Diagnostic of the Lévy area for geophysical flow models in view of defining high order stochastic discrete-time schemes

Pierre-Marie Boulvard, Etienne Mémin

Diagnostic of the Lévy area for geophysical flow models in view of defining high order stochastic discrete-time schemes.

Pierre-Marie Boulvard, Etienne Mémin

ODYSSEY, Centre Inria de l'Université de Rennes, France

IRMAR – UMR CNRS 6625, Rennes, France
Corresponding author: Pierre-Marie Boulvard,
pierre-marie.boulvard@inria.fr

October 4, 2023

Abstract

In this paper we characterize numerically through two criteria the Lévy area related to unresolved fluctuation velocities associated to a stochastic coarse-scale representation of geophysical fluid flow dynamics. We study in particular whether or not the process associated to the random unresolved velocity components exhibits a Lévy area corresponding to a Wiener process, and if the law of this process can reasonably be approached by a centered Dirac measure. This exploration enables us to answer positively to a conjecture made for the constitution of high-order discrete time evolution schemes for stochastic representation defined from stochastic transport.

1 Introduction

For several years there has been a strong interest in the development of stochastic models of geophysical flow dynamics. This interest is in particular motivated by climate change issues and the advent of ensemble methods for data assimilation and forecasting issues. In the latter case, one looks for a probabilistic quantification of likely scenario or events, while in the former case an optimal coupling is sought between a given dynamical model and available incomplete data of the system of interest. Stochastic dynamics are a natural way to describe this coupling in probabilistic data assimilation methods like ensemble Kalman filters [15] or particle filters [14]. For forecasting issues the objective is rather expressed in terms of modelling the unresolved components of the flow as well as the deviations ensuing from the diverse modelling approximations performed through random variables. The resulting dynamics is then generally formulated in terms of stochastic partial differential equations (SPDEs) rather than with partial differential equations (PDEs). Several schemes have been devised in that purpose. Physics based schemes have been proposed through a stochastic forcing of a negative Laplacian viscosity to inject dissipated energy back to the resolved flow [23]. Eddy parameterization based on a non-Newtonian stress defined on the partially resolved scales and their variability have been studied in [28, 33]. Various reduced order models for climate [3] have been also proposed on the basis of different methodological frameworks. More recently, rigorous derivations of stochastic fluid flow dynamical models relying on stochastic transport and Wiener processes [20, 27] have been devised. Such

models have been successfully numerically tested in various configurations and flows [1, 7, 5, 6, 25, 30]. Some of these models have also been theoretically analyzed in terms of well-posedness properties or in terms of convergence towards their deterministic counterparts [10, 11, 22]. For the stochastic Navier-Stokes equations derived through the modelling under uncertainty (LU), proposed in [27], existence of weak probabilistic solutions (uniqueness of strong solution in 2D) has been demonstrated, along with a convergence toward the deterministic equations in the limit of vanishing noise. This latter results brings strong consistency to the LU setting and highlights the ability to represent at large scales unresolved velocity components by Wiener processes. Additionally, it has been formally shown that such a Wiener process can be obtained as the limit of zero decorrelation of deterministic Lagrangian flow map [9]. A rigorous proof, based on the techniques developed in [13], showing that a diffusion limit of Navier-Stokes equations tends toward a stochastic model of LU type will be the subject of a forthcoming paper.

Numerically, the noise term in those models are often defined from high resolution data provided by a system of corresponding deterministic PDEs [4, 2, 8, 25, 29, 31, 30]. Fluctuations are computed between the original data and low-pass filtered data at a given scale. From these fluctuations, the basis functions for the noise are primarily inferred through the adaptation of techniques like the Karhunen-Loeve decomposition, such as the Proper Orthogonal Decomposition (POD)[26].

The question of the numerical representation of an Itô process is obviously central when one seeks to implement subgrid parameterization. In particular, with the objective of increasing the order in time of discrete evolution schemes in the stochastic case, we face the issue of simulating the iterated integrals of the considered stochastic process. The crux of these iterated integrals lies in the computation of the Lévy area, a feature that captures the finer behavior of the trajectories of the time processes at hand. The computation of the Lévy area is an intricate matter, which is theoretically unavoidable and leads to methods that can be time-consuming, even in approximated cases [17, 18]. Puzzlingly, it has been noticed in [16] for a surface quasi-geostrophic (SQG) stochastic model that the most accurate higher-order time stepping schemes were obtained by neglecting the associated Lévy area.This dynamical model was derived through the LU framework and the noise calibration was performed from high resolution data and a POD procedure. Hence, we advance the following conjecture: within the LU framework, the best way to model the deterministically-generated velocity fluctuations using an Itô process, is by neglecting the Lévy area in the corresponding high-order time stepping scheme.

If this conjecture is verified, it would constitute an incredibly positive property, as the cancellation of the Lévy area, leads in practice to simple high-order schemes that are *in fine* very similar to their deterministic counterparts.

Therefore, in this work, we focus on the study of the Lévy area, first as a way to refine our comparison between large scale fluctuations and their stochastic representation, but above all because the Lévy area is the centerpiece of high-order time schemes in stochastic modelling. Obviously, a theoretical answer to

this question would be ideal. However, we do not know yet how to formulate the problem. Instead, we propose to explore this question numerically with the help of several criterion proposed by Friz, Lyons, and Stroock [19]. We will show that indeed for a prototypical quasi-geostrophic model of ocean circulation the Lévy area associated to the velocity fluctuation between a high resolution model and a low resolution representation is weak. Giving hence some positive indications on our conjecture.

The paper is organized as follows. In section 2 we recall some generalities on the discretization of stochastic differential equations. Section 3 is a detailed explanation of the methods developed for the numerical computation of the Lévy area. While section 4 is a brief description of the pre-processing of the data prior to the Levy area computation, section 5 lays out the more sophisticated, diagnostic processing applied on our data and the results obtained in our experiments.

2 Discretization of a stochastic differential equation

We recall here some notions related to the simulation of stochastic differential equations. We recall in particular the definition of the Euler-Maruyama and of the Milstein schemes.

Let θ_t be a stochastic process over $\mathbb{R}^d$ defined by a stochastic integral equation under the following semi-martingale form:

$$\theta_t = \theta_0 + \int_0^t a(\theta_s)ds + \int_0^t b(\theta_s)dW_s, \tag{1}$$

where W_t is a Wiener process over $\mathbb{R}^d$. We may also write (1) under the following differential formulation:

$$d\theta_t = a(\theta_t)dt + b(\theta_t)dW_t. \tag{2}$$

The mathematical framework in which those equations and their solutions θ_t exist can be found in [21] among many text books. For the numerical integration of this equation we will use a few different limited expansions that we shall put forward below; we suppose that a and b verify the necessary regularity conditions in order for all the terms to be well defined.

Euler-Maruyama scheme: this forward Euler scheme is defined over a sequence of segments $[\tau_{n-1}, \tau_n]$, in the following way:

$$\widetilde{\theta}(\tau_n) := \widetilde{\theta}(\tau_{n-1}) + a\big(\widetilde{\theta}(\tau_{n-1})\big)(\tau_n - \tau_{n-1}) + b\big(\widetilde{\theta}(\tau_{n-1})\big)(W_{\tau_n} - W_{\tau_{n-1}}); \tag{3}$$

we henceforth adopt the following notations:

$$\psi_n := \psi_{\tau_n}, \quad \Delta t := \tau_n - \tau_{n-1}, \quad \Delta W_t := (W_{\tau_n} - W_{\tau_{n-1}}).$$

One has

$$\widetilde{\theta}_n = \theta(t) + \circ(\theta_t - \theta_{\tau_{n-1}}), \quad \forall t \in [\tau_{n-1}; \tau_n]$$

i.e. this scheme is first-order approximation —θ_t-wise— of (1) thus also a $\frac{1}{2}$-order approximation – time-wise –. With classical global Lipschitz and linear growth conditions on the coefficient of (1), the Euler-Maruyama scheme exhibits a strong convergence in $\sqrt{dt}$ – of order $1/2$ – toward a piecewise in time constant solution and a weak convergence in (dt) – of order 1 – toward the expectation of a piecewise time constant solution [21].

Milstein scheme: to increase the (strong) order of approximation to 1 – piecewise in time – we expand $b(\theta_s)$ to the higher order in (1):

$$\begin{aligned} b(\theta_s) &= b(\theta_{n-1}) + \frac{\partial b}{\partial \theta}\big|_{\theta=\theta_{n-1}}(\theta_s - \theta_{n-1}) + \circ(\theta_s - \theta_{n-1}), \\ &= b(\theta_{n-1}) + \frac{\partial b}{\partial \theta}\big|_{\theta=\theta_{n-1}}(W_s - W_{n-1}) + \circ(\Delta W_t), \\ &= b(\theta_{n-1}) + \frac{\partial b}{\partial \theta}\big|_{\theta=\theta_{n-1}}\left(\int_{\tau_{n-1}}^{s} dW_{s^*}\right) + \circ(\Delta W_t), \end{aligned}$$

this second line is a $\frac{1}{2}$-order time-wise approximation (the first line remaining of the first order). Using this approximation to discretize (1), we get the following scheme:

$$\widetilde{\theta}(\tau_n) := \widetilde{\theta}(\tau_{n-1}) + a(\widetilde{\theta}(\tau_{n-1}))(\tau_n - \tau_{n-1}) + b(\widetilde{\theta}(\tau_{n-1}))(\Delta W_t) \tag{4}$$

$$+\frac{\partial b}{\partial \theta}\big|_{\theta=\widetilde{\theta}(\tau_{n-1})}\left(\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dW_{s^*} \otimes dW_s\right). \tag{5}$$

where

$$\left(\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dW_{s^*} \otimes dW_s\right) = \left(\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dW^a_{s^*} \otimes dW^b_s\right)_{a,b\in\{1,\dots,d\}},$$

is an iterated integral of the Wiener process W_t and $f \otimes g$ stands for the tensor product of two vectors f and g. This scheme is a (weak) second order approximation – θ_t-wise – of (1) thus a (strong) first order approximation in time. However, it requires the computation of the iterated integral $\left(\int_{\tau_{n-1}}^{\tau_n}\int_{\tau_{n-1}}^{s} dW_{s^*} \otimes dW_s\right)$. To this end, we use the following decomposition result written for an arbitrary process ϕ:

Lemma-Definition 1. *Let ϕ be a temporal process on I over $\mathbb{R}^d$. The iterated integrals of the components of ϕ over a segment $[t, t'] \subset I$ can be laid out with a symmetric part and an anti-symmetric one; an integration by parts of the symmetric part, using Itô's formula, yields:*

$$\int_t^{t'}\int_t^{s} d\phi_{s'} \otimes d\phi_s = \frac{1}{2}\left((\phi_{t'} - \phi_t) \otimes (\phi_{t'} - \phi_t) - (t' - t)I_d\right) + \mathcal{A}_{[t,t']}(\phi), \tag{6}$$

where I_d is the identity matrix of $\mathbb{R}^d$ and the anti-symmetric part

$$\mathcal{A}_{[t,t']}(\phi) = -\frac{1}{2}\int_t^{t'}\int_t^s \left(d\phi_{s'} \otimes d\phi_s - d\phi_s \otimes d\phi_{s'}\right), \tag{7}$$

is called the Lévy area.

The numerical computation of the symmetric and the anti-symmetric part, and the definition of the Milstein scheme are detailed in [16] for the SQG stochastic system. A numerical estimate for the associated error of the investigated fluid flow models is also provided. In [16] it was surprisingly noticed that the accuracy of the Milstein scheme was improved when the Lévy area was neglected, giving rise to simple integration schemes with no computation of the Lévy area. Strictly speaking those schemes remains of order 1/2, but have an associated error constant that is significantly lowered.

The intuition in this article is, that the identification of the high-resolution part of our data with Wiener processes could induce an over-evaluation of the Lévy area. Inquiries in this direction, and particularly results laid out in [19] led us to set two criteria to characterize the behaviour of the Lévy area associated to high-resolution data of velocity fluctuations.

Asymptotic behaviour of the norm of the Lévy area: The result grounding this criterion was provided in [19] and reads as follows:

Lemma 1 ([19])**.** *Defining for any ϕ a time process over $\mathbb{R}^d$:*

$$\|\mathcal{A}(\phi)\|^{**}_{[0,T]} := \max_{0\leq t\leq T, 0,\leq s\leq t} \sqrt{\left|\mathcal{A}_{[s,t]}(\phi)\right|_{(\mathbb{R}^d)^2}}, \tag{8}$$

we have that for any Wiener process W_t,

$$\mathbb{P}\left(\|\mathcal{A}(W_t)\|^{**}_{[0,1]} \geq R\delta^{1/2} \big| \|W_t\|_{[0,1]} \leq \delta\right) \leq Ce^{-\gamma R^4}, \tag{9}$$

and

$$\mathbb{P}\left(\|\mathcal{A}(W_t)\|^{**}_{[0,1]} \leq \delta^{\eta} \big| \|W_t\|_{[0,1]} \leq \delta\right) \xrightarrow[\delta\to 0]{} 0, \quad \forall \eta > \frac{1}{2}. \tag{10}$$

Those two estimates taken together indicate that the Lévy area of a Wiener process W_t is of order $\frac{1}{2}$. For our numerical experiments we shall use criterion (10), which requires computing the value of $\|\mathcal{A}(\cdot)\|^{**}_{[0,T]}$ either from observation data or from high-resolution simulations. The first part of the results deals with this criterion and provides an answer, albeit an incomplete one in some cases. Moreover, it allows us to state whether the Lévy area of our data can be represented by a Wiener process, but it does not indicate if it is preferable to simulate the actual Lévy area of our data with the one of a Wiener process or simply to neglect it. To overcome this shortcoming we introduce a new criterion.

Conditionnal law of a Lévy area: we rely on the following result [19, 24]:

Lemma 2 ([19, 24])**.** *Let for two coordinates a and b, and $I = [0, T]$*

$$\mathcal{A}_{[0,t]}(\phi)^{a,b} = -\frac{1}{2}\left(\int_0^t \int_{t_0}^s d\phi_\tau^a d\phi_s^b - \int_0^t \int_{t_0}^s d\phi_\tau^b d\phi_s^a\right), \tag{11}$$

be the Lévy area between those two components, and

$$H[\phi^{a,b}](t) := \int_0^t (\phi_s^a)^2 + (\phi_s^b)^2 ds. \tag{12}$$

Let, at last $\|\phi\|_{C^0([0,T])}$ be the C^0 norm over $[0, T]$; then in the case of a Wiener process W_t, the law of $\mathcal{A}_{[0,t]}(W_t)^{a,b}$ conditioned by $\sigma(\|W_t\|_{C^0([0,T])} \leq \delta)$, where σ is the natural sigma-algebra associated to the Brownian motion, corresponds to the law of a Brownian motion with clock $H[W_t^{a,b}](t)$. In other words

$$\mathcal{L}\left(\mathcal{A}_{[0,t]}(W_t)^{a,b}\middle|\|W_t\|_{C^0([0,1])}\right) = \mathcal{L}\left(\mathcal{B}[H[W_t^{a,b}](t)]\middle|\|W_t\|_{C^0([0,1])}\right), \quad \forall t \in I. \tag{13}$$

We get from (13) the following criterion on the marginal laws of the Lévy areas of our observed processes:

$$\mathcal{L}\big(\mathcal{A}_{[0,t]}(W_t)^{a,b}\big) = \mathcal{L}\big(\mathcal{B}[H[W_t^{a,b}](t)]\big), \quad \forall t \in I. \tag{14}$$

The merits of this criterion is that it is usable for any values of the observed time instants whereas (10) requires to be close to 0. Besides, using such laws allows us to vindicate replacing our simulated Lévy area with a null Lévy area by comparing the distances between the law of the actual Lévy area associated to our data, the law of a Brownian motion with the appropriate clock and the Dirac measure in 0. The second part of our results will be specifically devoted to this criterion.

The first part of our article deals with the description of the methods we set for computing the Lévy area. The second part deals with the preliminary processing we performed on the data. In the third part we consider both criteria (10) and (14), first describing the data analysis used for each criterion then laying out our results.

3 Numerical estimation of the Lévy area

In the following, we describe different methods allowing us to evaluate criterion (10). The first one consists in the direct computation of the Lévy area, while the second one yields directly the norm of the Lévy area without computing explicitly the Lévy area itself, which is a costly undertaking, both in terms of computation and storage. Computations will be done on a time process u defined on I over $\mathbb{R}^d$, standing for the high-resolution part of our experimental data.

3.1 Direct computation of the Lévy area

In this section, we undertake to directly evaluate the Lévy area $\mathcal{A}_{[t,t']}(u)$ for all $t, t' \in I$. We approximate its definition (7) with the following discretization:

$$\widetilde{\mathcal{A}}_{[t,t']}(u) = \frac{1}{2} \sum_{s=t}^{t'-1} \sum_{s'=t}^{s} \left((u_{s'+1} - u_{s'}) \otimes (u_{s+1} - u_s) - (u_{s+1} - u_s) \otimes (u_{s'+1} - u_{s'}) \right), \tag{15}$$

first for any t and t', then for all $t, t' \in I$ simultaneously; hereafter we will denote this variable as:

$$\widetilde{\mathcal{A}}_I(u) := \left(\widetilde{\mathcal{A}}_{[t,t']}(u) \right)_{t,t' \in I}$$

3.1.1 Computation of $\widetilde{\mathcal{A}}_{[t,t']}(u)$

Let us now lay out $\widetilde{\mathcal{A}}_{[t,t']}(u)$ in a way that will make its computation easier:

Lemma 3. *The discrete Lévy areas may be written as:*

$$\widetilde{\mathcal{A}}_{[t,t']}(u) = \frac{1}{2} \sum_{s,s' \in I} \mathbb{1}_{t \leq s' < t'} \mathbb{1}_{s' < s} \big(PT_I(u) \big)_{s',s} \mathbb{1}_{t \leq s < t'}, \tag{16}$$

where

$$PT_I(u) := \Big((u_{s'+1} - u_{s'}) \otimes (u_{s+1} - u_s) - (u_{s+1} - u_s) \otimes (u_{s'+1} - u_{s'}) \Big)_{s,s' \in I}. \tag{17}$$

Thus to get every Lévy area on every subsegment $[t, t'] \subset I$ it is enough to compute once $PT_I(u)$, then to sum-up the components in different compositions to make up the Lévy areas. Each sum corresponds to the following inner product:

$$\widetilde{\mathcal{A}}_{[t,t']}(u) = \frac{1}{2} \big\langle M_{t,t'}, PT_I^{trig}(u) M_{t,t'} \big\rangle,$$

where $M_{t,t'} := (\mathbb{1}_{t \leq s < t'})_{s \in I}$ and $PT_I^{trig}(u)$ denotes the upper triangular matrix from $PT_I(u)$.

3.1.2 Computation of $\widetilde{\mathcal{A}}_I(u)$

To get $\|\widetilde{\mathcal{A}}(u)\|_I^{**}$, we need simultaneously to compute all the values of $\widetilde{\mathcal{A}}_{[t,t']}(u)$ for $t, t' \in I$. In order to do so we must assemble the matrix M_I, whose components are vector $M_{t,t'}$ of dimension I:

$$M_I := (M_{t,t'})_{[t,t'] \subset I};$$

then the matrix of all the Lévy areas over I is:

$$\widetilde{\mathcal{A}}_I(u) := \left(\widetilde{\mathcal{A}}_{[t,t']}(u) \right)_{[t,t'] \in I} = \left(\sum_{s,s' \in I} \left((M_I)_{t,t'} \right)_s \left(PT_I^{trig} \right)_{s,s'} \left((M_I)_{t,t'} \right)_{s'} \right)_{[t,t'] \in I}.$$

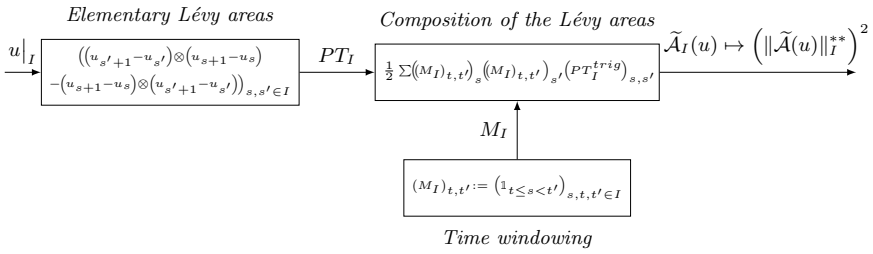

Figure 1: Flow chart of the direct computation of the Lévy areas

A flow chart gathering those processes is summed up in Fig.1

This process is necessary to evaluate criterion (14) and sufficient to evaluate (10). However, the computation of the Lévy area requires working with data of size $d^4 \times T^2$, which is not affordable for long times-series or data of big dimension. We present in the next section another technique enabling the direct computation of the norm of the Lévy area.

3.2 Computation of the norm of the Lévy area from the scalar products of fields

The purpose of this section is to propose a cheaper method for the evaluation of the Lévy area norm. Similar to the previous technique, we will first compute $\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2$ for $t, t' \in I$, before introducing an aggregation process for the simultaneous computation of the whole norm:

$$\left|\widetilde{\mathcal{A}}_{[I]}(u)\right|^2 := \left(\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2\right)_{t,t'\in I}.$$

3.2.1 Computation of $\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2$

To deduce a more efficient computation of $\widetilde{\mathcal{A}}_{[t,t']}(u)$ we start from lemma 3. Upon using once more (16) we may write $|\widetilde{\mathcal{A}}_{[t,t']}(u)|^2$ as:

$$\begin{aligned}\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2 =& \left\langle \widetilde{\mathcal{A}}_{[t,t']}(u), \widetilde{\mathcal{A}}_{[t,t']}(u)\right\rangle_{(\mathbb{R}^d)^2} \\ &= \frac{1}{4}\sum_{\substack{s,s'\\ \mathfrak{s},\mathfrak{s}'\in I}} \mathbb{1}_{t\leq s'<t'}\mathbb{1}_{t\leq s<t'}\mathbb{1}_{t\leq \mathfrak{s}'<t'}\mathbb{1}_{t\leq \mathfrak{s}<t'}\left\langle (PT_I)_{s,s'}, (PT_I)_{\mathfrak{s},\mathfrak{s}'}\right\rangle\end{aligned}$$

Let us now introduce a useful result which lightens the computation of the inner product in the right-hand side of the above equation:

Lemma 4. *Setting:*

$$(PS_I)_{s,\mathfrak{s}} := \langle u_{s+1}-u_s, u_{\mathfrak{s}+1}-u_{\mathfrak{s}}\rangle_{\mathbb{R}^d}, \tag{18}$$

we have

$$\begin{aligned}\left\langle (PT_I)_{s,s'}, (PT_I)_{\mathfrak{s},\mathfrak{s}'} \right\rangle =& (PS_I)_{s',\mathfrak{s}'} \times (PS_I)_{s,\mathfrak{s}} + (PS_I)_{s,\mathfrak{s}} \times (PS_I)_{s',\mathfrak{s}'} \\ & - (PS_I)_{s',\mathfrak{s}} \times (PS_I)_{s,\mathfrak{s}'} - (PS_I)_{s,\mathfrak{s}'} \times (PS_I)_{s',\mathfrak{s}} \end{aligned} \tag{19}$$

To make easier the proof of this lemma we use the following auxiliary result:

Lemma 5. *We have the following equality:*

$$\begin{aligned}&\left\langle (u_{s'+1} - u_{s'}) \otimes (u_{s+1} - u_s), (u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'}) \otimes (u_{\mathfrak{s}+1} - u_{\mathfrak{s}}) \right\rangle_{\mathbb{R}^d \times \mathbb{R}^d} \\ =& \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s+1} - u_s, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d},\end{aligned}$$

Proof.

$$\begin{aligned}&\left\langle (u_{s'+1} - u_{s'}) \otimes (u_{s+1} - u_s), (u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'}) \otimes (u_{\mathfrak{s}+1} - u_{\mathfrak{s}}) \right\rangle_{\mathbb{R}^d \times \mathbb{R}^d} \\ &\quad = \sum_{a=1}^{d} \sum_{b=1}^{d} \left(u^a_{s'+1} - u^a_{s'}\right) \left(u^b_{s+1} - u^b_s\right) \left(u^a_{\mathfrak{s}'+1} - u^a_{\mathfrak{s}'}\right) \left(u^b_{\mathfrak{s}+1} - u^b_{\mathfrak{s}}\right) \\ &\quad = \sum_{a=1}^{d} \sum_{b=1}^{d} \left(u^a_{s'+1} - u^a_{s'}\right) \left(u^a_{\mathfrak{s}'+1} - u^a_{\mathfrak{s}'}\right) \left(u^b_{s+1} - u^b_s\right) \left(u^b_{\mathfrak{s}+1} - u^b_{\mathfrak{s}}\right) \\ &\quad\quad = \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s+1} - u_s, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d}.\end{aligned}$$

□

Demonstration of lemma 4. Let us start with

$$\begin{aligned}\left\langle (PT_I)_{s,s'}, (PT_I)_{\mathfrak{s},\mathfrak{s}'} \right\rangle =& \Big\langle \left((u_{s'+1} - u_{s'}) \otimes (u_{s+1} - u_s) - (u_{s+1} - u_s) \otimes (u_{s'+1} - u_{s'}) \right), \\ & \left((u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'}) \otimes (u_{\mathfrak{s}+1} - u_{\mathfrak{s}}) - (u_{\mathfrak{s}+1} - u_{\mathfrak{s}}) \otimes (u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'}) \right) \Big\rangle_{\mathbb{R}^d \times \mathbb{R}^d},\end{aligned}$$

upon applying lemma 5, we get:

$$\begin{aligned}\left\langle (PT_I)_{s,s'}, (PT_I)_{\mathfrak{s},\mathfrak{s}'} \right\rangle =& \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s+1} - u_s, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \\ & + \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}'+1} - u_{\mathfrak{s}'} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s+1} - u_s, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \\ & - \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \\ & - \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \times \left\langle u_{s'+1} - u_{s'}, u_{\mathfrak{s}+1} - u_{\mathfrak{s}} \right\rangle_{\mathbb{R}^d} \\ =:& (PS_I)_{s',\mathfrak{s}'} \times (PS_I)_{s,\mathfrak{s}} + (PS_I)_{s,\mathfrak{s}} \times (PS_I)_{s',\mathfrak{s}'} \\ & - (PS_I)_{s',\mathfrak{s}} \times (PS_I)_{s,\mathfrak{s}'} - (PS_I)_{s,\mathfrak{s}'} \times (PS_I)_{s',\mathfrak{s}}\end{aligned}$$

□

Remark 1. *The former equality may be further simplified in:*

$$\left\langle (PT_I)_{s,s'}, (PT_I)_{\mathfrak{s},\mathfrak{s}'} \right\rangle = 2(PS_I)_{s',\mathfrak{s}'} \times (PS_I)_{s,\mathfrak{s}} - 2(PS_I)_{s',\mathfrak{s}} \times (PS_I)_{s,\mathfrak{s}'}, \tag{20}$$

but we shall use (19) *in all the following.*

Using those results we get a simplified expression of $\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2$

Proposition 1. *The discretization of the norm* $L^2\left(\left(\mathbb{R}^d\right)^2\right)$ *of the Lévy area may be computed as follows:*

$$\left|\widetilde{\mathcal{A}}_{[t,t']}(u)\right|^2 = \frac{1}{4}\sum_{\substack{s,s'\\ \mathfrak{s},\mathfrak{s}'\in I}} \mathbb{1}_{t\le s'<t'}\,\mathbb{1}_{t\le s<t'}\,\mathbb{1}_{t\le \mathfrak{s}'<t'}\,\mathbb{1}_{t\le \mathfrak{s}<t'}(\Pi_I)^{trig}_{s,s',\mathfrak{s},\mathfrak{s}'} \tag{21}$$

where

$$\Pi_I^{trig} := \Bigg(\mathbb{1}_{s'<s}\mathbb{1}_{\mathfrak{s}'<\mathfrak{s}}\Big[(PS_I)_{s',\mathfrak{s}'}\times(PS_I)_{s,\mathfrak{s}} + (PS_I)_{s,\mathfrak{s}}\times(PS_I)_{s',\mathfrak{s}'} - (PS_I)_{s',\mathfrak{s}}\times(PS_I)_{s,\mathfrak{s}'} - (PS_I)_{s,\mathfrak{s}'}\times(PS_I)_{s',\mathfrak{s}}\Big]\Bigg)_{s,s',\mathfrak{s},\mathfrak{s}'\in I}.$$

3.2.2 Evaluation of $\left|\widetilde{\mathcal{A}}_I(u)\right|^2$

We now assemble $\left|\widetilde{\mathcal{A}}_I(u)\right|^2$ from PS_I and the appropriate characteristics; they are all encased in the following matrix

$$M_I := (M_{t,t'})_{[t,t']\subset I}\,, \quad M_{t,t'} := (\mathbb{1}_{t\le s<t'})_{s\in I}\,.$$

Once PS_I and M_I are computed, we get $\left|\widetilde{\mathcal{A}}_I(u)\right|^2$ thanks to the following computations:

$$\left|\widetilde{\mathcal{A}}_I(u)\right|^2 = \left(\frac{1}{4}\sum_{\substack{s,s'\\ \mathfrak{s},\mathfrak{s}'\in I}}\Big((M_I)_{t,t'}\Big)_{s'}\Big((M_I)_{t,t'}\Big)_{s}\Big((M_I)_{t,t'}\Big)_{\mathfrak{s}'}\Big((M_I)_{t,t'}\Big)_{\mathfrak{s}}(\Pi_I)^{trig}_{s,s',\mathfrak{s},\mathfrak{s}'}\right)_{[t,t']\in I}. \tag{22}$$

These developments are summed up in the form of a flow chart presented in Fig.2.

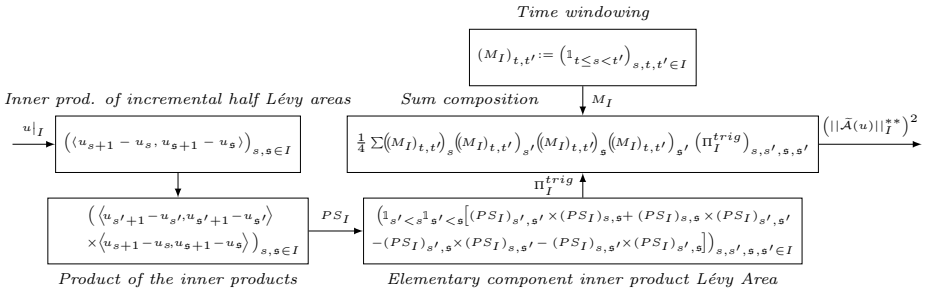

Figure 2: Flow chart of the direct computation of the Lévy areas.

In the next section we describe how we organize and structure our experimental data.

4 Initial structure of the experimental data

In the following the time-process u represents the small scales of observed data. Let us now define its structure. We consider $\mathcal{G}$ a Cartesian grid of a domain $\mathbb{R}^2$ and $\{0;T\}$ a regular sampling of times in $[0,T]$. Then our data is a set of instances over $\{0;T\}$ and $\mathcal{G}$ of the process u. In the first part we will describe the slicing of u in sub-trajectories which allows us to increase our set of realizations. In the second part we will discuss the organization of the data with respect to the way we interpret the raw data: either as the variable of interest or as its increments.

4.1 Slicing in independent trajectories

In order to analyze our data, we slice $\{0;T\}$ into set with length ℓ_{int} and consider that each sub-trajectory of the system is independent. The analysis of this set of sub-trajectories allows us to derive statistics for $\widetilde{\mathcal{A}}_I(u)$ and $\|\mathcal{A}(u)\|^{**}$. In what follows, we will see our data as defined on $I \subset [0,T]$.

4.2 Construction of the increment of the variable

According to definition (15), the computation of the Lévy area of a stochastic process u requires to compute its discrete increments Δu. However, in practice it may happen that the variables are directly available in terms of increments. For instance, in our case, the observed variable is a velocity field u and with respect to our model, we may either directly analyze it or its integral: the displacement field. We have thus two cases:

- The *differential* case: the variable of interest u is directly the observed variable and its increments are computed by differentiating the components of the trajectory.
- The *integral* case: the observed variable is the increment of the variable of interest which is recomputed by integration of the components of the observed trajectory increment.

The two tasks consisting of trajectory slicing and increment computation are pictured in Fig.3

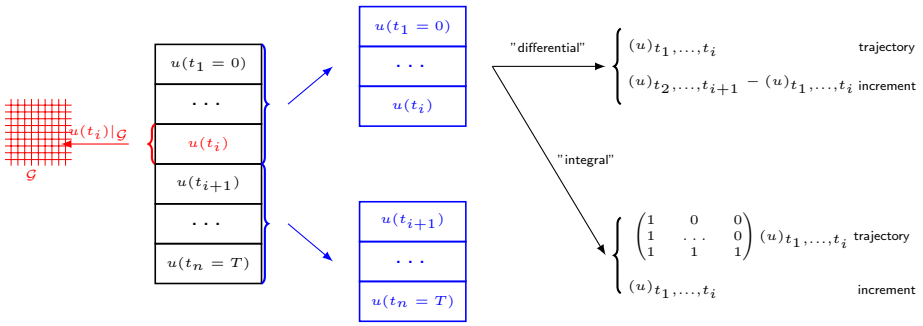

Figure 3: Data organization: the original sample in time and space over $\mathcal{G}$ is sliced in time in trajectory increments, each with length ℓ_{int}; next we define the trajectory under study and its corresponding increment in the differential/integral case.

5 Post-processing of data and results

5.1 Experimental data

The data used in this study corresponds to the simulation of a vertically discretized quasi-geostrophic model resulting in a three layered system, which allows for the development of (3D) baroclinic instabilities. This formulation is quite common in geophysical fluid dynamics (Vallis, 2017). The simulation data have been obtained from the code developed in [25], where the domain considered is a finite box ocean at mid-latitude driven by an idealized (stationary and symmetric) sinusoidal wind forcing. This so-called double-gyre configuration corresponds to an idealization of the North Atlantic ocean circulation, with two counter rotating gyres: a cyclonic circulation over the subpolar basin (the subpolar gyre) and an anticyclonic circulation in the subtropics (subtropical gyre). A western boundary courant in the form of meandering jet – and reminiscent of the Gulf stream – takes place at the interface. In the horizontal dimension, the system is discretized using Arakawa's method with a resolution ($\Delta x = 5$ km), high enough to be eddy resolving. Moreover the temporal scheme follows a third-order Runge-Kutta method. From this simulation we derived our large-scale fluctuations by sub-sampling the data on a coarse grid of 80km with a time sampling of 5 days. Snapshots on the low resolution grid of the meridional and zonal velocity components at the surface are plotted in Fig.4. The residual of the velocity components between the high resolution and the low-pass filtered one are shown in Fig.5.

As described above, the small scale fields u are obtained by a *coarse-graining* process which consists in subsampling the high-resolution data by a given factor followed by a low-pass filtering chosen in order for the data to respect the Nyquist-Shannon subsampling criterion and avoid the apparition of spurious aliasing [32]. In our case, the high-resolution data corresponds to data at 5km resolution while the low resolution is of 80km. Besides, as criterion (10) relates

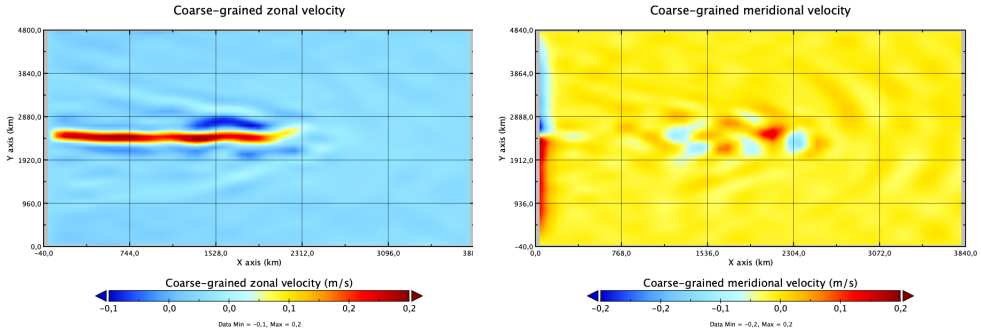

Figure 4: Snapshots on the low resolution grid of the meridional and zonal velocities

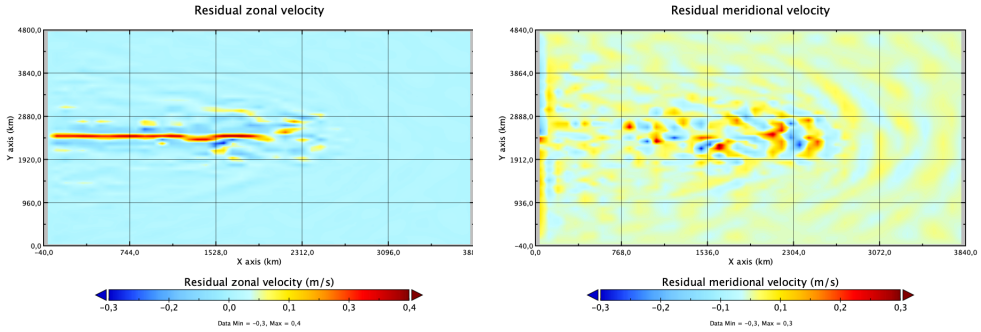

Figure 5: Snapshots of the residual meridional and zonal velocities

to the behaviour of the Lévy area when $\|u\|_{[0,1]} \to 0$, the study of our process will be done on zones which avoid too high-velocity jets while keeping patterns with high variability. The four regions of interest we focus on are delineated in Fig.6. These four regions located just above the jet stream exhibit four different situations. The first two show strong coherent vortices with stronger velocity amplitudes. The third region shows more intermittent fluctuating velocities of lower amplitude while the fourth region exhibits velocities with even lower fluctuations.

The data consists in a time-series of 2928 samples which span over 40 years; in the analysis presented below, the time-series was sliced into 366 trajectories of length $\ell_{int} = 8$. When done with bigger trajectories (namely with length $\ell_{int} = 20$) the analysis yields similar results (not shown here).

5.2 Criterion on $\|\mathcal{A}(\phi)\|^{**}_{[0,T]}$

In the following, we first study the Lévy area associated to process u using criterion (10).

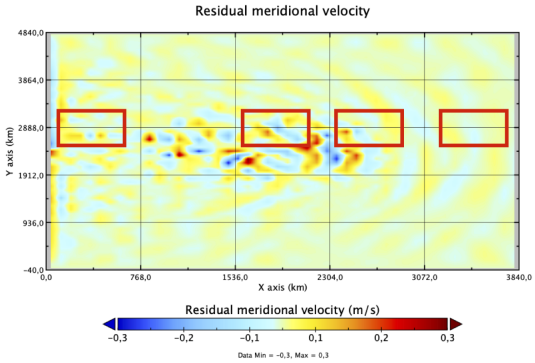

Figure 6: Analysis zones under consideration delineated with red rectangles on the residual meridional velocity component.

5.2.1 Processing of computed data

In order to compute the repartition function for $\mathbb{P}\big(\|\mathcal{A}(u)\|^{**}_{[0,1]} \leq \delta^{\gamma} \big| \|u\|_{[0,1]} \leq \delta\big)$, with instances conditioned on $\delta := \|u_0\|_{[0,1]}$ for all trajectories of u_0, we considered successively the following steps:

1. For u_0 a trajectory, set $\delta := \|u_0\|_{[0,1]}$,
2. Compute $Card(\|u\|_{[0,1]} \leq \delta)$,
3. Compute $Card(\|\mathcal{A}(u)\|^{**}_{[0,1]} \leq \delta^{\gamma} \cap \|u\|_{[0,1]} \leq \delta)$,
4. Evaluate

$$\mathbb{P}\left(\|\mathcal{A}(u)\|^{**}_{[0,1]} \leq \delta^{\gamma} \big| \|u\|_{[0,1]} \leq \delta\right) = \frac{Card(\|\mathcal{A}(u)\|^{**}_{[0,1]} \leq \delta^{\gamma} \cap \|u\|_{[0,1]} \leq \delta)}{Card(\|u\|_{[0,1]} \leq \delta)}.$$

5.2.2 Results

The model adressed in [16] relies on the assumption of representing the *coarse-grained* unresolved velocity fluctuations as Wiener processes, their Lévy area being estimated accordingly. We therefore first test criterion (10) on the Lévy area of the meridional velocity. We wish here to assess the conformity of modelling as a Wiener process the velocity fluctuation obtained from the discrepancy between a high resolution and a low resolution representation (subsampled in space and time) of the data.

The results are plotted Fig.7. The four frames under study are shown from left to right and from top to bottom; the order under consideration is $\gamma = \frac{3}{4}$ which is above the bound set in our criterion, that is $\frac{1}{2}$. We notice that even with a high value of γ the Lévy area is surely of order $\bigcirc(\delta^{\gamma})$ for $\gamma \to 0$, in all frames. Therefore the Lévy area associated to our velocity fluctuation cannot be identified with that of a Wiener process. It must be noticed that in the three

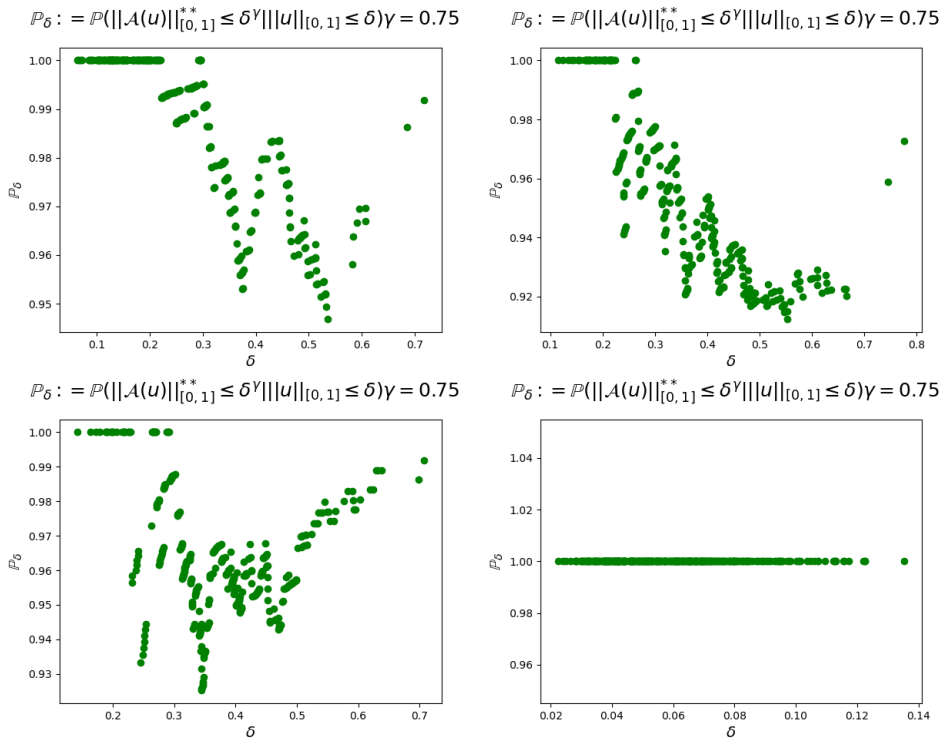

Figure 7: Results of the first criterion (10) on meridional velocities for the four studied areas (#1–4): from upper left (#1) to bottom right (#4)

first frames the δ are farther from zero as we are in a high-velocity regime area, which makes it more difficult to conclude. Our second criterion which is valid in all configurations will confirm those results.

Those results were obtained using the *"differential"* parameterization in the computation of the Lévy areas, in which we considered that our velocity fields are directly the variables that are identified to Wiener processes. The results obtained are gathered on the four graphs of Fig. 8. However, a more classical model would require rather to cope with displacements observations. Hence, considering our velocity fields as increments of the displacements (cf section 4.2) we then use criterion (10) on displacement fields.

Here again, an identification to a Wiener process model seems incorrect, in an even more radical way; maybe this is due to the fact that displacements being smoother than speeds, their Lévy areas may be thus smaller of a higher order γ around 0.

Conclusion on the first criterion: The direct identification of the velocity fluctuations between finely resolved scales and coarse scales as a Wiener process does not appear to be valid, with respect to their Lévy areas. The Lévy areas

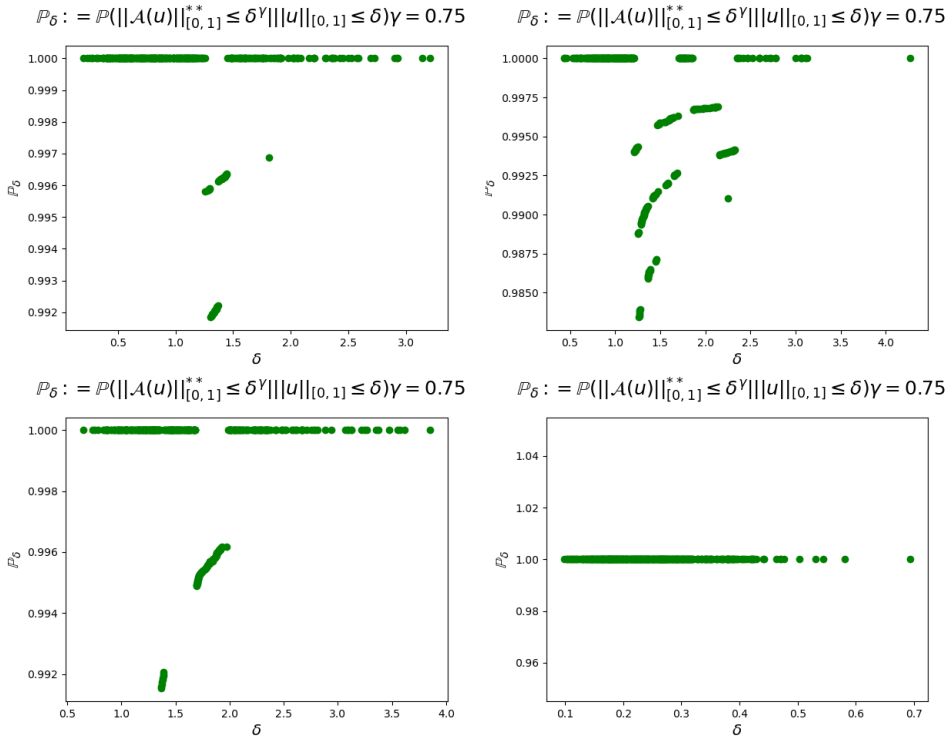

Figure 8: Results of the first criterion (10) on meridional displacements for the four studied areas (#1–4): from upper left (#1) to bottom right (#4)

associated to the velocity fluctuations are notably lower than the ones associated to Wiener processes. However, it is important to note that from a modeling perspective, such a model with uncorrelated velocity increments is perfectly justified in the limit of low noise magnitude [11] or at the decorrelation limit between low and fast-scale processes [9, 12, 13].

However, this discrepancy of the velocity fluctuation process with a Wiener process should be taken into account at some point in the simulation. As shown in [16] a cancellation of the Lévy area leads to simple and efficient time stepping schemes when one aims at representing at coarse-scale a PDE by a SPDE. Yet it is essential to check that such assumption is valid. This is the objective of the following section.

5.3 Criterion on $\mathcal{L}(\mathcal{A}_{[0,T]}(u)^{a,b})$

We next study the Lévy area of the data time-series according to criterion (14).

Let us here recall that this criterion relies on the comparison of the empirical distribution of the Levy area between any two coordinates a, b of a vector field

u and the law of a Brownian motion with a special clock. It aims to compare

$$\mathcal{L}(\mathcal{A}_{[0,T]}(u^{a,b})) \quad \text{and} \quad \mathcal{L}(\mathcal{B}[H[u^{a,b}](T)]).$$

We are therefore facing two tasks: estimating the empirical distribution and simulating the model law.

An important step to do so is to properly structure our data: it consists in a set of trajectories, each of which corresponds to a realization of the Levy area but also of the clock H. Our idea is to slice our set of trajectories with respect to H. Then for each slice we may determine from the trajectories, the empirical distribution of the Levy areas conditionned by the clock:

$$\mathcal{L}\big(\mathcal{A}_{[0,T]}(u)^{a,b}\Big|H[u^{a,b}](T)\big).$$

Furthermore, for each realization of the clock associated with the slice, we may build the law of a Brownian motion with this clock. This procedure will provide us with the law of a Brownian motions conditioned by the instances of the clock derived from our slices:

$$\mathcal{L}\big(\mathcal{B}[H[u^{a,b}](T)]\Big|H[u^{a,b}](T)\big).$$

Then the laws $\mathcal{L}(\mathcal{A}_{[0,T]}(u)^{a,b})$ and $\mathcal{L}(\mathcal{B}[H[u^{a,b}](T)])$, may be recomputed using the following Bayesian expansion:

$$\mathcal{L}\big(\mathcal{A}_{[0,T]}(u)^{a,b}\big) = \sum_{H[u^{a,b}](T)} \mathcal{L}\big(\mathcal{A}_{[0,T]}(u)^{a,b}\Big|H[u^{a,b}](T)\big) \times \mathcal{L}\big(H[u^{a,b}](T)\big)$$

$$\mathcal{L}\big(\mathcal{B}[H[u^{a,b}](T)]\big) = \sum_{H[u^{a,b}](T)} \mathcal{L}\big(\mathcal{B}[H[u^{a,b}](T)]\Big|H[u^{a,b}](T)\big) \times \mathcal{L}\big(H[u^{a,b}](T)\big)$$

We now describe the data structure which directly allows us to get our empirical distribution and then the more intricate simulation of the conditionnal Brownian motion laws.

5.3.1 Computing $\mathcal{L}\big(\mathcal{A}_{[0,T]}(u)^{a,b}\Big|H[u^{a,b}](T)\big)$

To build the conditionnal law $\mathcal{L}\big(\mathcal{A}_{[0,T]}(u)^{a,b}\Big|H[u^{a,b}](T)\big)$, we organize the trajectories and their Lévy areas $(u^a, u^b)_t$ in the two following structures:

1. An along **slices** structure $[H_i, H_{i+1}]$ depending on $H[u^{a,b}](T)$,
2. An along **segments** structure $[A_i, A_{i+1}]$ depending on $\mathcal{A}_{[0,T]}(u)^{a,b}$.

Such a segmentation of the data is pictured in figure 9.

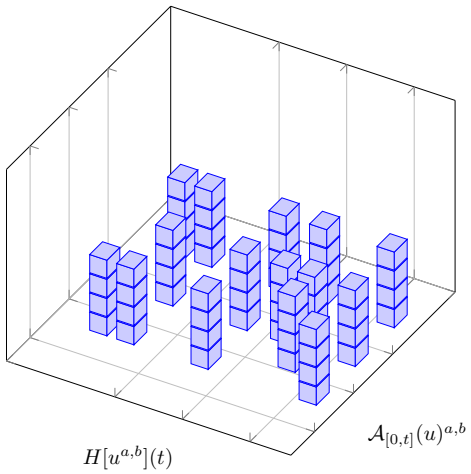

Figure 9: Structure of the set of trajectories in terms of along-slices and along-segments structures.

5.3.2 Computation of $\mathcal{L}(\mathcal{B}[H[u^{a,b}](T)]\Big|H[u^{a,b}](T))$ from instances of the clock.

We now turn to the simulation of the law of a Brownian motion with specific clock. The clock $H[u^{a,b}](t)\Big|_{t\in[0,T]}$ is discretized time-wise, it may thus be considered as a time sequence $(H[u^{a,b}](t_i))_{i\in\{1,\dots,n\}}$. As a consequence the Wiener process simulation of interest boils down to the simulation of $(\mathcal{B}[H[u^{a,b}](t_i)])_{i\in\{1,\dots,n\}}$ that will be performed iteratively.

Wiener process generation: Such a simulation relies on a Lebesgue approximation of the laws of $\mathcal{B}[H[u^{a,b}](t_i)]$, which is denoted as $\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])$ in the following:

$$\mathcal{D}\big(\mathcal{B}[H[u^{a,b}](t_i)]\big) \simeq \sum_{\delta_j\in[0,1]} \big[\mathcal{D}\big(\mathcal{B}[H[u^{a,b}](t_i)]\big)\big]^{-1}([\delta_j,\delta_{j+1}])\,\mathbb{1}_{[\delta_j,\delta_{j+1}]},$$

thus we may approximate $\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_{i+1})])$ by a weighted sum of Gaussians whose means are the center of the inverse image of the $[\delta_j,\delta_{j+1}]$ and whose variations are the $H[u^{a,b}](t_{i+1})$ i.e:

$$\mathcal{D}\big(\mathcal{B}[H[u^{a,b}](t_{i+1})]\big) \simeq \sum_{\delta_j\in[0,1]} \mathcal{N}\Big(\big[\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])\big]^{-1}\big([\delta_j,\delta_{j+1}]\big), H[u^{a,b}](t_{i+1})\Big) \times |\delta_{j+1}-\delta_j|.$$

Here we somewhat improperly refer to the center of the inverse image of the $[\delta_j,\delta_{j+1}]$ as $\big[\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])\big]^{-1}\big([\delta_j,\delta_{j+1}]\big)$. Moreover, in our case, the seg-

ments $[\delta_j, \delta_{j+1}]$ and their inverse images are more precisely defined as quantiles of $\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])$. In practice each distribution

$$\mathcal{N}\Big(\big[\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])\big]^{-1}\big([\delta_j, \delta_{j+1}]\big), H[u^{a,b}](t_{i+1})\Big)$$

is materialized by its associated cumulative distribution discretized on $n_{samples} = 1000$ instances.

The generation of the associated Wiener process consists hence in iterating on $i \in \{1, \ldots, n\}$ the two steps:

1. Evaluation of the centers of our quantiles $\big[\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_i)])\big]^{-1}([\delta_j, \delta_{j+1}])$, which constitutes new *auxiliary quantiles* ,
2. Computation of the *iterated distribution* $\mathcal{D}(\mathcal{B}[H[u^{a,b}](t_{i+1})])$.

Both steps taken together and iterated yield an approximation of the law of a Brownian motion with clock $(H[u^{a,b}](t_i))_{i\in\{1,\ldots,n\}}$. The whole procedure described above and used to evaluate criterion (14) is summed up in the following flow chart of Fig.10

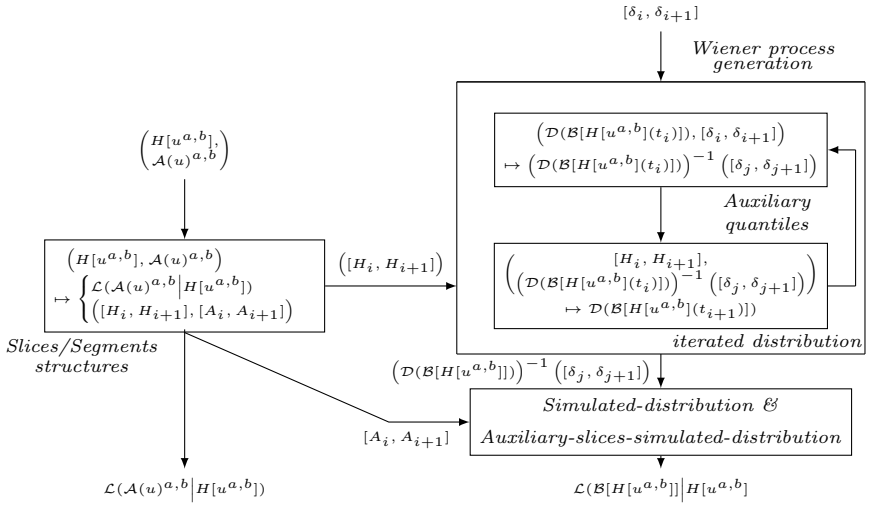

Figure 10: Flow chart for the evaluation of the conditionnal laws use in criterion (14)

5.3.3 Results

We now examine the results obtained for the second criterion (14). First of all, we display in Fig.11, for the four zones (in the same order as the for the previous figures), the laws of the Lévy areas associated to the observed data, the law of the Brownian motion with the appropriate clock and the Dirac distribution in

0. All the laws are evaluated on the support of the law for the Lévy area of the experimental data only. We observe that the simulated distribution related to

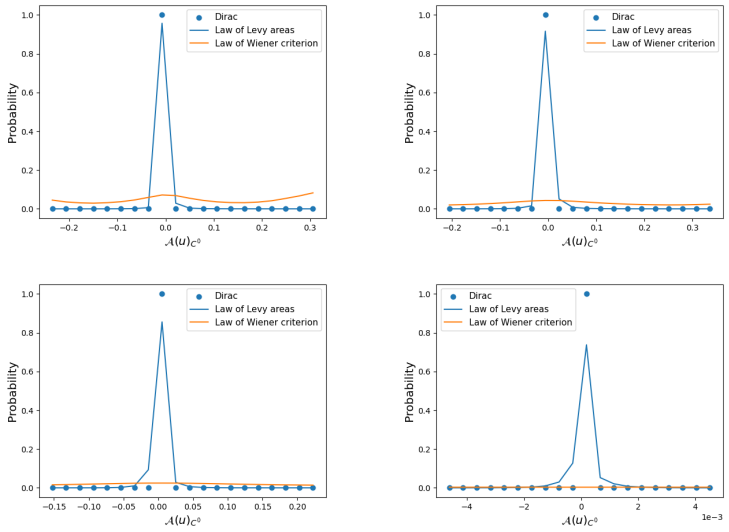

Figure 11: Marginal laws of the Lévy area of the velocity field for the four analyzed zones (#1–4): from upper left (#1) to bottom right (#4).

the Brownian model has a very high standard deviation and is thus too much spread out; the Dirac is therefore what most accurately depicts the behaviour of observed data. To better grasp the difference we display in Fig.12 the first Wasserstein distances between the measures. We display on the one hand (in blue bar) the distance between the law of the observed data and law of the clock-rescaled Brownian motion and on the other hand (in orange bar) the distance between the law of the observed data and the Dirac measure. Let us point out that our distributions are evaluated on the support of the empirical one, alone. Thus while the distance between the empirical distribution and the Dirac one is exact, the distance between empirical distribution and the criterion is underestimated as the latter has support beyond our area of study. Therefore we may not conclude anything from the apparent decrease of this last distance over the frames. What we may take from those results however is that the distribution of the Lévy area is spectacularly closer to the Dirac than to the criterion, with the distance to the Dirac being slightly higher in more turbulent areas. In order to provide the full picture we plotted on Fig.13 and Fig.14 the same criterion for the displacement field (the marginal laws of the Lévy areas and the histograms of the Wasserstein distance, respectively). Those results confirm those obtained from velocity fields.

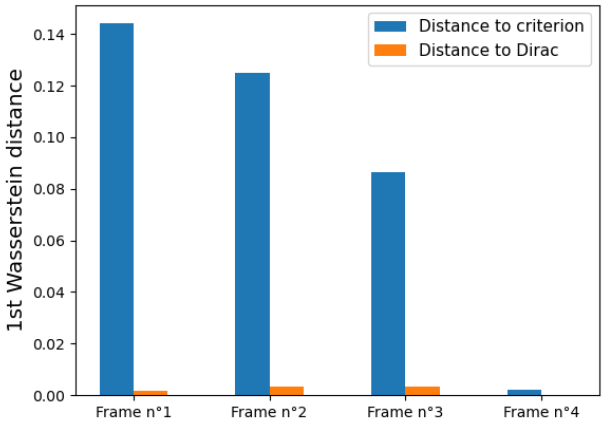

Figure 12: Histogram of the distance between the law of the Lévy area of observed data—the meridional velocity— and the theoretical models in the case of the velocity field for the four frames (from left to right).

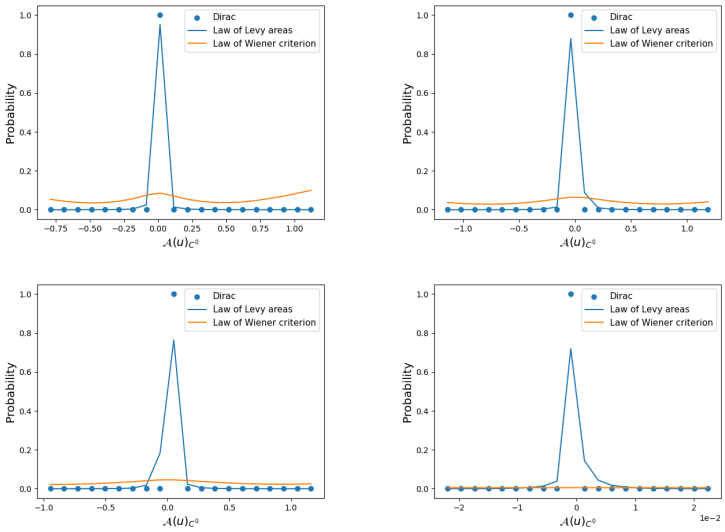

Figure 13: Marginal laws of the Lévy areas of the displacement field for the four analyzed zones (#1–4): from upper left (#1) to bottom right (#4).

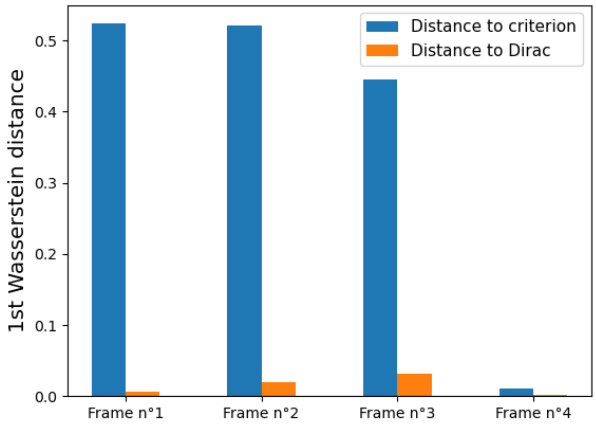

Figure 14: Histogram for the distance between the law of the observed data—the meridional displacement— and the theoretical models for the four frames

Conclusion on the second criterion: The results obtained for criterion (14) corroborate what has been observed for the first criterion. The Lévy area process associated to the velocity fluctuations computed between two resolution scales is small and does not correspond to the Lévy area of a Wiener process. The approximation of the law of the Lévy area process by a centered Dirac measure is justified with an error of order 2 on all the zones analyzed.

6 Conclusion

The two criteria explored here provide us some positive indications on the conjecture made in [16], where it was hypothesized that in the context of stochastic dynamics defined through the modelling under uncertainty formalism, the neglect of the Lévy area computation leads to efficient higher order time stepping schemes. Neglecting the Lévy area of the noise process can be interpreted here as a necessary correction to take into account the fact that (as shown by the first criterion) the stochastic process defined from the eigenfunctions of the empirical covariance of the high-resolution velocity fluctuation data exhibits a Lévy area that does not correspond to a Wiener process. The second criterion highlights that in this context the law of the Lévy area can be approximated by a centered Dirac measure. Bypassing the computation of the Lévy area, which constitutes a very costly process, yield simple integration schemes. These schemes are still theoretically of order 1/2. However, as shown in [16] they lead in practice to lower error constants. They authorized *de facto* to increase the time stepping value.

Obviously, the answer given here is incomplete as it has been explored nu-

merically only on a particular system, and it would be desirable to provide theoretical results on an extended family of dynamics. We don't know yet how to tackle this problem. This will be the objective of future research works.

Acknowledgments

The authors acknowledge the support of the ERC EU project 856408-STUOD. They also warmly thank Dan Crisan for many fruitful discussions and advices.

Data and code availability

For now, the code and data used in the experiments is available upon demand. We will make it accessible in a GitHub repository at a later stage.

References

[1] W. Bauer, P. Chandramouli, B. Chapron, L. Li, and E. Mémin. "Deciphering the Role of Small-Scale Inhomogeneity on Geophysical Flow Structuration: A Stochastic Approach". In: *Journal of Physical Oceanography* 50.4 (2020), pp. 983 –1003.

[2] W. Bauer, P. Chandramouli, L. Li, and E. Mémin. "Stochastic representation of mesoscale eddy effects in coarse-resolution barotropic models". In: *Ocean Modelling* 151 (2020), p. 101646.

[3] J. Berner and Coauthors. "Stochastic parameterization: Toward a new view of weather and climate models". In: *Bull. Amer. Meteor. Soc.* 98 (2017), pp. 565–588.

[4] R. Brecht, L. Li, W. Bauer, and E. Mémin. "Rotating shallow water flow under location uncertainty with a structure-preserving discretization". In: *Journal of Advances in Modeling Earth Systems* 13.12 (2021), pp. 1–28.

[5] P. Chandramouli, D. Heitz, S. Laizet, and E. Mémin. "Coarse large-eddy simulations in a transitional wake flow with flow models under location uncertainty". In: *Comp. & Fluids* 168 (2018), pp. 170–189.

[6] P. Chandramouli, E. Mémin, and D. Heitz. "4D large scale variational data assimilation of a turbulent flow with a dynamics error model". In: *Journal of Computational Physics* 412 (2020), p. 109446.

[7] B. Chapron, P. Dérian, E. Mémin, and V. Resseguier. "Large-scale flows under location uncertainty: a consistent stochastic framework". In: *QJRMS* 144.710 (2018), pp. 251–260.

[8] C. Cotter, D. Crisan, D. Holm, W. Pan, and I. Shevchenko. "Numerically Modeling Stochastic Lie Transport in Fluid Dynamics". In: *SIAM J. on Multiscale Modeling and Simulation* 17.1 (2019), pp. 192–232.

[9] C.J. Cotter, G.A. Gottwald, and D.D Holm. "Stochastic partial differential equations as a diffusive limit of deterministic Lagrangian multi-time dynamics". In: *Proc. R. Soc. A* 473: 20170388 (2017).

[10] D. Crisan, F. Flandoli, and D.D Holm. "Solution properties of a 3D stochastic Euler fluid equation". In: *J. Nonlinear Sci.* (2018), pp. 1–58.

[11] A. Debussche, B. Hug, and E. Mémin. "A Consistent Stochastic Large-Scale Representation of the Navier–Stokes Equations". In: *Journal of Mathematical Fluid Mechanics* 25.1 (2023), p. 19.

[12] A. Debussche and U. Pappalettera. *Second order perturbation theory of two-scale systems in fluid dynamics*. 2022. arXiv: 2206.07775 [math.PR].

[13] A. Debussche and J. Vovelle. "Diffusion limit for a stochastic kinetic problem". In: *Communications on Pure and Applied Analysis* 11.6 (2012), pp. 2305–2326.

[14] P. Del Moral. *Feynman-Kac Formulae Genealogical and Interacting Particle Systems with Applications*. Springer, New York; Series: Probability and Applications, 2004.

[15] G. Evensen. "The ensemble Kalman filter: Theoretical formulation and practical implementation". In: *Ocean Dynamics* (2003).

[16] C. Fiorini, P.-M. Boulvard, L. Li, and E. Mémin. "A Two-Step Numerical Scheme in Time for Surface Quasi Geostrophic Equations Under Location Uncertainty". In: *Stochastic Transport in Upper Ocean Dynamics*. Ed. by B. Chapron, D. Crisan, D. Holm, E. Mémin, and A. Radomska. Springer, 2023, pp. 57–67.

[17] Guy Flint and Terry Lyons. *Pathwise approximation of SDEs by coupling piecewise abelian rough paths*. 2015. arXiv: 1505.01298 [math.PR].

[18] J. Foster, T. Lyons, and H. Oberhauser. "An Optimal Polynomial Approximation of Brownian Motion". In: *SIAM Journal on Numerical Analysis* 58.3 (2020), pp. 1393–1421.

[19] P. Friz, T. Lyons, and D. Stroock. "Lévy's area under conditioning". In: *Annales De L'Institut Henri Poincare-probabilités Et Statistiques* 42 (Feb. 2006), pp. 89–101.

[20] D.D. Holm. "Variational principles for stochastic fluid dynamics". In: *Proc. R. Soc. A* 471.20140963 (2015).

[21] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Vol. 23. Applications of Mathematics. Berlin: Springer, 2010.

[22] O. Lang, D. Crisan, and E. Mémin. "Analytical Properties for a Stochastic Rotating Shallow Water Model Under Location Uncertainty". In: *Journal of Mathematical Fluid Mechanics* 25.2 (2023), p. 29. DOI: 10.1007/s00021-023-00769-9. URL: https://hal.inria.fr/hal-03832450.

[23] C. Leith. "Stochastic backscatter in a subgrid-scale model: plane shear mixing layer". In: *Phys. of Fluids* 2.3 (1990), pp. 1521–1530.

[24] P. Lévy. *Le mouvement brownien*. fr. Mémorial des sciences mathématiques 126. Gauthier-Villars, 1954.

[25] L. Li, B. Deremble, N. Lahaye, and E. Mémin. "Stochastic Data-Driven Parameterization of Unresolved Eddy Effects in a Baroclinic Quasi-Geostrophic Model". In: *Journal of Advances in Modeling Earth Systems* 15.2 (2023).

[26] J. L. Lumley. "The structure of inhomogeneous turbulent flows". In: *Atmospheric turbulence and radio wave propagation* 1.1 (1967), pp. 166–178.

[27] E. Mémin. "Fluid flow dynamics under location uncertainty". In: *Geophys. & Astro. Fluid Dyn.* 108.2 (2014), pp. 119–146.

[28] P. Porta Mana and L. Zanna. "Toward a stochastic parametrization of ocean mesoscale eddies". In: *Ocean Modelling* 79.1-20 (2014).

[29] V. Resseguier, L. Li, G. Jouan, P. Dérian, E. Mémin, and B. Chapron. "New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics". In: *Archives of Computational Methods in Engineering* 28.1 (2021), pp. 215–261.

[30] G. Tissot, A. V. G. Cavalieri, and E. Mémin. "Input-output analysis of the stochastic Navier-Stokes equations: application to turbulent channel flow". In: *Physical Review Fluids* (2023).

[31] G. Tissot, A. V. G. Cavalieri, and E. Mémin. "Stochastic linear modes in a turbulent channel flow". In: *Journal of Fluid Mechanics* 912 (2021), pp. 1–33.

[32] M. Unser. "Sampling-50 years after Shannon". In: *Proceedings of the IEEE* 88.4 (2000), pp. 569–587. DOI: 10.1109/5.843002.

[33] L. Zanna, P. Porta Mana, J. Anstey, T. David, and T. Bolton. "Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction". In: *Ocean Modelling* 111 (2017), pp. 66–80.