International Journal of Modern Physics B

STOCHASTIC MODELING AND FILTERING OF DISCRETE MEASUREMENTS FOR A TURBULENT FIELD. APPLICATION TO MEASUREMENTS OF ATMOSPHERIC WIND. *

C. BAEHR †

Météo-France-CNRS CNRM-GAME URA1357
42 Avenue Coriolis 31057 Toulouse Cedex 1, France
christophe.baehr@meteo.fr

Received 20 11 2008
Revised 22 12 2008

Non-linear filtering of local turbulent fluid measurements was an unexplored domain, in this paper we present original stochastic models and efficient filters to explore it. First we propose non-linear filters for processes of a mean-field law and give the convergence of their particle approximations. We then define the acquisition process of a vector field along a random path, and significantly modify the Lagrangian models of fluids proposed by the physicists to make them compatible with the problem of filtering. The closure of these equations is obtained by conditioning the dynamics to the observations and to the acquisition process. Our algorithm allowed us to filter velocity measurements of a real turbulent fluid in 3D flows.

Keywords: Nonlinear filtering; Lagrangian models; Atmospheric turbulence; Feynman-Kac formulæ

1. The origin of the problem to be considered

As soon as we make measurements of physical quantities, we face the problem of random errors naturally introduced in the signal. Until these random errors, called noise, can be considered negligible with respect to the physical part of the signal, no processing is necessary or only with some linear filters based on Fourier transform, moving averages, numerical filters, etc. Unfortunately, noises could be predominant, or sometimes the interest of our measurements is put on the thinnest structures of the signal where the noise is often the strongest. Then it's necessary to use an optimal filter to estimate the observed state. These filters require to develop mathematical models that represent the physical behavior of the measured system. The aptitude of such optimal estimations to extract the useful signal lies in the

*Proceedings of the workshop modelling geophysical systems by statistical mechanics methods, 27 April - 2 May 2008, Erice/Italy
†Associated member of the Laboratory of Statistics and Probability of the Toulouse Mathematics Institute (UMR 5219), 118 route de Narbonne 31062 Toulouse Cedex 9

pertinence of the signal's model.

For measurements at high frequency (> 10 Hz) in the atmosphere (i.e. in the turbulence domain where the phenomena are thin and fast), we are in a situation where the sensor perturbations could have the same level of energy as the physical signal. For mobile measurements (with balloons, ships, aircraft, ...), compounding to the perturbation of instruments acquired at high rate (for 25 to 200 Hz), there exists a supplementary difficulty in accounting the dynamics of the platform. In this particular case, it is a random excursion onto a random medium.

The nonlinear filter we present here will take advantage of a probabilistic description of the fluid and the trajectory of the measurement system. However we have to model the signal coming from a turbulent fluid as discrete stochastic processes and then propose specific algorithms. We suggest theoretical and practical solutions and their presentation is the aim of this contribution.

2. Filtering of perturbed measurements

As proven by P. Del Moral [3], the exact filtering problem has an integral solution as a Feynman-Kac probability distribution measure. To approximate this distribution we will use a particle genealogical algorithm with a genetic selection (see P. Del Moral [4] for details). These results range far and wide in the answer of nonlinear filtering problems, but do not provide solutions in the case of random media interacting with mean-field. We have developed [5,6] a nonlinear filtering for mean-field processes with specific algorithms, particle systems to approximate mean-field distribution or filtering laws and a strategy to learn hidden states from the observation process. We present these techniques in the next two sections.

Considering the discrete problem, let $\mathcal{X}_n$ be a point of the phase space at step $n > 0$ known by the observation process Y_n and π_n is the law of $\mathcal{X}_n$. In the sequel, for a process $\mathcal{X}$, the notation $\mathcal{X}_{[0,n]}$ stands for the trajectory of the process from initial step to time n. The filtering problem is to estimate the $Law(\mathcal{X}_{[0,n]} \mid Y_{[0,n-1]}) \overset{def}{=} \eta_n$ and the $Law(\mathcal{X}_{[0,n]} \mid Y_{[0,n]}) \overset{def}{=} \hat{\eta}_n$ using the differential stochastic system:

$$\begin{cases} d\mathcal{X}_n = B_n(\mathcal{X}_n, \pi_n)\Delta t + A_n(\mathcal{X}_n)\Delta W_n \\ \;\;Y_n = H_n(\mathcal{X}_n) + \sigma V_n \end{cases} \tag{1}$$

where V_n and ΔW_n are noises given, for example, by standard Wiener processes, A_n and B_n are measurable functions with locally bounded derivatives, H_n is the transfer function and σ is a non-negative parameter. The first equation is the stochastic dynamics model with Markov kernel M_{n+1,π_n} dependant of the mean-field law. We assume that the observation noise V_n is a random process with density. It is then possible to construct a function G_n (see [4]) which gives the potential of a state $\mathcal{X}_n$ with respect to an observation Y_n. This potential function gives a (non-unique) kernel of selection expressed by

$$S_{n,\eta_n}(x_n,\cdot) = G_n(x_n)\delta_{x_n}(\cdot) + [1 - G_n(x_n)]\frac{G_n(.)}{\eta_n(G_n)}\eta_n(\cdot) \tag{2}$$

where δ_{x_n} is the Dirac measure of the singleton set x_n. A well-adapted state $\mathcal{X}_n$ is preserved and an ill-adapted state is reallocated randomly with respect to filtering law η_n. The acceptation/rejection rate is only controlled by the potential G_n.

Therefore the algorithm of nonlinear filtering estimates the law η_n with the sequence (see [5]):

$$\eta_n \xrightarrow{S_{n,\eta_n}} \hat{\eta}_n = \eta_n\ S_{n,\eta_n} \xrightarrow{M_{n+1,\pi_n}} \eta_{n+1} = \hat{\eta}_n\ M_{n+1,\pi_n} \tag{3}$$

This algorithm has a particle approximation (see [5, 6]) using 2 systems; the first, with d elements, learns the mean-field law π_n, and the second, with N particles, learns the laws of filtering η_n and $\hat{\eta}_n$. We prove in [5] that these approximations lead an asymptotic convergence when the number N of particles tends to infinity and for any bounded measurable test function f,

$$\sup_{n\geq 0} \mathbb{E}(|\eta_n^N(f) - \eta_n(f)|^p)^{\frac{1}{p}} \leq (\frac{C(p)}{\sqrt{N}} + \frac{C'(p)}{\sqrt{d}})\|f\|$$

where η_n^N is the particle approximation of η_n, p is a non-negative number, and the constants $C(p)$ and $C'(p)$ depend only on the p parameter. Once the filtering law has been computed, the estimated (also called filtered) state $\overline{\mathcal{X}}_n = \int x\hat{\eta}_n(dx)$. Thus all the moments of the unknown state $\mathcal{X}_n$ are accessible.

3. Acquisition process of a random field along a random path

For measurements on random medium, especially when the sensor is mobile with respect to the medium, we do not have a Markovian model for $\mathcal{X}_n$. However when this environment can be described by a local (Lagrangian) model, an estimation of the signal can be achieved (see 5). We develop a new mathematical object, called the acquisition of a random field along a stochastic path. The acquisition process is simply the values taken by a random vector field of the phase space on each point of a path assumed to be stochastic in the configuration space. Considering the discrete case, this model leads to a prediction/update process well known in the nonlinear filtering methods and seen before in the scheme (3). The acquisition process could have very general purpose far from filtering, for example, in punctual modeling of random media.

For a random path X_n in the physical space E subset of the d dimensional real space and a random field $X'_{n,x}$ in the d' dimensional phase space E', we define the acquisition process of the field X' along the path X by $A_n = X'_{n,X_n}$. If the random path $X_n^{x_0}$, coming from x_0 at the initial step, is a flow of the vector field $X'_{n,x}$, we call $A_n = X'_{n,X_n^{x_0}}$ the Lagrangian acquisition. This definition is obvious when

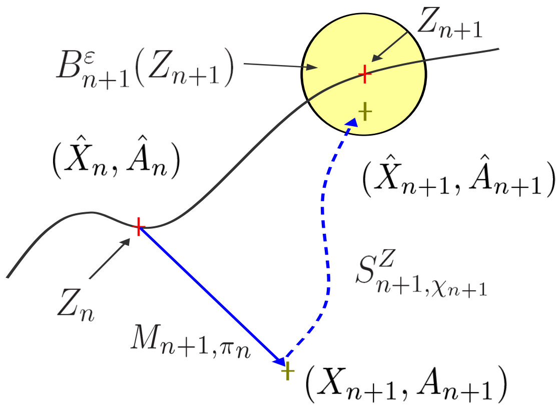

Fig. 1. Evolutionary scheme of a discrete acquisition process for a Lagrangian dynamics.

X' is the Eulerian field of fluid velocities. We will use this description and assume that the medium is locally homogeneous and the sensor path in given by $Z_n \in E$. The locally homogeneous hypothesis allows to define a sequence of balls following the acquisition path $B^{\varepsilon}_n(Z_n) = \{x \in E : |x - Z_n| \leq \varepsilon_n\}$. To solve the acquisition problem, we have to estimate the 2 laws of probability given for a test function by f :

$$\hat{\chi}_n(f) = \mathbb{E}(f(X_n, A_n) \mid X_0 \in B^{\varepsilon}_0(Z_0), \dots, X_n \in B^{\varepsilon}_n(Z_n)) \text{ and}$$
$$\chi_n(f) = \mathbb{E}(f(X_n, A_n) \mid X_0 \in B^{\varepsilon}_0(Z_0), \dots, X_{n-1} \in B^{\varepsilon}_{n-1}(Z_{n-1}))$$

In ref 5, we show that this measure can be learned by an iterative algorithm with prediction/selection step with the scheme

$$\hat{\chi}_n \xrightarrow{M_{n+1,\pi_n}} \chi_{n+1} \xrightarrow{S^Z_{n+1,\chi_{n+1}}} \hat{\chi}_{n+1}$$

where $S^Z_{n+1,\chi_{n+1}}$ is a selection kernel with the same form of (2) using the potential function with the value 1 into the ball $B^{\varepsilon}_{n+1}(Z_{n+1})$ and 0 elsewhere. This prediction/selection procedure of the discrete acquisition for the locally homogeneous case in term of physical state is illustrated with the figure 1.

Now we have the capacity to propose a filter for turbulent flows, coupling a Lagrangian model, the acquisition estimation, and the filtering problem.

4. Adapted Stochastic Lagrangian Model for Nonlinear Filter

To continue with the physical application, we can apply our acquisition process to the measurements of a random field and propose a method of calculating local averages and a nonlinear algorithm to filter data contaminated by noise of intruments.

In the case of turbulent fluids, we will use Stochastic Lagrangian Models (SLM) as proposed by S.B. Pope [2] in the homogeneous cases, or proposed by S. Das and P. Durbin for 3D stratified flows [1]. The filtering of such turbulent flows needs to condition the problem to the observations in order to give relevant estimation of the Eulerian quantities of large eddies. This conditioning to the observations of a Markovian kernel is a new method of closure for a stochastic model.

In its continuous version for 1D or 2D flows, the simplified SLM given by S.B. Pope for isotropic homogeneous turbulence is

$$\begin{cases} dX_t = V_t \; dt \\ dV_t = -\nabla_x < p > \; dt - (\frac{1}{2} + \frac{3}{4}C_0)\frac{\varepsilon_t}{k_t}(V_t - < v >) \; dt + \sqrt{C_0 \varepsilon_t} dB_t \end{cases} \tag{4}$$

where V_t is the Lagrangian velocity, $\nabla_x < p >$ is the gradient of mean pressure, ε_t is the turbulent dissipation rate, k_t the turbulent kinetic energy, $< \cdot >$ are Eulerian means, C_0 is the Kolmogorov constant and B_t is a Wiener process. The term $< v >$ will be approximated by the Lagrangian expectation conditioned to the observation and to the acquisition process $\mathbb{E}(V_t \mid X_{[0,t]} \in B_t^\varepsilon(Z_{[0,t]}), Y_t)$. The turbulent kinetic energy will be half of the variance associated with the same expectation. In this particular case, we choose to model $-\nabla_x < p > \; dt$ by the mean increment of velocities conditioned to the observation and ε_t will be modeled by the conditioned expectation of the square of the increment. We are well aware that this model for the gradient of mean pressure breaks the incompressibility of SLM, but it will be restored by the observation of an incompressible fluid and by the mechanism of acceptation/rejection of the filtering technique. Therefore the Markovian transition M_{n+1,π_n} associated to the real SLM is replaced by the transition $M_{n+1,\hat{\eta}_n}$ where $\hat{\eta}_n$ is the law of the optimal filter.

For 3D atmospheric flow, we derive from the model proposed by Das and Durbin [1] (established for dispersion in a stratified turbulent flow), a stochastic model which could be seen as the dispersion of a fluid particle with respect to a virtual one following the mean flow :

$$\begin{cases} dX_t = V_t \; dt \\ dV_{h,t} = -\nabla_h < p > .dt - C_1 \frac{\varepsilon_t}{k_t}(V_{h,t} - < V >_{h,t}).dt \\ \qquad + C_2.(W_t - < W >_t).\frac{d< V >_{h,t}}{dz}.dt + (C_0.\varepsilon_t)^{\frac{1}{2}} dB_t^{V_h} \\ dW_t = d < W >_t - C_1 \frac{\varepsilon_t}{k_t}(W_t - < W >_t).dt \\ \qquad + C_3.\beta.g.(\theta_t - < \theta >_t).dt + (C_0.\varepsilon_t)^{\frac{1}{2}} dB_t^W \\ d\theta_t = d < \theta >_t - C_4 \frac{\varepsilon_t}{k_t}(\theta_t - < \theta >_t).dt \\ \qquad -(W_t - < W >_t).\frac{d< \theta >_t}{dz}.dt + (C_\theta)^{\frac{1}{2}} dB_t^\theta \end{cases} \tag{5}$$

where $V_{h,t}$ is the horizontal speed, W is the vertical speed, θ is the temperature, $B_t^\bullet$

are Wiener processes and $C_\bullet$ are some constants (see [5]). We use the same methods to close the system by observations. The vertical gradient mean will be learned with the particle approximation using the gradient of the set of particles retained after the acceptation/rejection step of the filtering. After these adjustments, it is not really a model for turbulence anymore, but it may be useful to filter measurements.

We denote $\mathcal{X}_n = (X_n, V_n)$ or $\mathcal{X}_n = (X_n, V_n^h, W_n, \theta_n)$ and the superscript B_n means 'conditioned to be in the ball $B_n^\varepsilon(Z_n)$', the discrete filtering algorithm is therefore described by the following sequence :

$$\mathcal{X}_n^{B_n} \xrightarrow{\text{w.r.t. } S_{n,\eta_n}} \hat{\mathcal{X}}_n^{B_n} \xrightarrow{\text{w.r.t. } M_{n+1,\hat{\eta}_n}} \tilde{\mathcal{X}}_{n+1} \xrightarrow{\text{w.r.t } \mathcal{S}^Z_{n+1,\tilde{\eta}_{n+1}}} \mathcal{X}_{n+1}^{B_{n+1}}$$

where $\tilde{\eta}_{n+1} = Law(\tilde{\mathcal{X}}_{n+1} \mid \mathcal{X}_{[0,n]} \in B_n^\varepsilon(Z_{[0,n]}), Y_{[0,n]})$

5. Application to real 3D wind measurements

The quality of the estimations delivered by our filter needs to be evaluate with different points of view. The first one is the filtering of simulated data for fluid velocities. We do not present this result on 1D or 2D simulated fluids and prefer to check the ability of the method with real 3D flows. We propose to filter real clean data perturbed artificially with numerical noise. The purpose is to filter the corrupted signal and compare it to the signal of reference. First the comparison is visual with the examination of their respective series and their Power Spectral Density (PSD).

We use atmospheric data recorded at 5 Hz from an ultrasonic anemometer in the experimental field of the French Weather Research Center in Toulouse, France, on the 18th of July 2006 between 16h58 and 17h00 UTC. The choice of the date and the hour was only the quality of the measurements to become reference signals. We choose to add a random noise built using the local empirical variance of the reference signals with a Gaussian law. This noise is an upper limit for turbulence, and there is a memory effect which can give dramatic errors if the filter does not return the right parameter of the fluid.

The figure 2 (horizontal wind) and 3 (vertical wind and temperature) show in light blue the noisy signal of the 3 components of wind and temperature, the thick black line is the signal of reference to retrieve. The filter uses the algorithm described in the last section with a particle approximation for system of 800 particles. The red signal is the output of our filter using a SCILAB code.

Even if the perturbations are strong, the filtered signal compares well to the reference component. On the diagram of PSD , the correction is very sharp, with a decreasing of power following the reference spectrum, far from the perturbed signal. In detail, more there are slight differences between the filtered and the reference signals and come from our method itself. Indeed the algorithm estimates the charac-

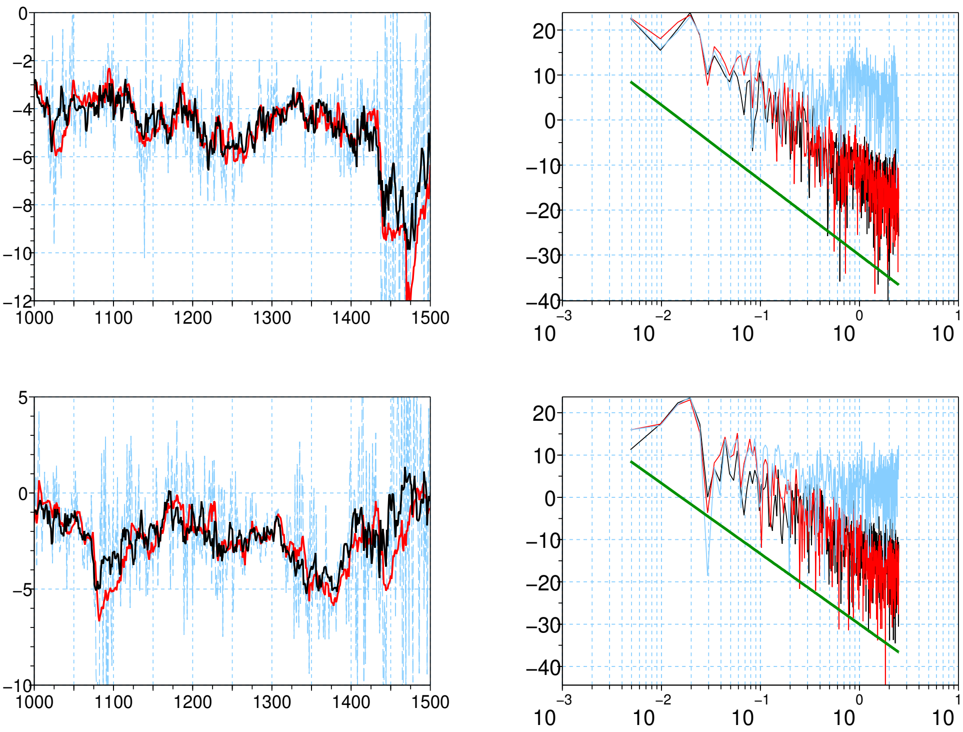

Fig. 2. Left, series of horizontal wind (velocity $(m.s^{-1})$ vs time step number) and right, PSD (with a log-log scale, power (dB) vs frequency (Hz)). The series on the top is the U componant and the bottom the V componant. In light blue the perturbed signal to be denoise, in black the reference signal and in red the filtered with 800 particles.

teristic parameters of the fluid through the model used by the filter, and the ouput is a possible realization of the medium. That's why the spectral correction is very good, as it is the spectrum of the model, while the series show differences but the noise is entirely subtracted out of the perturbed signal. The temperature is the least well-filtered parameter. There is two possible causes for this. First is a worse quality of the model of temperature, the second is a perturbed sensor which not gives a reference signal. At this time there is no answer and its one of our further work.

Our method, which estimates some characteristic terms, retrieves characterizations of turbulence parameters at high frequency which are accessible by the prediction model. It could be the turbulent dissipation rate or buoyancy coefficient or 3-d gradients of Eulerian averages, etc. and the acquisition path of the measurement sensor. The figure 4 gives an example of the series of turbulence dissipation rates and the vertical gradient of temperature.

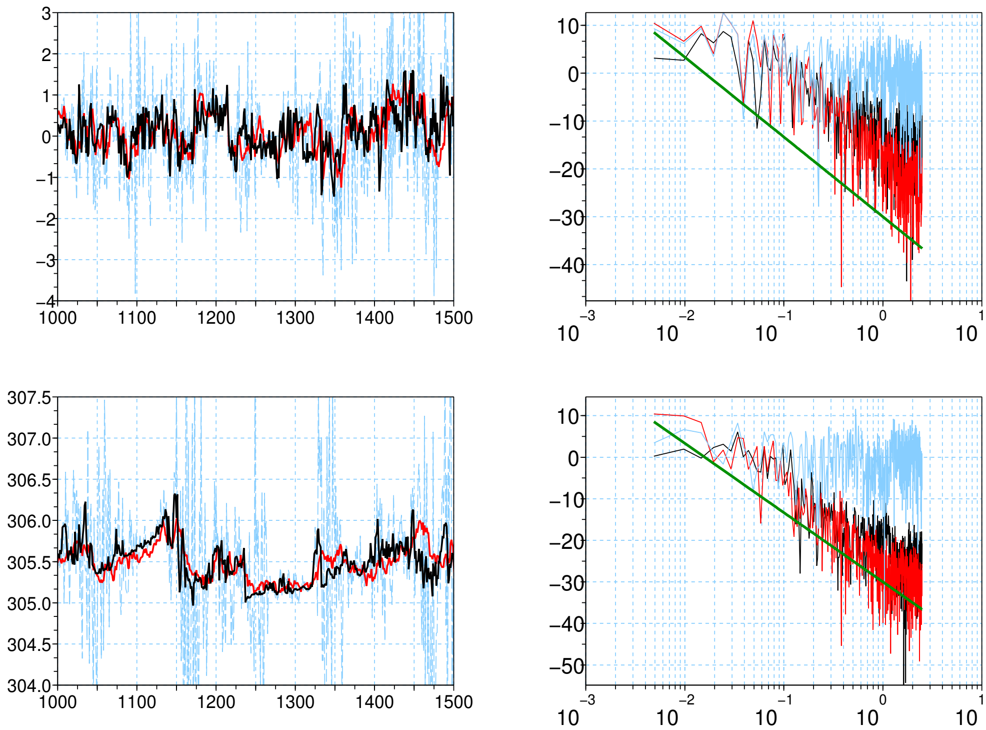

Fig. 3. Series (x-axis is time step number. On the top, the vertical velocity ($m.s^{-1}$), on the bottom, the temperature (K)) and PSD (power (dB) vs frequency (Hz)). In light blue the perturbed signal to be denoise, in black the reference signal and in red the filtered with 800 particles.

These results are very promising, now we have to perform systematic tests giving an objective evaluation of the technique.

6. Some systematic tests of the method

We conduct various tests on simulated or real data to understand the dynamics of our filter. To perform some statistic calculations, we need to run many times the same set of data with our stochastic filter and calculate for the absolute errors empirical mean or variance. The absolute error is defined here as the absolute difference between the reference signal and the filtered. Of course the reference is a realization of the random medium with respect to its law. This test is not really satisfactory, but it's the only one we are able to perform with real data. Because of the complexity of the calculations, we reduce our tests to 1D wind velocities (1 component horizontal of the 3D Wind), and the set of runs have 10 elements. For real data, we choose the measurements of an ultra-sonic anemometer sampled at

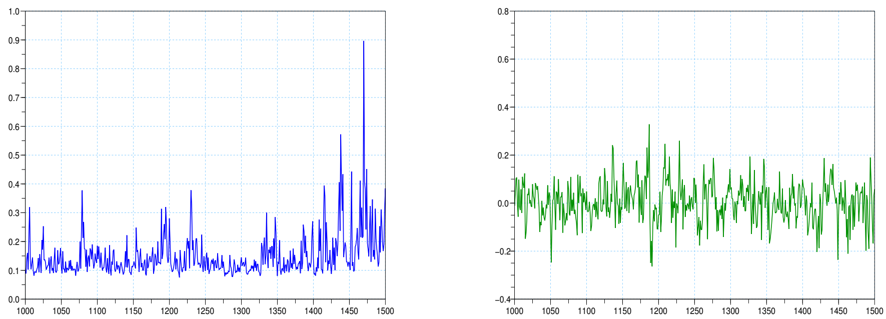

Fig. 4. Left, estimation of ε_n parameter $(m^2.s^{-2})$ and right gradient of mean temperature $(K.m^{-1})$ with 800 particles. X-axis is time step number.

10 Hz located at Saint-Sardos, France, on the 10th of January 2005 between 12h00 and 12h05 UTC. For all the tries, we use the same signal perturbed by an artificial Gaussian noise with standard deviation of 3/2. To test the sensitivity of the particle approximations, we have varied the number of the elements from 100 to 700. To test the pertinence of the adapted 1D SLM used to filter the data, we modified it first by suppressing the turbulent frequency $\frac{\varepsilon_n}{k_n}$ and secondly by omitting the mean-field term $\frac{\varepsilon_n}{k_n}[V_n - < v >]$.

The absolute error of the input data (due to observation noise) is in mean 1.22 ms^{-1} and variance 2.23. On the figure 5, the curve 1 is the absolute error filter with respect to the number of particles, in red a power law curve fit (curve 4). The decreasing of mean errors with the size of the set of particles goes from 0.38 ms^{-1} to 0.28 ms^{-1} (from 0.24 to 0.17 in variance). The shape of curve fit suggest a power law with exponent $-1/5$ instead of $-1/2$. The curve 2 is the modified model with a turbulent frequency put to 1, the curve 3 is the model without mean-field term. It's interesting to notice that for a few particles approximation is better to not take into account the mean-field term. The good representation of mean-field terms needs enough particles. In this case the complete model is the better than its 2 modifications.

Even if the method is not optimal, these numerical tests confirm that mean errors of our filter looks like the theoretical results, and show the necessity to take a set of particle big enough to correctly sample the mean-field law and finally the use of less realistic physical model gives worse results.

To conclude, we can mention that we have compared the calculation with classical methods [7] for turbulent fluxes using the reference signal and those deduced

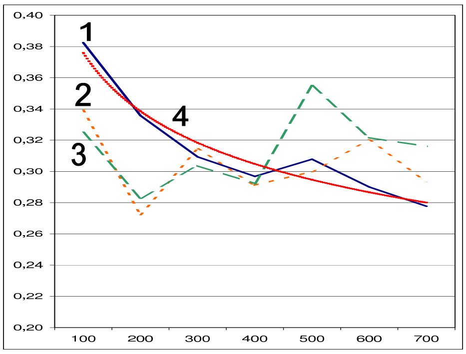

Fig. 5. Mean of absolute errors in ms^{-1} vs number of particles for the complete model (1), the model without turbulent frequency (2), without mean-field term (3) and (4) is a curve fit of (1).

with our filters. We do not present an illustration here, but the results are very comparable. If the classical methods give one number by step of few minutes (due to block averages), our method is able to estimate these turbulent parameters at high frequency (> 5 Hz).

7. Further developments

This work is a first stage, we hope to use in future developments these different techniques to estimate turbulent parameters for high resolution atmospheric forecasting model, or to design new integrated systems for airborne turbulence measurements or to filter numerically more complex atmospheric set of parameters including humidity content, chemical or aerosol concentration, droplets counting, etc. There is also some study to complete the understanding of this type of filter, in particular we need to rely the diameter of the redistribution ball of the acquisition process to physical quantities as the turbulent frequency. Theoretical works are also in progress to modify the algorithm and not break the incompressibility hypothesis as the model proposed here for the gradient of mean pressure could be.

References

1. S. Das and P. Durbin, “ A Lagrangian stochastic model for dispersion in stratified turbulence.”, *Phys. of Fluids* **17**, 025109 (2005).
2. S.B. Pope, *Turbulent Flows.*, (Cambridge University Press, 2000).
3. P. Del Moral, “ Measure valued processes and interacting particle systems. Application to nonlinear filtering problems.” *Ann. Appl. Probab.* **8**, 438-495 (1998).
4. P. Del Moral, *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications.*, (Springer-Verlag, 2004).
5. C. Baehr, *Modélisation probabiliste des écoulements atmosphériques turbulents afin d’en filtrer la mesure par approche particulaire.*, PhD thesis University of Toulouse III - Paul Sabatier, Toulouse Mathematics Institute, 2008.
6. C. Baehr and F. Legland, *Some Mean-Field Processes Filtering using Particles System Approximation.*, in prep.
7. X. Lee and W. Massman and B. Law, *Handbook of micrometeorology.*, (Kluwer Academic Publishers, 2004)