per day. To get a day profile, we have computed for each time $t = 1, ...144$, the mean and the standard deviation through the 366 days of year 2004. The following figures represent the evolution of the mean of the data computed at each time. Clearly, S and P are highly correlated.

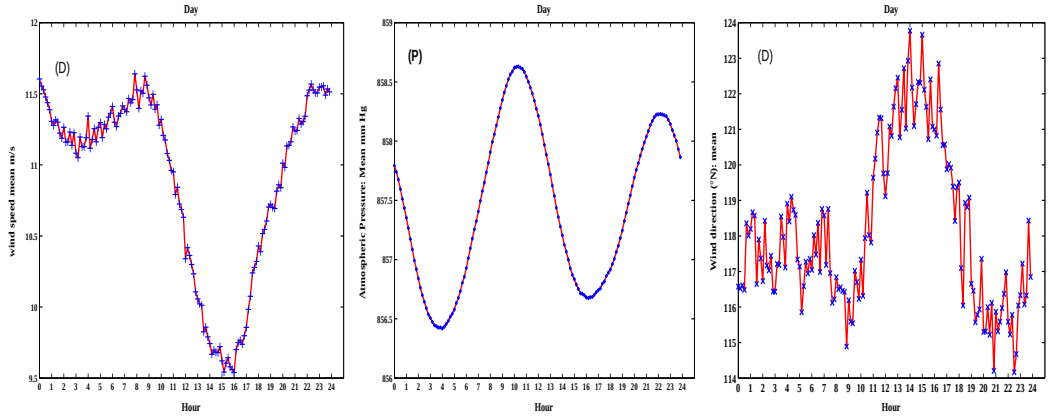



Fig. 6. Average Day profile of (S,P,D) for 2004.

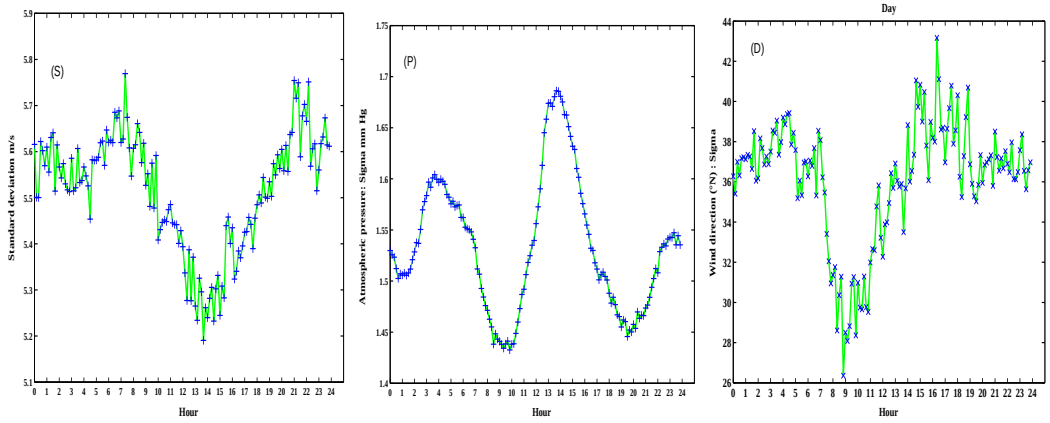



Fig. 7. Standard deviation for a day profile of S, P and D in 2004.

Similarities between these evolution curves yield some empirical rules such as the following ones which can be of interest for future measurement experiments. Between 0 and 10 a.m., wind speed varies around 11 m $s^{-1}$. From 10 a.m. to 3:00 p.m. it decreases to reach a minima of 9.5 m $s^{-1}$ between 3:00 and 4:00 p.m.. Then it regularly increases to reach a maxima at 10 p.m..

Concerning the pressure we observe a well shaped "pressure wave" with an amplitude of 2 hPa and a 12 hour phase. We observe two maxima at 10 a.m. and 10 p.m. local time alternated by two minima at 4 a.m. and 4 p.m.. These results agree with published meteorological data [Dhonneur (1978)], [Brevignon et al.(2006)]. Concerning wind direction, it is constant from 0:00 to 9:00 a.m.. A rotation of about 10 towards the south is observed from 9:00 a.m. to 2:00 p.m.,then wind direction comes back to its initial direction at 6:00 p.m.. Fig.6 represents standard deviations calculated at each time of the day-profile for the three variables S, P and D. The measure of the distibution dispersion is evaluated by the coefficient of variation. For S, it is 50%, for P it is less than 1% and for D it is about 25 %. Thus S is the most chaotic variable and P is the most stable one.

### *3.3.4 Profiles at a fixed hour*

It is of interest to know how the data vary when they are observed at a fixed hour every day, say e.g. at 8 a.m. The following figures yield some empirical rules. On the 5th april 2004, at 8 a.m., the wind speed modulus has a first peak, then it decreases from 5th to 10th, and from 10th it increases up to reach a second peak on 12th. Next, the speed at 8 a.m. decreases from 12th to 15th and then increases up to the largest peak on 23th. It decreases down on 24th to slightly increase again. This profile is quite the same at any fixed hour during this month.

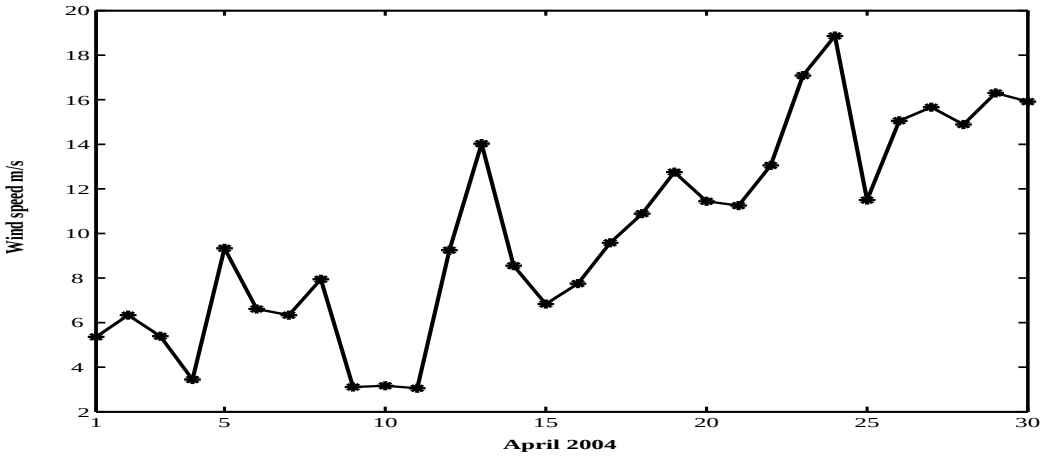



Fig. 8. Evolution of the wind speed at 8 a.m in April 2004

## 4 Fourier transform and autocorrelation

Fourier transform and autocorrelation function provide us a deep insight into the data.

### *4.1 Fourier transform*

Spectral analysis is the most popular tool for observing cyclical phenomena and highliting lead-lag relation within a time series. It is based on Fourier decomposition wich is a way of separating time series into different frequency components that contribute to or influence the time series. This separating procedure is important since dynamic processes operate on different frequencies. In our case, the fast Fourier transform is performed in order to get a periodogram.

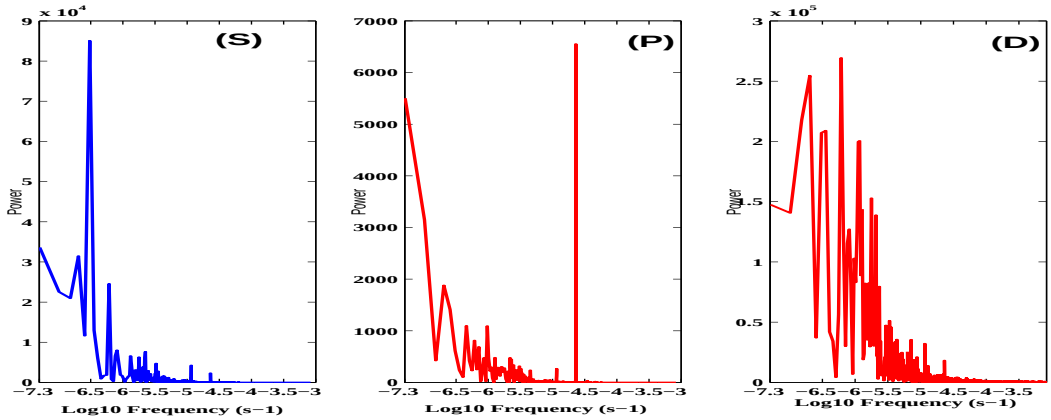



Fig. 9. Power spectrum density of (S,P,D)

In this wind speed periodogram, we observed a first peak corresponding to a 12-hours cycle (2.309 e-5 Hz) and a second one corresponding to a daily cycle (1.155 e-5 Hz) . These two periodic peaks are observed at 0.5-day and 1-day. Concerning the atmospheric pressure, we observed two additional peaks corresponding to a 8-hours cycle (3.4688 e-5 Hz) and a 6-hours cycle (4.623 e-5 Hz).

### *4.2 (P,V) autocorrelation*

The normalized auto-correlation function (NACF) is defined as

$$R(\tau) = \frac{\overline{X(t_i)X(t_i+\tau)}}{\overline{X(t_i)^2}},$$

where the time series $X(t_i)$ is centered, $\tau$ denoting the delay time parameter. The decay of the NACF against the delay time $\tau$ from 0 is an indicator of the relaxation time scale of the system. For the wind speed, it is observed that we have to wait for a long time before flow renewal. The extinction time for the wind flow is about 3,000 minutes, that is 2 days and 2 hours. This specific extinction time and the mean value of the wind speed intensity yield an idea

of the whirl size. For the atmospheric pressure, the extinction is softer. The observed time period is 720 minutes, that is a half-day.

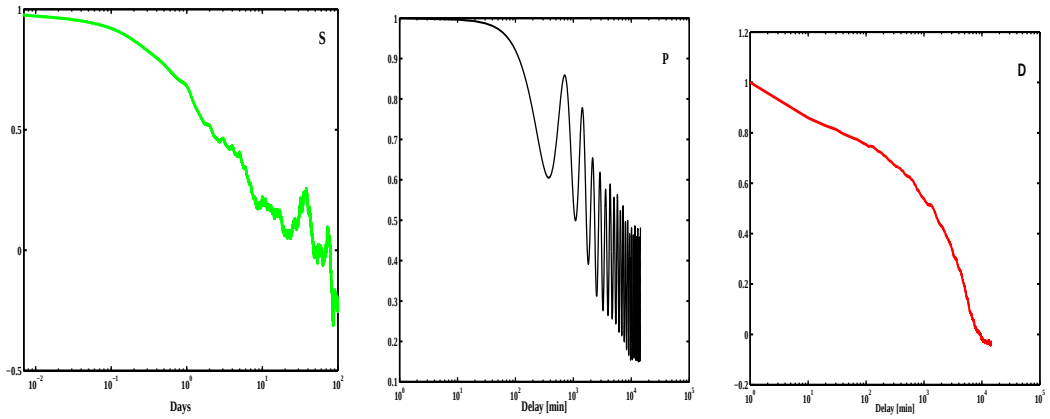



Fig. 10. Autocorrelation function of (S, P, D)

## 5 Wind speed stochastic process

### *5.1 FARIMA model*

The slow decay of the autocorrelation function plotted in section 4 shows a long-term memory behaviour of the wind intensity, confirming a well-known property of meteorological time series. Since [Raftery et al. (1989)] and [Motanari et al (1996)], FARIMA and GARMA models are usually fit to such data. A process $(X_t)_{t>0}$ is FARIMA$(p, d, q)$ [Hosking (1981)] if:

$$\phi(B)\Delta^d X_t = \theta(B)\epsilon_t,$$

where $d$ is a real number $(-1/2 < d < 1/2)$, $\phi(B)$ and $\theta(B)$ are AR and MA polynomials respectively, and $\epsilon_t$ is a white noise.
Its autocorrelation function $\rho(h)$ verifies $\rho(h) \approx c_d h^{2d-1}$ as $h \to \infty$. If

$$log(|\rho(h)|) \equiv \widehat{\alpha} + \widehat{\beta} log(h)$$

is a linear regression model estimation, then $d$ can be estimated by $\widehat{d} = \left(\widehat{\beta} + 1\right)/2$. For our wind speed time series, we obtain $\widehat{\alpha} = 0.9069$, and $\widehat{\beta} = -0.2794$, therefore $d = 0.360$.

*5.2 Periodic trend plus noise model*

We now consider the following model:

$$X_t = T(t) + N_t,$$

where $N$ is the noise process and $T$ is a deterministic (one year period) periodic trend which can be expressed as a linear combination of sine and cosine functions:

$$T(t) = \alpha_0 + \sum_{j=1}^{m} \left[ \alpha_j Cos\left(\frac{2\pi jt}{K}\right) + \beta_j Sin\left(\frac{2\pi jt}{K}\right) \right], \tag{1}$$

where $\alpha_j$ and $\beta_j$ for $j = 1, ..., m$ are real constants given by the least squares estimation method and defined as $\alpha_0 = \frac{1}{K}\sum_{n=1}^{K} X(n)$,

$$\alpha_j = \frac{2}{K}\sum_{n=1}^{K} X(n) Cos\left(\frac{2\pi jn}{K}\right), \quad \beta_j = \frac{2}{K}\sum_{n=1}^{K} X(n) Sin\left(\frac{2\pi jn}{K}\right),$$

the integer $K$ denoting the number of data recorded during one year, here $K = 52706$.
The main problem is the estimation of $m$, the number of harmonics to be considered. When $m$ is too large we get an overfit model so that we penalize large $m$ by using the Akaike Information Criterion AIC [Akaike (1974)] or the Bayesian Information Criterion BIC [Schwarz (1978)] defined respectively by

$$AIC(p) = Ln(\hat{\sigma}_p^2) + \frac{2p}{K} \quad and \quad BIC(p) = Ln(\hat{\sigma}_p^2) + \frac{p.Ln(K)}{K},$$

where $\hat{\sigma}_p^2 = \frac{1}{K}\sum_{t=1}^{K} e_{tp}^2$, $e_{tp}^2 = (X(t) - T_p(t))^2$, and $T_p$ is the tendancy on $p$ harmonics. The value of $m$ is then the value $p$ that minimizes the choosen criterion. Unfortunately, the above formulas show that the larger the number K is the higher $\hat{p}$ will be. In our case, this leads to very large value of $m$ so we turned to the *mean deviation* criterion. We decided that a polynomial of type (1) is a good approximation of the wind speed process when its mean deviation is less than 10. The best solution in this way is obtained for $m = 72$.

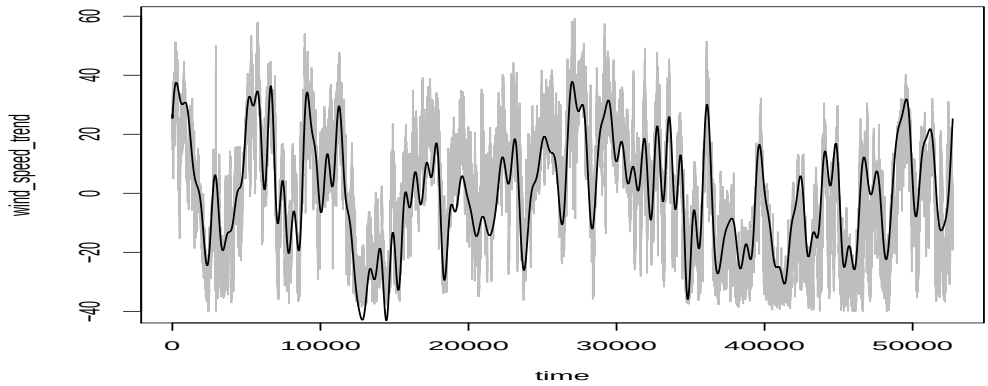



Fig. 11. Trend on 72 harmonics

*5.3 Multifractional noise*

Figure 11 below shows the noise's time series of the wind speed after removing the trend on 72 harmonics.

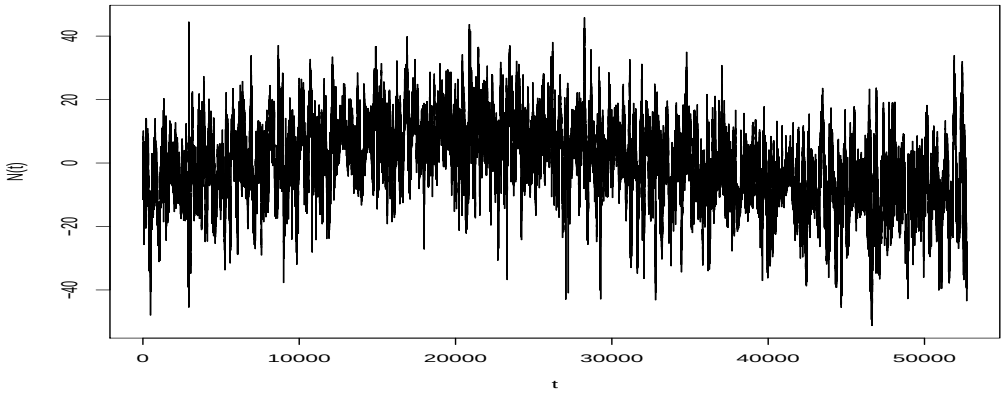



Fig. 12. Noise time series of the wind speed

The noise process $(N_t)_{t\geq 0}$ is defined as $N_t = X_t - T(t)$. In terms of infinitesimal increments we have $dX_t = T(t)dt + dN_t$. Usually the noise process is a standard white noise that is the derivative in the sense of Ito calculus of a standard Brownian motion: $dN_t = dB_t$. However, while Brownian motion is memoryless (the increments are independent), the measurements show that the process of wind speed modulus has a long term memory behaviour. So, in a first approach we have estimated the noise by using a fractional Brownian motion. A fractional Brownian motion (fBm) $(B_t^H)_{t\geq 0}$ with (Hurst) parameter $H$, $0 \leq\ H \leq 1$ is a zero-mean Gaussian process such that

$$Cov[B_t^H, B_u^H] = \frac{|t|^{2H} + |u|^{2H} - |t-u|^{2H}}{2}.$$

The parameter $H$ can be estimated by

$$H_K = \frac{1}{2} log_2 \left( \frac{V_{K/2}}{V_K} + 1 \right),$$

where $V_K$ is the generalized quadratic variation defined by

$$V_K = \sum_{p=0}^{K-2} \left( B^H_{\frac{p+2}{K}} - 2B^H_{\frac{p+1}{K}} + B^H_{\frac{p}{K}} \right)^2.$$

Unfortunately the above estimations applied to the wind speed data show that this parameter is not constant as it depends on the time parameter $t$. Hence we have considered a multifractional Brownian motion (mfBm) $B_t^{H(t)}$, a richer model for which the Hurst parameter is a function of $t$. The estimation of the function $H$ is achieved by using the formula

$$H_{\epsilon,K}(t) = \frac{1}{2} log_2 \left( \frac{V_{\epsilon,K/2}(t)}{V_{\epsilon,K}(t)} + 1 \right)$$

for $0 \leq t \leq 1$, noticing that the process is sampled at discrete time $i = 1, 2, ....., K$. The process $V_{\epsilon,K}(t)$ is the local generalized quadratic variation defined on a neighborhood $\nu_{\epsilon,K}(t) = \left\{ p \in Z, \left| \frac{p}{K} - t \right| \leq \epsilon \right\}$ of $t$ by

$$V_{\epsilon,K}(t) = \sum_{p \in \nu_{\epsilon,K}(t)} \left( B^{H(t)}_{\frac{p+2}{K}} - 2B^{H(t)}_{\frac{p+1}{K}} + B^{H(t)}_{\frac{p}{K}} \right)^2.$$

In our study we have taken $\epsilon = 300/K$.

The results reveal that the maximal value of $H(t)$ does not exceed 0.5. This confirms the high variability of the noise series, since H(t) is an indicator of the smothness of the sample paths of a mBm. The more H(t) is closed to 1 the smoother the sample paths are. Futher H(t) is not a regular function since it presents several variations in short intervals. A close look at the results (figure 13) sugest that H(t) might be well fited by a stochastic process.

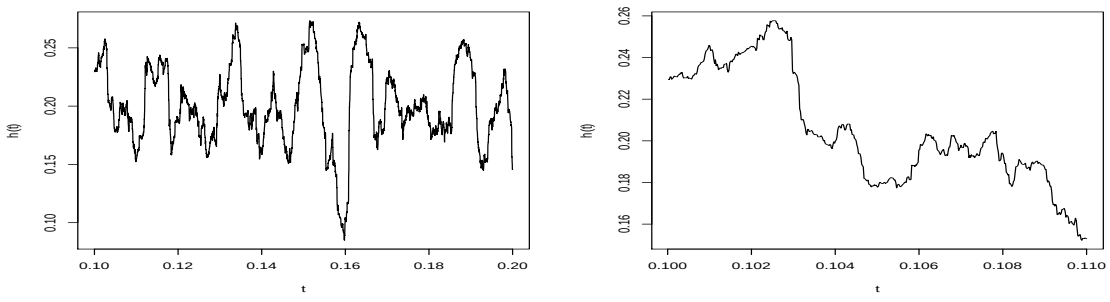



Fig. 13. Evolution of the estimated $H(t)$ on $[0.1, 0.2]$ and $[0.1, 0.11]$

## 6 Continuous cascades and stochastic equation model

The results in the preceding section has put in evidence the local high variability of the wind process but due to the difficulties of identifying a model for the Hurst function $H(t)$, we turned to a model which has a deep interpretability in physics: the continuous cascades model with fractional stochastic integral drift and standard noise [Schmitt (2001)] [Schmitt (2003)] [Schmitt (2005)]. Our motivation has come from the similarity between our measurements time series representation (Fig. 1) and the simulations in the papers of F.G. Smith [Schmitt (2003)] . The estimation of the parameters are derived from some nice computations of F.G. Schmitt [Schmitt (2003)] whose notations are kept to make comparisons easier. We will also show how our numerical results suggest an extension of this continuous cascade model.
Discrete multiplicative cascades model of Yaglom [Yaglom (1966)] appeared after the celebrated pioneering works of Richardson, Kolmogorov and Obukhov on turbulence modelling. Yaglom model can be seen as a weighted random tree depending on an i.i.d. family of random variables (r.v.) $W_{(p,i)}$: the root node is associated with $W_{1,1}$ and with a scale number $L = l_0\lambda_1^n$, each node having $\lambda_1 > 1$ children nodes and all the branches having length $n$. Each of the $\lambda_1^p$ edges $i$ in layer $p$ is associated with $W_{p,i}$ and with a scale number $L/\lambda_1^p = l_0\lambda_1^{n-p}$. Hence each node is associated with the product of the $W_{p,i}$'s which are in the path from the root to this node, and also to a scale number. In particular, each leaf node $x$ is associated with a r.v.

$$\varepsilon(x) = \prod_{p=1}^{n} W_{p,x}.$$

Similar constructions are done for random Polya trees [Paddock et al. (2003)] or in Kraft construction of a continuous random distribution [Kraft (1964)]. Continuous cascades are obtained by letting the depth $n \to +\infty$ while the total scale ratio $\lambda = L/l_0 = \lambda_1^n$ is kept fixed (but large) so that $\lambda_1 = \lambda^{1/n} \to 1^+$. When taking all the $W$'s log-normal, that is

$$W = \exp(\sqrt{\mu \ln 2} g_o - \frac{\mu}{2} \ln 2),$$

where $g_0$ is a standard Gaussian, we obtain a dissipation stochastic process:

$$\varepsilon_\lambda(t) = \lambda^{-\frac{\mu}{2}} \exp(\mu^{\frac{1}{2}} \int_{t+1-\lambda}^{t} (t+1-u)^{-1/2} dB(u)),$$

and its generator,called *singularity* process:

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

has a stochastic drift:

$$d\gamma_\lambda(t) = -\frac{\mu^{1/2}}{2}\left(\int_{t+1-\lambda}^{t}(t+1-u)^{-3/2}dB(u)\right)dt+\mu^{1/2}dB(t)-\lambda^{-1/2}dB(t+1-\lambda).$$

For large scale ratios $\lambda \gg 1$ we have

$$d\gamma_\lambda(t) \approx -\frac{\mu^{1/2}}{2}\left(\int_{t+1-\lambda}^{t}(t+1-u)^{-3/2}dB(u)\right)dt + \mu^{1/2}dB(t).$$

These processes are both stationary with long-range correlations so that under some standard hypothesis, the ergodic theorem ensures the existence of

$$< \varepsilon_\lambda(.) >= \lim_{T\to+\infty}\frac{1}{T}\int_0^T \varepsilon(s)ds.$$

F.G. Schmitt [Schmitt (2003)] (pages 89-90) has computed the moments of the dissipation process

$$< \varepsilon_\lambda^q(.) >= \lambda^{-q\mu/2}\lambda^{\mu q^2/2}$$

$$< \varepsilon_\lambda(.) >= 1,$$

while the autocorrelation function of the singularity process satifies

$$< \gamma_\lambda(.)\gamma_\lambda(.+\tau) >\approx A_\lambda - \mu\ln\tau$$

under the assumption $1 \ll \tau \ll \lambda$.

### *6.1 Estimation of the continuous cascades model*

Our one-year 52706 measurements of the wind speed modulus during 2004 will be represented by a vector denoted $S[1..52706]$. Although the wind process is not stationnary, we will assume that we can fit the above process $\gamma_\lambda(.)$ during a time interval of our study period. However as we need $< \varepsilon_\lambda(.) >= 1$, the vector $S$ has to be translated from a constant $c$ determined by the equation:

$$1 =< \exp(\gamma_\lambda(.)) >=< \exp(S+c)(.) >$$

which gives

$$c = -\ln(\lim_{n\to+\infty}\frac{1}{n}\sum_1^n \exp(S[i])).$$

In our series, the limit is reached for $n \approx 15000$.
The estimation of the *intermittency* parameter $\mu$ can be obtained as follows:

$$< \gamma_\lambda(.)\gamma_\lambda(.+\tau) >\approx A_\lambda - \mu\ln\tau$$

and

$$< \gamma_\lambda(.)\gamma_\lambda(.+3\tau) > \approx A_\lambda - \mu \ln \tau - \mu \ln(3)$$

imply that

$$\mu \approx (< \gamma_\lambda(.)\gamma_\lambda(.+\tau) > - < \gamma_\lambda(.)\gamma_\lambda(.+3\tau) >) / \ln(3),$$

$$\mu \approx \lim_{n\to+\infty} \left( \frac{1}{n-1} \sum_1^{n-1} \gamma[i]\gamma[i+1] - \frac{1}{n-3} \sum_1^{n-3} \gamma[i]\gamma[i+3] \right) / \ln(3).$$

In our case we have taken $1 \ll \tau = 10' = 600$" since the logarithmic decrease is observed during $3\tau = 30' = 1800$".
The estimation of the parameter $\lambda$ is achieved by observing that

$$< \varepsilon_\lambda(t)^2 > = < \exp(2\gamma_\lambda(t)) >$$

implies

$$\lambda^\mu = \lim_{n\to+\infty} \frac{1}{n} \sum_1^n \exp(2\gamma[i]).$$

As $\mu$ is already estimated, we get an estimate of $\lambda$ and that of the depth $d = \ln(\lambda)/\ln(2)$ of the continuous cascade model.

### *6.2 Numerical results on continuous cascades model*

Using $S[1..52706]$, we first plot the autocorellation of the $\gamma$ process with a logarithmic scale on the $x$-axis for the parameter $\tau$. As expected we observe a logarithmic decrease. The above estimations yield $\mu \approx 0.3378$ and $\lambda \approx 20,000,000$

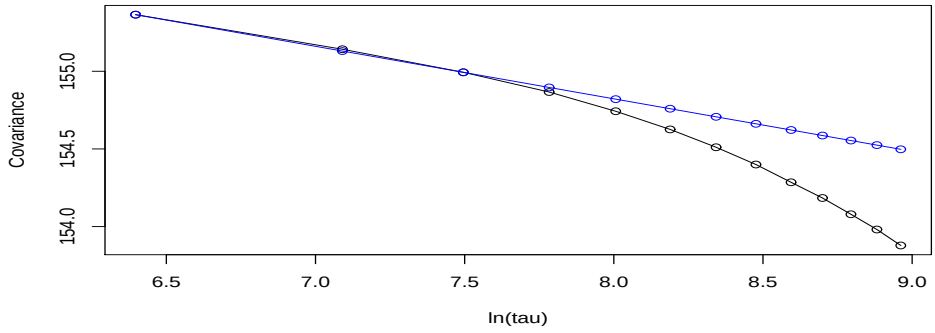



Fig. 14. Autocorrelation function and logarithmic decrease

that is $d = 23$ levels of cascades.
Further, as the wind speed is not stationnary we have also considered sliding windows of size 25000 by considering $S[1..25000], S[2..25001] \ldots S[27706..52706]$
It is important to observe that during some periods the parameters $\mu$ and $\lambda$ are quite constant but that they evoluate in response to environment change,

putting in evidence some regime changes. Similar behaviour have been observed when taking windows of size 30000, 35000, 40000 and 45000.

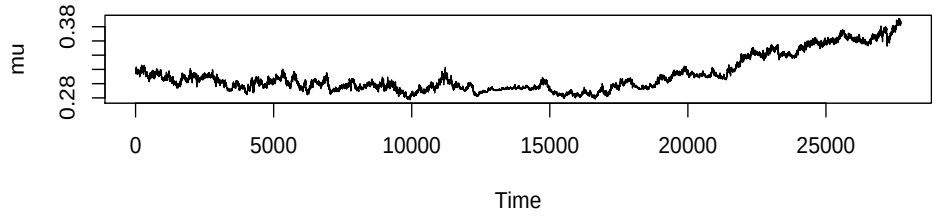



Fig. 15. Evolution of $\mu[1+n..25000+n]$ for $n = 0, 1, 2, ...$

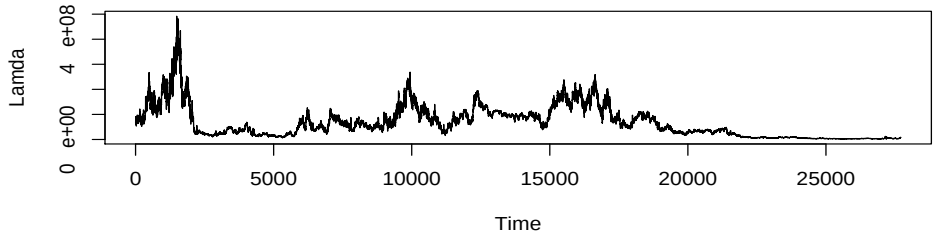



Fig. 16. Evolution of $\lambda[1+n..25000+n]$ for $n = 0, 1, 2, ...$

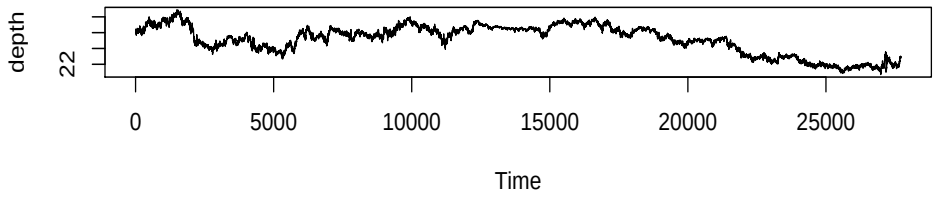



Fig. 17. Levels of cascades [1+n..25000+n] for $n = 0, 1, 2, ...$

### *6.3 Continuous mixtures and hierarchical models*

The above estimations suggest that deeper models should be considered: time series with regime changes or, if the parameters are themselves considered random, hierarchical mixture models. It should be pointed out that the estimations of such models need a more accute sampling, we think of 10 or 20

measurements per second.
We make precise our idea on mixture models. Let $M(\mu, \lambda)$ denote the above cascades model. In order to take in account the environment changes, we assume that the parameters $\mu$ and $\lambda$ are themselves stochastic processes $\mu = (\mu_t)_{t\geq 0}$, $\lambda = (\lambda_t)_{t\geq 0}$. This is connected to the refined notion of stochastic process in *random environment*. We think of the following hierarchical model (as usual $\sim$ indicates the distribution and $|$ conditionality):

$$\begin{aligned} WindSpeed(t)_{|\mu,\lambda} &\sim M(\mu_t, \lambda_t) \\ (\mu, \lambda) &\sim D_0, \end{aligned}$$

where the mixing distribution $D_0$ is the distribution of a pair of stochastic processes, that is, in our case, a distribution on a product of two sets of continuous functions. This model can even be complexified by using a nonparametric Bayesian approach via a continuous-time Ferguson - Dirichlet prior:

$$\begin{aligned} WindSpeed(t)_{|\mu,\lambda} &\sim M(\mu_t, \lambda_t) \\ (\mu, \lambda)_{|P} &\sim P \\ P &\sim Dirichlet(cD_0), \end{aligned}$$

where $P$ denotes a random distribution drawn from a Dirichlet prior and $c$ is a positive constant representing the confidence in that the mean distribution of $P$ is $D_0$.
This model is a generalization to continuous time of some hierarchical models used in classification (see e.g. [Ishwaran (2002)]).

## 7 Conclusion

We have analyzed the behaviour of some atmospheric parameters at the top of La soufrière volcano giving some useful profiles and interpretations which can be of interest for future experiments. We have specially put in evidence the complexity of the wind speed modelling by fitting various models: FARIMA, periodic trend plus multifractal noise, and a continuous cascades model with fractional stochastic integral drift, this last one being fit on some specific periods. The estimations suggest the use of more complex models such as time series with regime changes or stochastic processes in random environment, for example continuous time hierarchical mixture models with Dirichlet prior. This is a challenge for future works since a nice model for wind speed is a preliminary step for getting nice models for pollution prediction.

## 8 Acknowledgments

This research was supported by the European Regional Development Funds (ERDF/FEDER), French government and the Guadeloupe Regional council. The authors would like to thank Dr. Beauducel from The observatoire Volcanologique et Sismique de Guadeloupe for the data set.

## References

[Akaike (1974)] Akaike, H., 1974. A new look at the statistical identification model. IEEE Transactions on Automatic Control 19, 717-723.

[Bernard et al.(2006)] Bernard, M.L., Molinié, J., Petit, R.H., Beauducel, F.,Hammouya, G., & Marion, G., 2006. Remote and in situ measurements of aid gas release from La Soufrière volcano, Guadeloupe. Journal of Volcanology and Geothermal Research 150, 395-409.

[Brevignon et al.(2006)] Brevignon, 2006. Atlas Climatique de la Guadeloupe 150, 395-409.

[Dhonneur (1978)] Dhonneur, G., 1978. Météorologie Tropicale.

[Hosking (1981)] Hosking, J.R.M., April 1981. Fractional differencing. Biometrika 83(1), 165-176.

[Kraft (1964)] Kraft, C. H., 1964. A class of distribution function processes which have derivatives. Journal of Applied Probability 1, 384-388.

[Ishwaran (2002)] Ishwaran H., Lancelot F. J., 2006. Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. Journal of Computational and Graphical Statistics 11(3), 1-26.

[Motanari et al (1996)] Montanari, A., Rosso, R., & Taqqu, M., 1996. Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation, Water Resources Research 33, 1035-1044.

[Paddock et al. (2003)] Paddock, S., Ruggeri F., Lavine M., West M., 2003. A randomised Polya tree models for nonparametric Bayesian inference. Statistica Sinica 13(2), 443-460.

[Raftery et al. (1989)] Raftery, AE., Haslett, J., 1989. Space - time modeling with long memory dependence: assessing irland's wind power resource. Applied Statistics 38(1), 1-50.

[Schwarz (1978)] Schwarz, G., 1978. Estimating the dimension of a model. The Annal of statistics 6, 461-464.

[Schmitt (2001)] Scmitt F. G., Marsan, D., 2001. A causal Stochastic equations generating continuous multiplicative cascades. The European Physical Journal B 20, 3-6.

[Schmitt (2003)] Scmitt F. G., 2003. A causal multifractal stochastic equation and its statistical properties. The European Physical Journal B 34, 85-98.

[Schmitt (2005)] Scmitt F. G., Equations stochastiques continues pour gérer des champs multifractals, 17ème congrès de mécanique, Troyes, 2005.

[Stephanakos (1999)] Stephanakos, C.N, 1999. Nonstationary stochastic modelling of time series with applications to environmental data. PhD, Thesis NTAU.

[Walton et al (1990)] Walton, T.L., Borgman, L.E., 1990. Simulation of non-stationary, non-Gaussian water levels on the great lakes. Journal of Waterways, Ports, Coastal and Ocean Division, ASCE 116(6), 664-685.

[Yaglom (1966)] 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.