



**HAL**  
open science

# New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics

Valentin Resseguier, Long Li, Gabriel Jouan, Pierre Dérian, Etienne Mémin, Bertrand Chapron

## ► To cite this version:

Valentin Resseguier, Long Li, Gabriel Jouan, Pierre Dérian, Etienne Mémin, et al.. New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics. Archives of Computational Methods in Engineering, 2021, 28 (1), pp.215-261. 10.1007/s11831-020-09437-x . hal-02558016v1

**HAL Id: hal-02558016**

**<https://inria.hal.science/hal-02558016v1>**

Submitted on 29 Apr 2020 (v1), last revised 15 Apr 2024 (v2)

**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.

# New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics

V. Resseguier · L. Li · G. Jouan · P.  
Dérian · E. Mémin · B. Chapron

Received: date / Accepted: date

**Abstract** Numerical simulations of industrial and geophysical fluid flows cannot usually solve the exact Navier-Stokes equations. Accordingly, they encompass strong local errors. For some applications – like coupling models and measurements – these errors need to be accurately quantified, and ensemble forecast is a way to achieve this goal. This paper reviews the different approaches that have been proposed in this direction. A particular attention is given to the models under location uncertainty and stochastic advection by Lie transport. Besides, this paper introduces a new energy-budget-based stochastic subgrid scheme and a new way of parameterizing models under location uncertainty. Finally, new ensemble forecast simulations are presented. The skills of that new stochastic parameterization are compared to that of the dynamics under location uncertainty and of randomized-initial-condition methods.

**Keywords** Uncertainty quantification · ensemble forecast · fluid dynamics · data assimilation · model under location uncertainty

## 1 Introduction

Understanding, modeling and tracking high Reynolds flows remain main challenges in current researches. Indeed, beyond economical applications related to

---

V. Resseguier · G. Jouan  
L@b SCALIAN Espace Nobel, 2 Alle de Becquerel, Rennes 35700, France  
Tel.: +33 2 53 48 31 86  
E-mail: valentin.resseguier@scalian.com

P. Dérian · L. Li · E. Mémin  
Inria, Fluminance group, Campus universitaire de Beaulieu, Rennes Cedex 35042, France

B. Chapron  
Ifremer, LOPS, Pointe du Diable, Plouzané 29280, France

weather forecasting and industrial flows analysis, accurate climate projections have become a societal need. The complexity of such systems is mainly due to the non-linear and non-local nature of the evolution laws. These features make large-scale flow structures interact with smaller ones. As such, the large-scale flow components cannot be simulated without the small-scale components. However, the computational expense of solving all the hydrodynamical scales is still today beyond of reach even for turbulent flows of moderated complexity. As an example, today's most accurate oceanic currents numerical simulations use a horizontal mesh resolution of about 1 km (Klein et al., 2008; Gula et al., 2015), whereas solving the "real" equations of fluid dynamics, say the Navier-Stokes equations, would require a grid cell of about 1 cm.

The effects of the unresolved so-called turbulent small-scale fluctuations have to be modeled. Turbulent dissipations, advection corrections and backscatterings need to be introduced to mimic the action of the small-scale processes on the large-scale components. They respectively reduce, move and increase energy of large-scale tracers. But, even using the best subgrid model, the true dynamics cannot be fully recovered by a large-scale model. Strong errors remain as only the unresolved dissipation is in general taken into account. The unresolved subgrid components of tracer or velocity remains by definition unknown. In other words, they are *uncertain*. In this review, uncertainty and stochasticity does not mean noise but rather something that is only known through a probability distribution. Uncertainty can be attached to coherent structures in time and space. To better express this idea, let us introduce a comparison. When looking at a tree, you cannot see all the branches because they are either too small or hidden by leaves. The particular shapes of these branches are uncertain and could be modeled by random processes. However, you know that they respect some features. For instance, each of them is linked to the trunk through one and only one path. These branches are uncertain coherent structures in the same way as unresolved small-scale vortices are. Both must respect appropriate physical laws. Since resolved and subgrid fluid dynamics are coupled, the large-scale resolved dynamics must explicitly take into account this stochastic nature. Understanding, modeling and simulating this randomness or errors is the subject of Uncertainty Quantification (UQ). For sake of concision, this review will not deal with the random models specifically introduced to study extreme events since there is a full literature on this subject (e.g. Franzke, 2017).

The paper is organized as follows. We first motivate the need for UQ in CFD through the data assimilation application. Then, section 3 describes some widely-used CFD UQ tools which are not *a priori* related to the numerical scale truncation. Section 4 reviews the historical stochastic subgrid parameterizations, which were not originally introduced for UQ purposes. Next, section 5 enumerates the more recent schemes designed for UQ purposes in particular the physically-based ones. In this section, we also propose a new stochastic subgrid parameterization adapted to the CFD dissipation scheme through an imposed energy budget. Section 6 focuses on a specific physically-based family of stochastic subgrid tensors: the models under Location Uncertainty (LU)

and the Stochastic Advection by Lie Transport (SALT). After this, section 7 details the metrics used to assess the ensemble forecasts' qualities. Finally, new numerical results are presented for short-term and long-term ensemble forecasts. Several stochastic subgrid parameterizations are compared on a simplified model of geophysical fluid flow dynamics.

## 2 Data assimilation, a main motivation

To introduce this review topic, let us first give a brief description about data assimilation and filtering method. Even beyond fluid dynamics applications, coupling numerical model simulations and measurements is of great interest. This is called data assimilation. Some of these methods, derived from variational principle and optimal control theory, are deterministic (Le Dimet and Talagrand, 1986). A functional criterion is optimized in terms of control variables (such as initial conditions) to drive the model as closely as possible to the observations. For instance, in 4D-Var algorithms solutions are functions of time and space but also of the initial conditions. Variational optimization performed using the adjoint tangent dynamical model allows to infer an initial condition with a trajectory that matches at best the observations (with respect to a given distance). This review will instead be placed in the framework of probabilistic data assimilation methods, such as smoothing and above all filtering (Doucet et al., 2001; Doucet and Johansen, 2009; Candy, 2011; Papadakis et al., 2010). These methods allow combining a random dynamical model together with noisy and partial observations of the system of interest to drive an ensemble forecast – a set of realizations – along time. Hybrid approaches also exist, such as 4D-EnVar (Buehner, 2005; Liu et al., 2009; Yang, 2014; Bocquet and Sakov, 2014; Yang and Mémin, 2017, 2018; Sakov et al., 2017).

In probabilistic data assimilation algorithms, the randomness of the model is meant to capture the errors of the dynamical system whereas randomness of observations represents the measurement errors. If the model noise is prominent, the estimated filtered variable relies principally on the observations. Conversely, when the observation noise is dominant, the filtered variable trajectory is mainly driven by the model. For this reason, an accurate design of the model errors is crucial in weather and climate communities (Allen and Stainforth, 2002; Penland, 2003a,b; Berner et al., 2015). Furthermore, due to the huge state-space dimension ( $\propto 10^{7-9}$ ), the ensemble size is usually very small in comparison ( $\propto 10^{1-2}$ ). Thus, the randomness of the dynamical model has to be very efficient. Ensemble members, also called particles, have to quickly spread in the phase space. At the same time, each particle should remain a “physically plausible realization” to focus on meaningful regions of the state space.

We will now describe several ways to represent the randomness of the fluid dynamical model. That randomness can have several sources: e.g. wrong

physics, unknown parameters or initial conditions. We first detail the approaches which are not *a priori* related to the unresolved components of the dynamics. Then, we will review random parameterizations which model energy backscatterings and/or errors induced by the subgrid dynamics.

### 3 UQ not directly related to the coarse-resolution-induced errors

#### 3.1 UQ from parameters, boundary conditions and forcings

Some authors inject randomness through the parameters. Indeed, parameters like viscosity, initial and boundaries conditions are often assumed random (e.g. Le Maitre et al., 2002; Sapsis and Lermusiaux, 2012). The chaotic nature of fluid dynamics increases quickly the eventual errors associated to these parameters. Other authors study the uncertainty arising from forcings. In particular,  $CO_2$  concentration conditions are difficult to specify in climate sciences. Lucarini et al. (2014) approach this problematic with Ruelle response theory.

#### 3.2 Random initial conditions

In fluid dynamics, random initial conditions have been widely used for both UQ and predictability studies (e.g. Métais and Lesieur, 1986). For a long time, operational weather forecast centers had relied on random perturbations of initial conditions to spread the ensemble forecasts (Ehrendorfer, 1997; Buizza, 2016b). Different types of perturbations were proposed to that end. For instance, the European center (ECMWF) relied on the so-called singular vectors method (SV) while the American center (NCEP) used the bred vectors (BV). The common idea is to perturb the initial condition along the few directions that will lead to a maximum ensemble variance in the next future. The SV are obtained by maximizing the ensemble variance after a finite-time linearized simulation. The linearization enables to solve the optimization problem with the adjoint equations and a singular value decomposition. The amplitude of the perturbation is then tuned by inflation (see later in this section for a description of the inflation method). In contrast, the bred vectors computation does not need an adjoint code nor a linearization. Two simulations – perturbed and not perturbed – are launched from a previous time. At the current time of interest – the time of the forecast initial condition – the difference between the two simulations is rescaled to give the bred vectors.

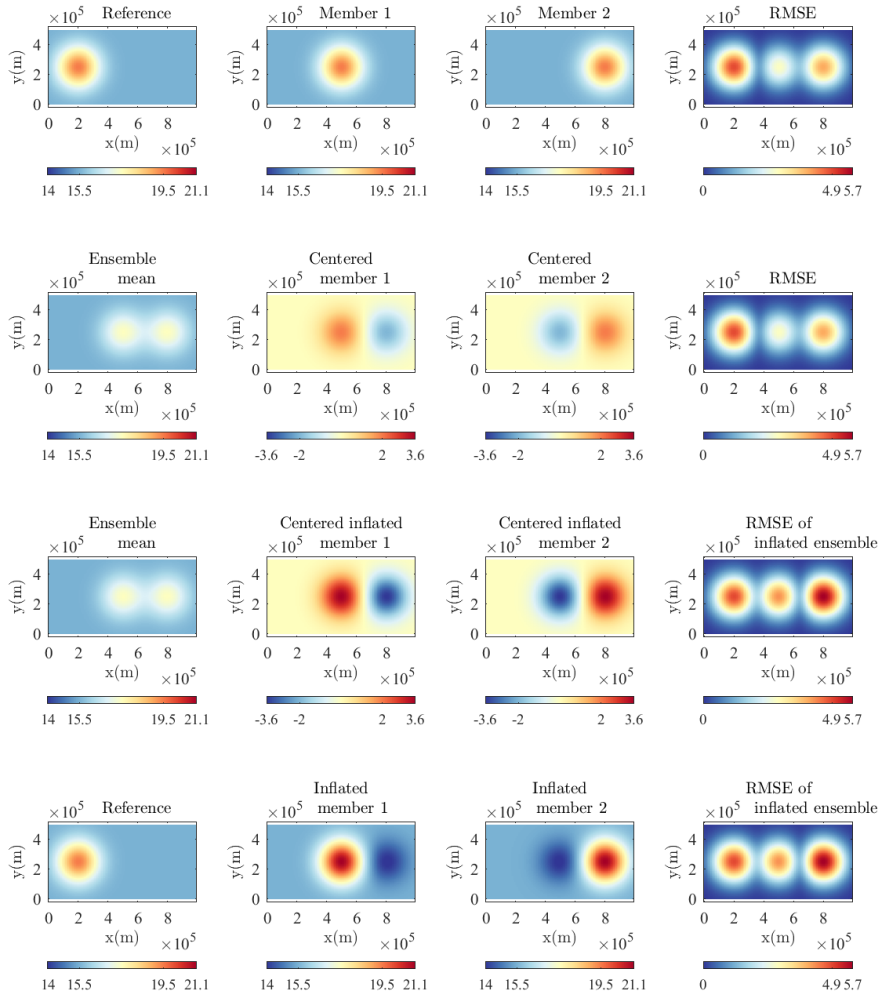
In the geophysical data assimilation communities, the initial condition random perturbation is now known to be underdispersive, i.e. it underestimates the errors related to the coarsening of the state variable dynamics (Berner et al., 2011; Mitchell and Gottwald, 2012; Gottwald and Harlim, 2013; Franzke et al., 2015). As such the model is overconfident. When an observation – often far from the ensemble – is assimilated, only very few particles – and in the worst case only one – are considered relevant. The others are discarded. This

degeneracy is referred to as filter divergence. The weakness of the method can be explained by at least two facts. At the initial time, the random perturbations live in a huge state space. Computational limitation leads to the setup of only a small-size ensemble, thus spanning only a small part of the state-space (Mitchell and Gottwald, 2012; Gottwald and Harlim, 2013). As a consequence, without any adhoc compensation, the ensemble variance is underestimated. Moreover, the initial condition random perturbations are injected at small scales and are hence quickly diffused by the subgrid tensor. From a dynamical system point of view, subgrid tensor makes small-scale components of the solution more stable. Without the fully resolved non-linear mechanisms yielding an energy redistribution, the particles tend to asymptotically align with the most unstable directions (Trevisan and Uboldi, 2004; Trevisan and Palatella, 2011; Mitchell and Gottwald, 2012; Sapsis, 2013; Gottwald and Harlim, 2013). Thus, as time evolves the ensemble spans a smaller and smaller space.

To mitigate the ensemble variance underestimation, a famous compensation method exists. It is called covariance inflation (Anderson and Anderson, 1999). The ensemble covariance is increased (in an additive or multiplicative way) by a carefully-tuned parameter. This scalar can be identified through statistics estimation in the observation space (Tandeo et al., 2018). Indeed, in this space, the innovations are the differences between the predicted and the actual observations. By simple algebra, the variance estimation of that innovation – before the analysis step – can be related to the missing inflation factor, the badly-scaled state-vector covariance, the measurement-error covariance, and the matrix linking state vector and observations. Some operational weather forecasts centers rely on inflation (Raynaud et al., 2016). In any cases, the ensemble covariance is often erroneous and such a compensation may lead to nonphysical behaviors, as exemplified in figure 3.2. Indeed, an ensemble spread underestimation – says a factor 2 – due to for instance an overconfidence in the position of an eddy would lead to a (variance) inflation factor of 4. Thus, in one realization of the ensemble, if the eddy does not overlap with mean eddy, it will stay in the same place but will become 2 times stronger. Therefore, the ensemble mean square error will increase without pushing any realization closer to the truth. A popular inflation variant was introduced by Desroziers et al. (2005). From the *a priori* innovation covariance and the cross-covariance between *a priori* and *a posteriori* innovations, it is possible to correct both the badly-scaled state-vector covariance and the measurement-error covariance. Nevertheless, similar drawbacks may be expected.

#### 4 Stochastic backscattering

Other works preferably address the modeling of errors related to wrong dynamics. Indeed, as long as all the scales are not resolved, subgrid dynamics are modeled rather than resolved. This introduces errors which grow in time due to chaotic behaviors. A natural way to address this UQ is to continuously introduce noise in the dynamics. Evolution laws no longer rely on bulk pa-



**Fig. 1** Inconsistency introduced by covariance inflation for a 2-member ensemble of temperature fields (in Celsius degrees). Here, the optimal inflation factor is about 1.8. The inflation increases the variance without reducing the bias (even though the bias will be modified in the following data assimilation analysis step). Thus, it increases the point-wise Root Mean Square Error (RMSE, see section 7.2 for a definition). Here, inflation makes ensemble members further away from the reference. Moreover, the inflation artificially creates water at  $21.1^\circ$  and at  $14^\circ$  from a water at temperature  $T \in [15.5^\circ, 19.5^\circ]$ .

parameterizations *i.e.* deterministic models. They become Langevin equations expressed through Stochastic Differential Equations (SDE) or Stochastic Partial Differential Equations (SPDE). Note that the Langevin equations can be non-linear with complicated random forcing and memory terms.

Langevin equations are not new in fluid dynamics. Stochastic subgrid parameterizations have been developed before the use of ensemble forecasts in CFD. Nevertheless, most of these historical stochastic schemes have not been developed for UQ purposes – even though they can often be used to that end.

These first stochastic subgrid parameterizations generally try to mimic the intermittent back-scattering of energy from small scales toward larger scales. Several recent UQ methods are also based or inspired from the stochastic backscattering literature. We briefly describe this literature below.

#### 4.1 Stochastic Lagrangian methods

First, let us mention the stochastic Lagrangian models introduced by S. B. Pope and coauthors (see Pope, 1994, for a review). The Lagrangian fluid particles represent independent realizations. Each particle moves with a local velocity which is itself the solution of a randomized Navier-Stokes model in a Lagrangian form. In addition to the mean pressure forcing, the authors assume a relaxation mechanism toward the mean velocity and an additive space-time white noise. Its amplitude is proportional to the dissipation  $\epsilon_D$ . In order to compensate the induced mean kinetic energy increase, the relaxation term includes a term proportional to  $\epsilon_D/\text{TKE}$  where TKE denotes the turbulent kinetic energy. Depending on the complexity of the relaxation matrix, the full model is referred to as simplified Langevin model (SLM) or generalized Langevin model (GLM) (Pope, 1983). In the second case, the relaxation matrix also depends on the mean shear and mean Reynolds stress tensor. The mean terms can be computed by averaging over particles since they are independent. As all coefficients of the SLM and GLM are deterministic, the solution is a Gaussian process. The Langevin models are simulated through Monte Carlo Markov Chain (MCMC). The refined Langevin model (Pope and Chen, 1990) is a variant which replaces the mean dissipation by a stochastic one in the SLM and GLM. The dissipation is assumed to follow another stochastic equation which makes the dissipation probability law close to a log-normal distribution (*i.e.* the logarithm of the dissipation is approximately Gaussian). Sawford (1991) also proposes a Lagrangian model where it is the Lagrangian acceleration which follows a (Gaussian) SLM-like model. Note that considering an acceleration evolution equation is formally equivalent to the inclusion of memory terms in a velocity evolution equation.

Berloff and McWilliams (2002); Veneziani et al. (2004) also consider Stochastic Lagrangian models. The slightly non-linear resulting evolution models are defined on empirical grounds and Gaussian assumptions.



## 4.2 Eulerian Gaussian backscattering

The most famous Eulerian backscattering method is perhaps the Eddy-Damped Quasi Normal Markovian (EDQNM) model introduced by Orszag (1970) and Leith (1971). It closes the large-scale Navier-Stokes equations in the Fourier space by neglecting some phase correlations in the non-linear terms while keeping the energy constant. Chasnov (1991) uses this framework to set up a forced-dissipative Navier-Stokes model where the Eddy Viscosity is scale-dependant and the forcing is Gaussian, homogeneous and isotropic in space and uncorrelated in time. Replacing non-linear interactions by a damping term and a Gaussian forcing is now common practice. This is in particular used to setup evolution laws of subgrid scales where accuracy is of lower concern. The solutions are in this case Gaussian processes. Structural Stability Theory (S3T) is one example in that spirit (Farrell and Ioannou, 2014). The Quasi-Linear (QL) approximation separates the non-linear deterministic dynamics of the mean field and the linearized randomly-forced dynamics of fluctuations. Then, stability analyses are applied to the augmented state-vector of mean and covariance in order to reveal and characterize various turbulent phenomenon. Stochastic superparameterization (SSP) proposes a similar model (Grooms and Majda, 2014). The point approximation separates the large-scale and the small-scale dynamics. The small-scale evolution law is linearized and corrected by the introduction of noise and damping terms. To ensure energy conservation in the stationary regime, noise variances and damping coefficients are related. Then, the second order moments of the solution are known analytically and can feed the subgrid tensor expression of the mean large-scale evolution law. Without involving any theoretical closure, Berloff (2005) considers a Gaussian forcing as well. Yet its noise is inhomogeneous in space and correlated in time. Well specified inhomogeneity brings phase information making the model more accurate and the forcing more efficient.

## 4.3 Eulerian non-Gaussian backscattering

Phase information can also be encoded by multiplicative noises. Besides, multiplicative noises are the most common non-Gaussian forcing in CFD. Leith (1990) multiplies a white Gaussian noise by a function of the resolved local strain rate. Schumann (1995) uses a quadratic function of a homogeneous Gaussian noise. Brankart (2013) adds at each time step a multiplicative noise to the active tracers (salinity and temperature) before computing the corresponding density. After this, the density is averaged over realizations. The non-linearity of the state-equation makes this transitional variability non-negligible and improves the simulation results.

In contrast, Mana and Zanna (2014) use a non-Gaussian noise process. It is uncorrelated in time and space with a variance that depends on the resolved potential vorticity (PV) gradient.

Except Schumann (1995), all these methods defined on empirical grounds have a common characteristic: the factor of the multiplicative noise – or the noise variance – is a function of the gradient of the transported quantity. This comment is also valid for the stochastic Lagrangian models of S. B. Pope and coauthors. This suggests a link between stochastic backscattering and turbulent dissipation, but the justification of this link is generally either unclear or arbitrary.

## 5 UQ for errors induced by a coarsening process

### 5.1 Gaussian additive noise

The simplest random dynamics are defined from linear Langevin equations with additive Gaussian noise. This has already been discussed above for stochastic backscattering methods. We may add the linear inverse models (Peland and Matrosova, 1994; Penland and Sardeshmukh, 1995). Keating et al. (2012) also use a linear evolution model with Gaussian noise for a filtering purpose. However, the parameters of the models are themselves Ornstein-Uhlenbeck (OU) processes. An OU process is a gaussian process with an exponential covariance in the stationary regime. Such a process corresponds to an Auto-Regressive (AR) process in discrete time. In Keating et al. (2012), OU parameters make the evolution model solution not Gaussian and more flexible. This method – well known in the filtering community – is referred to as Stochastic Parameterized Extended Kalman Filter (SPEKF) (Gottwald and Harlim, 2013).

In geophysical fluid dynamics data assimilation, dynamics are often non-linear. Nevertheless, Gaussian additive random forcings are still widely used. Tandeo et al. (2018) review the additive-Gaussian-noise-covariance estimation procedures. Moment-based, maximum likelihood, and Bayesian estimations are detailed. Several authors also define correlation matrices as solutions of diffusion equations parameterized from observed correlation lengths (Pannekoucke and Massart, 2008). Others learn a spatial mapping which transforms the heterogeneous and anisotropic spatial covariance into a homogeneous and isotropic one using wavelets (Michel, 2013a,b) or differential geometry tools (Pannekoucke et al., 2014). Then, a classical covariance model can be fitted.

### 5.2 Empirical non-Gaussian noises

Among other empirical stochastic models, the Stochastic Perturbed Physics Tendency scheme (SPPT) introduces a correlated multiplicative noise (Buizza et al., 1999). All parameterized tendencies are multiplied by the same uniform variable in  $[0.5, 1.5]$ . The same random variable is used in a fixed space-time window. This window sets the noise spatio-temporal correlations. Shutts

(2005); Berner et al. (2009, 2011) have proposed another method constructed from a spatially homogeneous and isotropic Gaussian noise. Then, they multiply it by the dissipation rate. This method is called Stochastic Kinetic Energy Backscatter (SKEBS). As in Schumann (1995) and Brankart (2013), the noise is defined from an Ornstein-Uhlenbeck (OU) process with a very small correlation time. As for many stochastic subgrid parameterization, the factor of the multiplicative noise is again a function of the gradient of the transported quantity.

We may also cite a recent variant of SKEBS referred to as Stochastic Convective Backscatter (SCB) scheme (Shutts, 2015). This scheme is focused on atmospheric flow applications. A Gaussian noise white in space and time is weighted by the vertical variation of the parameterized convective updraught and downdraught mass fluxes. Then, a spatio-temporal smoothing imposes the spatial and temporal correlations.

SPPT and SKEBS methods have been successfully applied in operational weather and climate forecast centers (Franzke et al., 2015; Buizza, 2016b; Pegion et al., 2016; Separovic et al., 2016; McCabe et al., 2016; Buizza, 2016a). Nevertheless, many drawbacks of these methods have been reported. Above all, conservation laws (e.g. energy and mass conservations) are violated (Reynolds et al., 2016; Leutbechner et al., 2016). Precipitation biases have also been observed (Buizza, 2016a). Moreover, the random forcing scales differ from the scales of the error sources that are meant to be quantified (Reynolds et al., 2016).

### 5.3 Energy-budget-based noises

Energy conservation and redistribution being fundamental aspects of physics, several authors have developed ad hoc schemes to deal with conservation laws.

#### 5.3.1 Existing schemes

Sapsis and Majda (2013c) introduce the Modified Quasilinear Gaussian (MQG) model for – but not restricted to – dimensionally reduced systems (Sapsis and Majda, 2013b,a). This model approximates the third-order moment in the covariance evolution law in order to redistribute the right amounts of energy between modes. Based on stationary regime information, dampings and noises are specified in that way.

Earlier in this paper, we have mentioned the Stochastic superparameterization (SSP) (Grooms and Majda, 2014), the EDQNM approximation and other related methods. In these approaches, the random forcing is also specified from an imposed mean energy balance. Nevertheless, these methods are not *a priori* designed for an UQ purpose.

Similar energy mean budgets have recently been discussed by several authors. Majda (2015) refers to this energy mean as the statistical energy. The

author derives the evolution law of this energy by adding the evolution equations of the mean energy and the evolution equation of the integrated variance. However, Majda (2015) does not specify the random forcing. This is why the latter does not *a priori* balance the turbulent diffusion. Farrell and Ioannou (2014) also study the energy mean of stochastic fluid dynamics systems especially under quasi-linear approximations and with an additive Gaussian forcing. Gugole and Franzke (219) enforce the energy conservation by projecting their random forcing into the space of constant energy.

As cited previously, SKEBS parameterization (Shutts, 2005; Berner et al., 2009, 2011) backscatters a given portion of the global dissipated energy using a flow-dependent correlated noise. Jansen and Held (2014) work with a similar idea. Yet, they consider a hyper-viscous diffusion and a noise white in space and time. Dwivedi et al. (2019) also consider hyper-viscosity and neglect spatial correlations but they keep time correlations.

### 5.3.2 A new random forcing derived from large-scale subgrid dissipation models

We here propose a new energy-budget-based stochastic subgrid model.

In practice, large-scale fluid dynamics models involve dissipation operators called subgrid models. They stabilize the numerical simulations and mimic the action of the unresolved small scales by draining the energy at high wavenumbers. Simple subgrid models often take the form of classical Laplacian operator or of higher-order hyperviscosity operators (typically, some power of a Laplacian). Let us consider the dynamic of a fluid property  $q$ . It is assumed to be transported up to a dissipation operator denoted  $\mathcal{L}$  and a stochastic forcing  $\eta$  – which is a centred process uncorrelated in time:

$$\frac{Dq}{Dt} = \mathcal{L}[q] + \eta, \quad (1)$$

where  $D/Dt = \partial_t + \mathbf{v} \cdot \nabla$  stands for the material derivative operator.

Through Itô calculus, a conservation of the energy mean –  $\mathbb{E} \int_{\Omega} \frac{1}{2} q^2$  – would imply:

$$0 = \mathbb{E} \int_{\Omega} \frac{d}{dt} \left( \frac{1}{2} q^2 \right) = \mathbb{E} \int_{\Omega} \left( q \frac{d}{dt} q + \frac{1}{2} \frac{d}{dt} \langle q, q \rangle \right) = \mathbb{E} \int_{\Omega} \left( q \mathcal{L}[q] + \frac{1}{2} (\mathcal{H}[q])^2 dt \right), \quad (2)$$

where

$$(\mathcal{H}[q](\mathbf{x}, t))^2 \triangleq \mathbb{E}(\eta^2(\mathbf{x}, t) | q(\bullet, t)), \quad (3)$$

is the stochastic forcing variance conditioned on field  $q$  at time  $t$  and  $\langle q, q \rangle$  denotes the quadratic variation term. In order to maintain a desired amount of energy dissipation, we introduce a scaling factor  $\zeta > 0$  in the above balance. Specifically, we want to impose:

$$\int_{\Omega} q \left( -\frac{2\zeta}{dt} \mathcal{L}[q] \right) = \int_{\Omega} (\mathcal{H}[q])^2. \quad (4)$$

Now, for a given dissipation operator  $\mathcal{L}$ , we will construct a noise  $\eta$  which meets the desired balance (4). If  $\mathcal{L}$  is a negative auto-adjoint operator – which is generally the case – we can define an operator  $\mathbf{H}$  which is such that

$$-\frac{2\zeta}{dt}\mathcal{L} = \mathbf{H}\mathbf{H}^*, \quad (5)$$

with  $\mathbf{H}^*$  denoting the adjoint of operator  $\mathbf{H}$ . More precisely, for the Laplacian, bi-Laplacian or higher-order hyperviscosity operators, this operator  $\mathbf{H}$  reads,

$$\mathbf{H}[q] \triangleq \begin{cases} \tilde{\zeta}\boldsymbol{\alpha}\nabla q & \text{if } \mathcal{L}[q] = \nabla \cdot (\boldsymbol{\alpha}\boldsymbol{\alpha}^T\nabla q), \\ \tilde{\zeta}\nabla \cdot (\boldsymbol{\alpha}\nabla q) & \text{if } \mathcal{L}[q] = -\nabla \cdot (\boldsymbol{\alpha}\nabla(\nabla \cdot (\boldsymbol{\alpha}\nabla q))), \\ \tilde{\zeta}\boldsymbol{\alpha}\nabla\Delta^p q & \text{if } \mathcal{L}[q] = \alpha^2\Delta^{2p+1}q, \\ \tilde{\zeta}\boldsymbol{\alpha}\Delta^p q & \text{if } \mathcal{L}[q] = -\alpha^2\Delta^{2p}q, \end{cases} \quad (6)$$

where  $\tilde{\zeta} = \sqrt{\frac{2\zeta}{dt}}$ ,  $p$  is a positive integer,  $\alpha$  a constant and  $\boldsymbol{\alpha}$  a – possibly spatially-varying – matrix.

Then, for specific deterministic subgrid tensor  $\mathcal{L}$ , if the stochastic forcing (conditional) variance integral,  $\int_{\Omega}(\mathcal{H}[q])^2$ , can be set to the integral  $\int_{\Omega}\|\mathbf{H}[q]\|^2$ , it is possible then to control the energy dissipation and meet the desired balance (4). Indeed, in that case, we have:

$$\int_{\Omega}(\mathcal{H}[q])^2 = \int_{\Omega}\|\mathbf{H}[q]\|^2 = \int_{\Omega}q\mathbf{H}^*\mathbf{H}[q] = \int_{\Omega}q\left(-\frac{2\zeta}{dt}\mathcal{L}[q]\right). \quad (7)$$

Nevertheless, to simulate the stochastic forcing, knowing its global variance is not enough. We also need to choose its spatial structure (e.g. the local variation of the variance, its spatial correlations). In this purpose, we express the noise on a convenient orthonormal basis of  $L^2(\Omega)$  (e.g. a wavelet basis) denoted  $\{e_k\}_{k \geq 0}$ :

$$\eta(\mathbf{x}, t) = \sum_{k \geq 0} \lambda_k(t)\xi_k(t)e_k(\mathbf{x}), \quad (8)$$

where the  $\xi_k$  are independent Gaussian white noises of unit variance and:

$$\lambda_k = \int_{\Omega}\|\mathbf{H}[q]\| e_k. \quad (9)$$

Then, by definition of  $\mathcal{H}$  and by the Parseval theorem, the random forcing (conditional) variance integral is given by:

$$\int_{\Omega}(\mathcal{H}[q])^2 = \int_{\Omega}\mathbb{E}(\eta^2(\mathbf{x}, t)|q(\bullet, t)) = \sum_{k \geq 0}|\lambda_k|^2 = \int_{\Omega}\|\mathbf{H}[q]\|^2. \quad (10)$$

Therefore, equation (7) is valid and the stochastic forcing  $\eta$  defined by (8) and (9) meets our target energy balance (4).

Therefore, given a numerical dissipation  $\mathcal{L}$  together with a chosen orthonormal basis, it is always possible to define a noise (8) that respects the assumed

energy balance (4). Note however that this balance constitutes only a global balance. Indeed, locally in space, the random forcing variance is:

$$(\mathcal{H}[q])^2 = \sum_{k \geq 0} |\lambda_k|^2 |e_k|^2 \neq \left| \sum_{k \geq 0} \lambda_k e_k \right|^2 = \|\mathbf{H}[q]\|^2, \quad (11)$$

and moreover,

$$\|\mathbf{H}[q]\|^2 \neq q \mathbf{H}^* \mathbf{H}[q] = q \left( -\frac{2\zeta}{dt} \mathcal{L}[q] \right). \quad (12)$$

The choice of a given basis  $\{e_k\}_{k \geq 0}$  in this model is of crucial importance, as it influences in particular the variance heterogeneity and the correlation lengths of the random forcing. In order to be close to a local energy balance, basis functions with small supports can be selected. This enables us to restrain the inequality (11) and make it closer to an equality. For instance, Fourier modes have a large support and would lead to a homogeneous random forcing (since  $|e_k| = \text{cst}$ ). In this case, the inequality (11) would be  $(\mathcal{H}[q])^2 = \frac{1}{\mu(\Omega)} \int_{\Omega} \|\mathbf{H}[q]\|^2 \neq \|\mathbf{H}[q]\|^2(\mathbf{x}, t)$ , where  $\mu(\Omega)$  is the measure of the spatial domain  $\Omega$ . In contrast a basis of regularized Dirac functions  $e_k(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_k)$  – defined on the grid points  $\mathbf{x}_k$  – have near zero-measure supports. The inequality (11) would hence become an equality. Nevertheless, such an infinitesimally small basis function support would induce an almost zero random forcing correlation length. We consider this behavior as nonphysical. According to that analysis, a wavelet basis seems a promising trade-off and will be used in the numerical section 8.

#### 5.4 Physically-based noises

We now come back to the literature review. In fluid dynamics, due to nonlinearities, the targeted probability density of the solution is highly non-Gaussian and relevant stochastic dynamical models are difficult to derive. In this perspective, an attractive path would be to infer randomness from physics (Berner et al., 2015; Craig et al., 2016). Yet as Navier-Stokes equations are deterministic, this path is not straightforward.

##### 5.4.1 Time-scale separation

###### *Averaging and homogenization*

To derive large-scale fluid dynamical model, the time-scale separation assumption is convenient. In the seventies, Hasselmann (1976) already relied on it for geophysical fluid dynamics. In his seminal work, the large-scale dynamics were encoded by both mean terms and noise terms. However, eventually only simple multidimensional OU processes were considered.

The time-scale separation assumption is also the foundation of the more rigorous averaging and homogenization theories (Kurtz, 1973; Papanicolaou and

Kohler, 1974; Givon et al., 2004; Mitchell and Gottwald, 2012; Gottwald and Melbourne, 2013; Gottwald and Harlim, 2013; Franzke et al., 2015; Gottwald et al., 2015). As the time-scale separation goes to infinity, the large-scale dynamics will converge according to averaging or to homogenization depending on the structure of the global model. The global dynamics as well as the limit large-scale dynamics can be differential equations or SDEs. In the large-scale equation, terms which are only functions of the small-scale variable often tend to converge to a white-noise-in-time term in the Stratonovich sense with a covariance of the Green-Kubo type. Nevertheless, it is not always true for nonlinear dynamics. Sometimes, the noise has to be understood in the sense of Ito or Marcus stochastic integral. In the last case, the noise is a Levy process (Gottwald and Melbourne, 2013; Gottwald et al., 2015).

A successful application of the homogenization theory in geophysics is the MTV scheme (Majda et al., 1999, 2001; Franzke et al., 2005; Majda et al., 2008). MTV refers to the names of the three main authors: Majda, Timofeyev and Vanden-Eijnden. In practice, the non-linearity of the small-scale equation is empirically replaced by a noise term and a damping term before the homogenization procedure. The homogenized dynamics obtained are cubic with correlated additive and multiplicative (CAM) noises. Even without dealing with Levy processes, this structure is able to produce intermittency and extreme events especially because of the CAM noise. This specific form has also been used to infer data-driven models. Peavoy et al. (2015) proposed an example of such a model which uses energy constrained Bayesian estimators and artificial additional observations through Brownian bridge.

Another method called invariant manifold theory also invokes a time-scale separation. Yet it relies directly on the SDE solution rather than on its probability density function (Givon et al., 2004; Gottwald and Harlim, 2013). This scheme provided good UQ skills especially because of the multiplicative noise appearing in the limit dynamics. Chekroun et al. (2017) describes a similar problem for geophysical fluid dynamics on the so-called parameterized manifolds with small Rossby number.

To conclude on these methods, for complicated non-linear dynamical systems, it is still not clear how to perform homogenization and when this is possible. Moreover, the theory does not make the noise covariance explicit enough and it has to be estimated on data. During this step, some Gaussian approximations are usually done when estimating the coefficients of the model. Some homogenization methods like the MTV algorithm may suffer from energy-conservation issues. Nevertheless, workarounds exist (Frank and Gottwald, 2013; Jain et al., 2014). In addition, the homogenization methods have shown successful results in the context of reduced order models and suggest that geophysical stochastic fluid dynamic models should involve CAM noises.

### *Edgeworth expansion*

In order to alleviate the time scale gap assumption of averaging and homoge-

nization procedures, Edgeworth expansions goes to higher order in time scale ratio (Wouters and Gottwald, 2018).

#### *Velocity time scale separation and skew-symmetric noises*

The Kraichnan model is an idealization of passive tracer turbulence (Kraichnan, 1968, 1994; Gawędzki and Kupiainen, 1995). A tracer is forced and advected by two independent spatially-homogeneous time-uncorrelated Gaussian processes. In contrast to the previous section, the time scale gap between the unresolved and the resolved dynamics is directly assumed, without going to a limit. The forcing of the Kraichnan model lives at large spatial scales whereas the energy of the random advecting velocity is distributed over the spatial scales in a self-similar way. Since this velocity is also assumed divergence-free, the advecting term is here a skew-symmetric multiplicative noise of the tracer dynamics. It is a well-studied and well-known model in statistical physics (Gawędzki and Kupiainen, 1995; Klyatskin et al., 1996; Majda et al., 1999; Falkovich et al., 2001). In particular, it is known that the tracer gradient norm has a log-normal point-wise law. Some authors have also added a deterministic advecting velocity to the delta-correlated random term (e.g. Klyatskin et al., 1996).

More recently, several authors have considered the random transport of fluid dynamics quantities by the combination of two velocity components: a time-correlated one and a time-uncorrelated one. Unlike the Kraichnan model, these stochastic subgrid parameterizations are not restricted to the linear dynamics of passive tracers and can consider large-scale time-correlated flows. In the same time, these new schemes still encompass skew-symmetric multiplicative noises. The mathematical theory relies on stochastic calculus and Ito-Wentzell formula (Kunita, 1997). Brzeźniak et al. (1991) first introduces the idea. Then, Mikulevicius and Rozovskii (2004) and Flandoli (2011) modify the formula. Their works have focused on pure mathematical aims: existence and uniqueness of SPDE solutions. Neves and Olivera (2015) also studied the well-posedness of similar SPDE using the derivation of the previous authors. Models under location uncertainty (LU) (Mémin, 2014; Resseguier et al., 2017a,b; Resseguier, 2017; Chapron et al., 2018; Cai et al., 2018; Resseguier et al., 2019) and Stochastic Advection by Lie Transport (SALT) (Holm, 2015; Crisan et al., 2017; Gay-Balmaz and Holm, 2018; Cotter et al., 2018c,a,b; Resseguier et al., 2019) are new types of stochastic subgrid tensors based on that stochastic transport. These schemes will be deeply detailed in section 6. Sardeshmukh and Sura (2009); Sardeshmukh et al. (2015) also highlight the relevance of skew-symmetric multiplicative noises and its link with an advecting velocity time scale separation assumption. Besides, this separation assumption can be rigorously derived from homogenization technique (Cotter et al., 2017). With similar assumptions but different mathematical tools (Taylor series instead of stochastic calculus) Dukowicz and Smith (1997) obtain results for the mean tracers which coincide with LU framework. Nevertheless, equations for the mean is not sufficient for UQ purposes.



### 5.4.2 Memory effects

When there is no time-scale separation, the large-scale system can become non-Markovian. For deterministic dynamics, this is readily shown by the Mori-Zwanzig equation (Givon et al., 2004; Gottwald et al., 2015). Indeed, this explicit expression of the large-scale observables of interest involves a memory term. Accordingly, some authors have proposed non-Markovian stochastic parameterizations (e.g. Chekroun et al., 2011; Kondrashov et al., 2015; Kondrashov and Berloff, 2015; Lu et al., 2017). Even though these memory effects are realistic, it is important to note that their simulations imply an increase of the state-space dimension and hence an additional computational cost.

### 5.4.3 Other approaches based on statistical physics

Using Ruelle response theory, Wouters and Lucarini (2012) have proposed a systematic way to approximate non-linear dynamical systems. Their main assumption is the weak coupling between two terms of the original evolution equation. In this approximated dynamics, a deterministic first-order correction appears. At second order, there is also a memory term and a stochastic forcing.

In quasi direct interaction closure (DIA), Frederiksen (1999) proposed to simplify the equations for the mean flow and the fluctuations by assuming that the fluctuation two-time covariance function and the Green function of the diffusion equation (associated with eddy viscosity) are quasi-diagonal in Fourier space. We recall that a diagonal covariance in Fourier space means homogeneity in spatial space. One limitation of the method is the restriction to Fourier-based numerical simulations.

Plant and Craig (2008) have proposed a physically-based stochastic subgrid parameterization to model mass flux induced by clouds updrafts and downdrafts. Nevertheless, its application is *a priori* limited to weather forecasts.

## 6 Dynamics under Location Uncertainty and Stochastic Advection by Lie Transport

In this section, we focus on a new family of stochastic subgrid parameterizations. These schemes mainly rely on a stochastic transport. Therefore, in this section, we first explain the principle of this random transport and review its main properties. Most of these theoretical results come from Mémin (2014) and Resseguier et al. (2017b). The Stochastic Advection by Lie Transport can be introduced alternatively through stochastic differential geometry (e.g. Holm, 2015). Then, we present the randomized Euler equations of this model family. Finally, we review the existing parameterization choices in this stochastic framework and propose a new one.

## 6.1 Stochastic transport

### 6.1.1 Informal description

CFD introduces *de facto* a coarse scale truncation of the system. Those dynamical models emanate from physical deterministic representations whose solutions are assumed to be smooth (*i.e.* differentiable) in time. Although small-scale fluid flow velocities can be characterized by local and intermittent energy bursts, possibly associated with infinitesimal characteristic time-scales, it is generally assumed that these unresolved flow components remain smooth in time. At the model resolution, the resolved (large-scale) flow can thus be considered as a coarse-grained representation of the actual Eulerian flow, with the unresolved flow component rapidly varying in time. From an observer point of view, such sub-grid dynamics can be conveniently modeled by a delta-correlated process. The smooth velocity field, denoted  $\mathbf{w}$ , represents a large-scale, possibly random, component continuous in time. The unresolved contribution, expressed as  $\sigma \dot{\mathbf{B}}$ , is then assumed Gaussian, volume preserving (divergence-free) and uncorrelated in time. This contribution can be non-homogeneous and anisotropic in space. Due to the irregularity of the resulting flow, the transport of a conserved quantity,  $q$ , by the whole velocity, defined as

$$q(\mathbf{X}_{t+\Delta t}, t + \Delta t) = q(\mathbf{X}_t, t) \quad (13)$$

reads in an informal stochastic way as

$$\partial_t q + \underbrace{\mathbf{w}^* \cdot \nabla q}_{\text{Corrected advection}} = \underbrace{\nabla \cdot \left( \frac{1}{2} \mathbf{a} \nabla q \right)}_{\text{Diffusion}} - \underbrace{\sigma \dot{\mathbf{B}} \cdot \nabla q}_{\text{Random forcing}}, \quad (14)$$

with a drift velocity corrected as

$$\mathbf{w}^* = \mathbf{w} - \frac{1}{2}(\nabla \cdot \mathbf{a})^T + \sigma(\nabla \cdot \sigma)^T. \quad (15)$$

Hence, the deterministic (for a fixed realization of velocity field  $\mathbf{w}$ ) evolution equation is replaced by a stochastic equation with respect to  $\sigma \dot{\mathbf{B}}$ . As a result, the conserved quantity is now advected by an “effective” velocity,  $\mathbf{w}^*$ , taking into account the possible spatial variation of the small-scale velocity variance and the possible small-scale velocity divergence. This modified advection is indeed essential to take into account essential physical effects due the small-scale velocity inhomogeneity (Chandramouli et al., 2019; Pinier et al., 2019). The random forcing term in (14) relates to the advection by the unresolved velocity  $\sigma \dot{\mathbf{B}} = \sigma \frac{d\mathbf{B}_t}{dt}$ . This term continuously backscatters random energy to the system. The diffusion term then accounts for the mixing effect of the small-scale random velocity. This term plays a role similar to the eddy diffusivity models introduced in classical large-scale representations (Gent and McWilliams, 1990; Germano et al., 1991; Lilly, 1966; Smagorinsky, 1963), in analogy with

the molecular diffusion mechanism (Boussinesq, 1877). In particular, for a homogeneous and isotropic small-scale velocity the turbulent diffusion is also homogeneous and isotropic:

$$\mathbf{a} = a_0 \mathbb{I}_d \text{ and } \nabla \cdot \left( \frac{1}{2} \mathbf{a} \nabla q \right) = \frac{1}{2} a_0 \Delta q. \quad (16)$$

The inhomogeneous and anisotropic diffusion coefficient matrix,  $\mathbf{a}$ , is then defined by the one-point one-time covariance of the unresolved displacement per unit of time:

$$\mathbf{a} = \frac{\mathbb{E} \{ \boldsymbol{\sigma} d\mathbf{B}_t (\boldsymbol{\sigma} d\mathbf{B}_t)^T \}}{dt}. \quad (17)$$

This ensures an exact energy balance between the amount of diffusion and the random forcing.

### 6.1.2 Stochastic flow

To derive more formally the evolution law of a scalar quantity transported by a stochastic flow, the stochastic Lagrangian description of the infinitesimal displacement associated with a particle trajectory  $\mathbf{X}_t$  writes:

$$d\mathbf{X}_t = \mathbf{w}(\mathbf{X}_t, t) dt + \boldsymbol{\sigma}(\mathbf{X}_t, t) d\mathbf{B}_t. \quad (18)$$

In this equation, the second term explicitly figures the flow location uncertainty. Formally, this random field is defined over the fluid domain,  $\Omega \subset \mathbb{R}^d$ , from a  $d$ -dimensional Brownian function  $\mathbf{B}_t$ . Such a function can be interpreted as a white noise process in space and a Brownian process in time. Formally it is a cylindrical  $I_d$ -Wiener process (see Da Prato and Zabczyk, 1992; Prévôt and Röckner, 2007, for more information on infinite dimensional Wiener process and cylindrical  $I_d$ -Wiener process). The time derivative of the Brownian function, in a distribution sense, is informally denoted  $\boldsymbol{\sigma} \dot{\mathbf{B}} = \boldsymbol{\sigma} \frac{d\mathbf{B}_t}{dt}$ , and is a white noise distribution. The spatial correlations of the flow uncertainty are specified through the diffusion operator  $\boldsymbol{\sigma}(\cdot, t)$ , defined for any vectorial function,  $\mathbf{f}$ , through the matrix kernel  $\check{\boldsymbol{\sigma}}(\cdot, \cdot, t)$ :

$$\boldsymbol{\sigma}(\mathbf{x}, t) \mathbf{f} \triangleq \int_{\Omega} \check{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{z}, t) \mathbf{f}(\mathbf{z}, t) d\mathbf{z}. \quad (19)$$

This quantity is assumed to have a finite norm. More precisely, the operator  $\boldsymbol{\sigma}$  is assumed to be Hilbert-Schmidt. We also assume that the above expression have periodic or null boundary conditions on the domain frontier. The resulting  $d$ -dimensional random field,  $\boldsymbol{\sigma}(\mathbf{x}, t) d\mathbf{B}_t$ , is a centered vectorial Gaussian function, correlated in space and uncorrelated in time with covariance tensor:

$$\text{Cov}(\mathbf{x}, \mathbf{y}, t, t') \triangleq \mathbb{E} \{ (\boldsymbol{\sigma}(\mathbf{x}, t) d\mathbf{B}_t) (\boldsymbol{\sigma}(\mathbf{y}, t') d\mathbf{B}_{t'})^T \}, \quad (20)$$

$$= \int_{\Omega} \check{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{z}, t) \check{\boldsymbol{\sigma}}^T(\mathbf{y}, \mathbf{z}, t) d\mathbf{z} \delta(t - t') dt. \quad (21)$$

Hereafter, the diagonal of the covariance tensor,  $\mathbf{a}$ , – also denoted as  $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$  – will be referred to as the variance tensor:

$$\mathbf{a}(\mathbf{x}, t)\delta(t - t')dt = \mathbf{Cov}(\mathbf{x}, \mathbf{x}, t, t').$$

Cotter et al. (2017) have rigorously shown that the decomposition (18) corresponds to the limit of a deterministic flow when the correlation time of the small-scale velocity goes to zero.

### 6.1.3 Scalar advection

For a fluid flow defined by equation (18), the material derivative (expressed in Ito form) of a quantity  $q$  writes:

$$\begin{aligned} D_t q &\triangleq (q(\mathbf{X}_{t+dt}, t+dt) - q(\mathbf{X}_t, t))|_{\mathbf{x}_t=\mathbf{x}} = & (22) \\ &\underbrace{d_t q}_{\substack{\triangleq q(\mathbf{x}, t+dt) - q(\mathbf{x}, t) \\ \text{Time increment}}} + \underbrace{(\mathbf{w}^* dt + \boldsymbol{\sigma} d\mathbf{B}_t) \cdot \nabla q}_{\text{Advection}} - \underbrace{\nabla \cdot \left(\frac{1}{2}\mathbf{a}\nabla q\right) dt}_{\text{Turbulent diffusion}} + \underbrace{\text{tr}((\boldsymbol{\sigma}^T \nabla) \mathbf{H}^T) dt}_{\text{Coupling turbulence-forcing}}, \end{aligned}$$

where the quantity  $q$  is forced as follows

$$D_t q = F dt + \mathbf{H}^T d\mathbf{B}_t. \quad (23)$$

The time increment  $d_t q$  can be interpreted as the analog of the partial time derivative  $\partial_t q$  in deterministic partial differential equations. The above expression – derived by Resseguier et al. (2017a) – is a reformulation of the Ito-Wentzell formula (theorem 3.3.1 page 91 Kunita, 1997; Chow, 2014).

#### Stratonovich notations:

Equations (22) and (23) rely on the so-called Ito notations. They are convenient to derive ensemble mean – since terms in  $d\mathbf{B}_t$  are always centered – and for the numerical implementations. From these notations, we also get more insights on the involved physical processes (e.g. the turbulent diffusion). Nevertheless, it is also possible to write the very same equations in Stratonovich notations (i.e. with " $\circ d\mathbf{B}_t$ " instead of " $d\mathbf{B}_t$ "):

$$\begin{aligned} D_t^S q &= \underbrace{d_t q}_{\substack{= q(\mathbf{x}, t+dt/2) \\ - q(\mathbf{x}, t-dt/2) \\ \text{Centered} \\ \text{time increment}}} + (\mathbf{w}_S dt + \boldsymbol{\sigma} \circ d\mathbf{B}_t) \cdot \nabla q, & (24) \end{aligned}$$

where

$$D_t^S q \triangleq \left( q\left(\mathbf{X}_{t+\frac{dt}{2}}, t+\frac{dt}{2}\right) - q\left(\mathbf{X}_{t-\frac{dt}{2}}, t-\frac{dt}{2}\right) \right) \Big|_{\mathbf{x}_t=\mathbf{x}}, \quad (25)$$

$$= D_t q - \frac{1}{2} \text{tr}((\boldsymbol{\sigma}^T \nabla) \mathbf{H}^T) dt, \quad (26)$$

$$= \left( F - \frac{1}{2} \frac{d}{dt} \langle \mathbf{H}^T, \mathbf{B}_t \rangle - \frac{1}{2} \text{tr}((\boldsymbol{\sigma}^T \nabla) \mathbf{H}^T) \right) dt + \mathbf{H}^T \circ d\mathbf{B}_t, \quad (27)$$

is the Stratonovich material derivative and

$$\mathbf{w}_S = \mathbf{w} - \frac{1}{2} \sum_{i=1}^d \partial_i \boldsymbol{\sigma} \boldsymbol{\sigma}_{i\bullet}^T = \mathbf{w}^* - \frac{1}{2} \boldsymbol{\sigma} (\boldsymbol{\nabla} \cdot \boldsymbol{\sigma})^T, \quad (28)$$

is the Stratonovich drift of the flow (18). The Stratonovich material derivative,  $D_t^S q$ , coincides with the Ito material derivative,  $D_t q$ , for  $\mathbf{H} = 0$ , i.e. for time correlated forcing in the transport equation (23) – which is generally the case. The Stratonovich drift,  $\mathbf{w}_S$ , coincides with the corrected drift,  $\mathbf{w}^*$ , for divergence-free unresolved velocity –, which is the most common case. Equation (24) is the Ito-Wentzell formula for Stratonovich notations (Kunita, 1997; Chow, 2014). The reader not familiar with stochastic partial differential equations can also refer to section 10.1.2 of Resseguier (2017) for more details. Equations (27) and (26) are derived in Appendix B by application of the Stratonovich-Ito-notation-change formula (Kunita, 1997, theorem 3.2.5 page 60).

#### 6.1.4 Energy

##### *Energy conservation:*

To ensure a stochastic isochoric flow, incompressibility constraints on the modified drift,  $\boldsymbol{\nabla} \cdot \mathbf{w}^* = 0$  and on the small-scale velocity,  $\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = 0$ , are required. One can show (Resseguier et al., 2017a) that those two constraints enable to establish a strong energy conservation property for any realizations of a tracer (i.e. with zero forcing in (23)):

$$\frac{d}{dt} \int_{\Omega} \frac{1}{2} q^2 = 0. \quad (29)$$

The noise energy intake is exactly compensated by the diffusion term. In particular, the mean energy,  $\mathbb{E} \int_{\Omega} \frac{1}{2} q^2$ , is also conserved. So, the LU-SALT also meets the balance (4) of section 5.3.2 (with  $\zeta = 1$ ,  $\mathcal{L}[q] = \boldsymbol{\nabla} \cdot (\frac{1}{2} \mathbf{a} \boldsymbol{\nabla} q)$  and  $(\mathcal{H}[q])^2 = \|\mathcal{H}[q]\|^2 = \frac{1}{dt} \boldsymbol{\nabla} q^T \mathbf{a} \boldsymbol{\nabla} q$ ) even though the SALT-LU random forcing is not of the form (8) in general.

The energy conservation (29) also implies that the decrease of the energy of the mean always goes with a variance increase:

$$\frac{d}{dt} \int_{\Omega} \frac{1}{2} \text{Var}(q) = - \frac{d}{dt} \int_{\Omega} \frac{1}{2} (\mathbb{E}(q))^2, \quad (30)$$

This process is useful for UQ. While the tracer interacts with the unresolved scales, the tracer is continuously randomized.

##### *Local energy in spatial space:*

Since  $D_t$  is a derivative (considering Lagrangian coordinates), it is straightforward that:

$$D_t \left( \frac{1}{2} q^2 \right) = q D_t q = 0. \quad (31)$$

Hence, all the discussions about the tracer  $q$  are also valid for the local energy  $\frac{1}{2} q^2$ .

*Local energy in Fourier space (spectrum):*

Here,  $\hat{q}(\mathbf{k}, t) \triangleq \int_{\Omega} q(\mathbf{x}, t) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}$  stands for the Fourier transform of  $q$ . In order to get some insights on where the uncertainty goes in Fourier space, we can derive the time evolution of the tracer spectrum  $\Gamma(\mathbf{k}, t) = \mathbb{E}|\hat{q}(\mathbf{k}, t)|^2$ . Indeed, if the small-scale velocity is homogeneous and isotropic in space, we can show (see appendix A) that:

$$\partial_t \Gamma(\mathbf{k}, t) = \underbrace{\frac{2}{(2\pi)^d} \mathbb{E} \Re \left\{ \widehat{w \cdot \nabla q}(\mathbf{k}, t) \overline{\hat{q}(\mathbf{k}, t)} \right\}}_{\text{Usual spectrum variation}} + \underbrace{\int_{\mathbb{R}^d} F_{\sigma}(\mathbf{k}, \mathbf{k}', t) d\mathbf{k}'}_{\text{Noise-induced spectrum flux}}, \quad (32)$$

where  $\Re\{z\}$  is the real part of  $z$  and the spectrum flux – induced by the small-scale velocity – from the wave-vector  $\mathbf{k} - \mathbf{k}'$  toward the wave-vector  $\mathbf{k}$  is:

$$F_{\sigma}(\mathbf{k}, \mathbf{k}', t) = \frac{\|\mathbf{k}\|^2}{(2\pi)^d} \Gamma_{\sigma}(\mathbf{k}', t) \left( \sin^2(\widehat{(\mathbf{k}, \mathbf{k}')})) \Gamma(\mathbf{k} - \mathbf{k}', t) - \frac{1}{d} \Gamma(\mathbf{k}, t) \right), \quad (33)$$

where

$$\Gamma_{\sigma}(\mathbf{k}', t) = \frac{1}{dt} \mathbb{E} \|\widehat{\sigma d\mathbf{B}_t}(\mathbf{k}', t)\|^2. \quad (34)$$

The first term of the budget (32) encompasses the usual large-scale non-linear effects. The second term is the tracer spectrum increase which is due to the small-scale velocity. This energy flux does not depend on the tracer phases, but only on the tracer spectrum. Furthermore, the tracer mode  $\hat{q}(\mathbf{k} - \mathbf{k}', t)$  gives energy to  $\hat{q}(\mathbf{k}, t)$ , by the intermediate of  $\sigma d\mathbf{B}_t$  if and only if:

$$\Gamma_{\sigma}(\mathbf{k}', t) \neq 0 \text{ and } \frac{|\hat{q}(\mathbf{k}, t)|}{|\hat{q}(\mathbf{k} - \mathbf{k}', t)|} < \sqrt{d} \left| \sin(\widehat{(\mathbf{k}, \mathbf{k}')})) \right|. \quad (35)$$

In this case, the energy transfer from  $|\hat{q}(\mathbf{k} - \mathbf{k}', t)|^2$  toward  $|\hat{q}(\mathbf{k}, t)|^2$  is mostly a random process, since it is due to the positive difference between the noise energy intake – which is a random process – and the turbulent dissipation. Moreover, if we assume that the (bidirectional or tridirectional) spectrum is decreasing with the wave number  $\|\mathbf{k}\|$  and that the dimension  $d = 2$ , a backscattering of energy from the small-scale Fourier mode  $\hat{q}(\mathbf{k} - \mathbf{k}', t)$  to the large-scale Fourier mode  $|\hat{q}(\mathbf{k}, t)| > |\hat{q}(\mathbf{k} - \mathbf{k}', t)|$  implies:

$$\left| \cos \left( \frac{\pi}{2} - \widehat{(\mathbf{k}, \mathbf{k}')} \right) \right| > \frac{1}{\sqrt{2}} = \cos \left( \frac{\pi}{4} \right) \quad (36)$$

$$\text{i.e. } \left( \widehat{(\mathbf{k}, \mathbf{k}')} \right) \text{ modulo } \pi = - \left( \widehat{(\mathbf{k}, \mathbf{k} - \mathbf{k}')} \right) \text{ modulo } \pi \in \left[ \frac{\pi}{4}, \frac{3\pi}{4} \right]. \quad (37)$$

Note that this result holds even if the statistics of the small-scale velocity depends on the tracer  $q$ , as long as the small-scale velocity is homogeneous and isotropic.

*Moments of passive tracers:*

Consider now that the expectation corresponds to a conditional expectation given the effective drift. This applies to passive scalar transport for which the drift does not depend on the tracer. In equation (22), terms in  $d\mathbf{B}_t$  have zero-mean, and the mean passive scalar evolution can be immediately derived taking the conditional expectation of the stochastic transport:

$$\partial_t \mathbb{E}(q) + \underbrace{\mathbf{w}^* \cdot \nabla \mathbb{E}(q)}_{\text{Advection}} = \underbrace{\nabla \cdot \left( \frac{1}{2} \mathbf{a} \nabla \mathbb{E}(q) \right)}_{\text{Diffusion}}. \quad (38)$$

Since  $\mathbf{w}^*$  is divergent-free, it has no influence on the energy budget. The mean field energy decreases with time due to diffusion. As for the variance, its evolution equation, derived in Resseguier et al. (2017a), reads:

$$\partial_t \text{Var}(q) + \underbrace{\mathbf{w}^* \cdot \nabla \text{Var}(q)}_{\text{Advection}} = \underbrace{\nabla \cdot \left( \frac{1}{2} \mathbf{a} \nabla \text{Var}(q) \right)}_{\text{Diffusion}} + \underbrace{(\nabla \mathbb{E}(q))^T \mathbf{a} \nabla \mathbb{E}(q)}_{\text{Variance intake}}. \quad (39)$$

This is also an advection-diffusion equation, with an additional source term. Integrating this equation on the whole domain, with the divergent-free condition, and considering the divergence form of the first right-hand term, we obtain

$$\frac{d}{dt} \int_{\Omega} \text{Var}(q) = \int_{\Omega} (\nabla \mathbb{E}(q))^T \mathbf{a} \nabla \mathbb{E}(q) \geq 0. \quad (40)$$

It shows that the stochastic transport of a passive scalar creates variance. The dissipation that occurs in the mean-field energy equation is exactly compensated by a variance increase. This mechanism is very relevant for ensemble-based simulations. The uncertainty modeling directly incorporates a large-scale dissipating sub-grid tensor, and further encompasses a variance increase mechanism to balance the total energy dissipation. Such a mechanism is absent in ensemble-based data assimilation development (Berner et al., 2011; Gottwald and Harlim, 2013; Snyder et al., 2015).

### 6.1.5 Extensive properties : Reynolds-transport theorem

Similar to the deterministic case, the stochastic Reynolds transport theorem shall describe the time differential of a scalar function,  $q(\mathbf{x}, t)$ , integrated over a material volume,  $\mathcal{V}(t)$ , transported by the random flow (18):

$$d \int_{\mathcal{V}(t)} q = \int_{\mathcal{V}(t)} \left[ D_t q + \nabla \cdot (\mathbf{w}^* dt + \boldsymbol{\sigma} d\mathbf{B}_t) q + d \left\langle \int_0^t D_{t'} q, \int_0^t \nabla \cdot \boldsymbol{\sigma} d\mathbf{B}_{t'} \right\rangle \right]. \quad (41)$$

This expression, rigorously derived in Resseguier et al. (2017a), was first introduced in a slightly different version by Mémin (2014). In most cases, the unresolved velocity component,  $\boldsymbol{\sigma} \dot{\mathbf{B}}$ , is divergence-free and, the source of variations of the extensive property  $\int_{\mathcal{V}(t)} q$  is time-differentiable, *i.e.* with a differential of the form  $d \int_{\mathcal{V}(t)} q = \mathcal{F} dt$ . In such a case, for an arbitrary volume, the transport theorem takes the form  $D_t q = f dt$ , and the material derivative can be replaced by the equation (22) to provide an intrinsic expression of this stochastic transport theorem.

## 6.2 Euler models

We will now express the incompressible Euler models existing in this family of random schemes. The Euler model under Location Uncertainty (LU) differs from Stochastic Advection by Lie transport (SALT) Euler model. For sake of simplicity and in contrast to the rest of the paper, we present them only in their Stratonovich forms.

Additionally, in this review, we will not consider the compressible case, since it brings other complexities. Accordingly, the density  $\rho$  is assumed to be constant. Using the mass conservation, this implies the following divergence-free constraints (Mémin, 2014; Resseguier et al., 2017a):

$$\nabla \cdot \mathbf{w}^* = 0, \quad (42)$$

$$\nabla \cdot \boldsymbol{\sigma} = 0. \quad (43)$$

After presenting SALT and LU Euler models, we will present a deterministic LES-like model referred to as pseudo-stochastic model. It has also been referred to as Euler model under Location Uncertainty (Mémin, 2014; Resseguier et al., 2015; Harouna and Mémin, 2017; Resseguier et al., 2017a,c; Chandramouli et al., 2018; Cintolesi and Mémin, 2019a,b). Therefore, we briefly explain it here for disambiguation.

### 6.2.1 Euler LU

The Euler LU assumes stochastic transport of the Ito drift  $\mathbf{w}$  up to some forcings:

$$D_t \mathbf{w} = D_t^S \underbrace{\mathbf{w}}_{\substack{\text{Due to the} \\ \text{transport} \\ \text{of } \rho \mathbf{w}}} = \left( \mathbf{g} - \frac{1}{\rho} \nabla p \right) dt. \quad (44)$$

The forcings can also encompass other terms (e.g. viscous terms, time-uncorrelated pressure forcing). By the properties of the stochastic transport, the Navier-Stokes LU conserves the (Ito) kinetic energy,  $\int_{\Omega} \frac{1}{2} \|\mathbf{w}\|^2$  (chapter 8 of Resseguier, 2017).

### 6.2.2 SALT Euler

In contrast, the SALT Euler is derived from differential geometry arguments with a Kelvin circulation preservation constraint and reads:

$$D_t^S \underbrace{\mathbf{w}^*}_{\substack{\text{Due to the} \\ \text{transport} \\ \text{of } \rho \mathbf{w}^*}} + \underbrace{\sum_{j=1}^d \nabla(\boldsymbol{\sigma} \circ d\mathbf{B}_t)_j w_j^*}_{\text{Additional term}} = \left( \mathbf{g} - \frac{1}{\rho} \nabla p \right) dt. \quad (45)$$

The reader may refer to (Holm, 2015) for a full presentation of the derivation. The SALT Euler does not conserve kinetic energy (chapter 10 of Resseguier,



2017) but conserves the (Stratonovich) helicity,  $H = \int_{\Omega} \mathbf{w}^* \cdot \nabla \times \mathbf{w}^*$  (Holm, 2015; Resseguier et al., 2019). The advantage of this model is that it leads to a classical form of the vorticity equation. However, this at the price of an additional term in the momentum equation and a loss of energy conservation. Crisan et al. (2017) have proofed the local well-posedness of the 3D incompressible SALT Euler model. In this review, we will not consider the compressible case, since it brings other complexities.

### 6.2.3 Pseudo-stochastic Euler LU

The pseudo-stochastic Euler model under location uncertainty writes:

$$\partial_t \mathbf{w} + (\mathbf{w}^* \cdot \nabla) \mathbf{w} - \nabla \cdot \left( \frac{1}{2} \mathbf{a} \nabla \mathbf{w}^T \right)^T = \mathbf{g} - \frac{1}{\rho} \nabla p. \quad (46)$$

Compared to the (stochastic) Euler model under location uncertainty (44), the pseudo-stochastic Euler model under location uncertainty (46) does not involve the stochastic forcing  $(\sigma d\mathbf{B}_t \cdot \nabla) \mathbf{w}$ . This is because it is assumed that the drift  $\mathbf{w}$  is a smooth function of time (more precisely a finite variation process) or, equivalently, that an uncorrelated pressure forcing compensates the stochastic forcing  $-(\sigma d\mathbf{B}_t \cdot \nabla) \mathbf{w}$ , which ensues from the unique decomposition in terms of martingales and finite variation terms. Accordingly, this pseudo-stochastic model is deterministic and hence easier to simulate for LES-like applications. The pseudo-stochastic model still retains the action of the unresolved small-scale components in terms of dissipation and modified advection, but does not include the backscattering effect brought by the random component. As a consequence, for UQ purposes, the (stochastic) Euler model under location uncertainty (44) may be more suitable (Resseguier et al., 2017a). Note that this pseudo-stochastic model has been successively used to define efficient reduced order models (Resseguier et al., 2017c), in which the subgrid dissipation is directly defined from the neglected modes. This representation has the advantage to provide also new diagnosis enabling to quantify local energy dissipation as well as the effect of the turbulence inhomogeneity on the large-scale flow (Pinier et al., 2019; Resseguier et al., 2017c). This ability is a strong asset of the LU formalism for reduced order modeling.

## 6.3 Parameterization of LU-SALT models

LU-SALT models are parameterized by the unresolved velocity spatial covariance (20). It can be an anisotropic, heterogeneous, time-dependant or even be a function of the large-scale quantities (e.g.  $q$ ). Choosing this covariance is equivalent to choosing the linear operator  $\sigma$  or choosing its kernel  $\check{\sigma}$ . In this section, we review the different parameterizations that have been considered so far. At the end of the section, we also propose a new one. It enables in particular time-dependant heterogeneous unresolved velocity statistics without relying on off-line high-resolution simulation outputs.

### 6.3.1 Kraichnan model

Similarities between LU-SALT dynamics and the Kraichnan model (Kraichnan, 1968, 1994; Gawędzki and Kupiainen, 1995; Majda et al., 1999) were already highlighted by Mémin (2014). Kraichnan (1968) focuses on the passive scalar transport by a homogeneous ( $\mathbf{Cov}(\mathbf{x}, \mathbf{y}, t, t') = \mathbf{Cov}(\mathbf{x} - \mathbf{y}, t, t')$ ), isotropic ( $\mathbf{Cov}(\mathbf{x}, \mathbf{y}, t, t') = \mathbf{Cov}(\|\mathbf{x} - \mathbf{y}\|, t, t')$ ) and divergence-free velocity field uncorrelated in time. With LU notations, it implies:

$$\mathbf{w} = 0, \quad (47)$$

$$\nabla \cdot \boldsymbol{\sigma} = 0, \quad (48)$$

$$\boldsymbol{\sigma} d\mathbf{B}_t = \check{\boldsymbol{\sigma}} * d\mathbf{B}_t, \quad (49)$$

where  $*$  denotes a convolution and  $\check{\boldsymbol{\sigma}}$  is a particular (isotropic) filter. The homogeneity condition (49) makes the parameterization easier in Fourier space:

$$\mathbb{E} \left\{ \left( \widehat{\boldsymbol{\sigma} d\mathbf{B}_t} \right) (\mathbf{k}_1, t) \left( \widehat{\boldsymbol{\sigma} d\mathbf{B}_{t'}} \right)^T (\mathbf{k}_2, t') \right\} = \boldsymbol{\Gamma}_\sigma(\mathbf{k}_1) \delta(\mathbf{k}_1 - \mathbf{k}_2) \delta(t - t') dt, \quad (50)$$

where  $\boldsymbol{\Gamma}_\sigma$  can be taken to mimic the self-similarities observed in real turbulence flow (Gawędzki and Kupiainen, 1995):

$$\boldsymbol{\Gamma}_\sigma(\mathbf{k}) = A^2 \widehat{\mathcal{P}}(\mathbf{k}) \left( 1 + \frac{\|\mathbf{k}\|^2}{\kappa_m^2} \right)^{-2\alpha}, \quad (51)$$

with  $A$  is a constant,  $1/\kappa_m$  specifies the velocity correlation length and

$$\widehat{\mathcal{P}}(\mathbf{k}) = \mathbb{I}_d - \frac{\mathbf{k}\mathbf{k}^T}{\|\mathbf{k}\|^2}, \quad (52)$$

is the projection – expressed in Fourier space – onto the space of solenoidal functions. The Fourier transform of the associated filter  $\check{\boldsymbol{\sigma}}$  (appearing in (49)) is:

$$\widehat{\check{\boldsymbol{\sigma}}}(\mathbf{k}) = A \widehat{\mathcal{P}}(\mathbf{k}) \left( 1 + \frac{\|\mathbf{k}\|^2}{\kappa_m^2} \right)^{-\alpha}. \quad (53)$$

Finally, the Fourier transform of the small-scale velocity writes:

$$\widehat{\boldsymbol{\sigma} \dot{\mathbf{B}}}(\mathbf{k}) = \frac{A}{\sqrt{\Delta t}} \widehat{\mathcal{P}}(\mathbf{k}) \left( 1 + \frac{\|\mathbf{k}\|^2}{\kappa_m^2} \right)^{-\alpha} \widehat{\frac{d\mathbf{B}_t}{\sqrt{\Delta t}}}(\mathbf{k}), \quad (54)$$

where  $\widehat{d\mathbf{B}_t}$  is the spatial Fourier transform of  $d\mathbf{B}_t$ , with  $d\mathbf{B}_t/\sqrt{\Delta t}$ , a vector of  $d$  independent discrete white noise processes of unit variance in space and time. To sample the small-scale velocity, we first sample  $d\mathbf{B}_t/\sqrt{\Delta t}$ , to get  $\widehat{d\mathbf{B}_t}/\sqrt{\Delta t}$ , and finally  $\widehat{\boldsymbol{\sigma} \dot{\mathbf{B}}}(\mathbf{k})$  with the above equation.

Note that the same unresolved simulation is possible without relying on Fourier transform. Indeed, one can use the fact that  $\widehat{\mathcal{M}}(\mathbf{k}) \triangleq \left( 1 + \frac{\|\mathbf{k}\|^2}{\kappa_m^2} \right)^{-\alpha}$  is

the Fourier transform of a Matérn covariance  $\mathcal{M}$  of degree  $\nu = \alpha - 1$  and of range  $\rho = \sqrt{2(\alpha - 1)}/\kappa_m$ . Therefore, the unresolved velocity is:

$$\boldsymbol{\sigma} d\mathbf{B}_t = A\mathcal{P} \{\mathcal{M} * d\mathbf{B}_t\}, \quad (55)$$

with

$$\mathcal{P} = \mathbb{I}_d - \Delta^{-1} \nabla \nabla^T. \quad (56)$$

### 6.3.2 The homogeneous stationary model

Assuming zero large-scale velocity is obviously not possible in CFD. However, combined with a resolved large-scale velocity, a homogeneous and stationary small-scale velocity can be a relatively good approximation for some flows. Note that even a homogeneous small-scale velocity leads to a heterogeneous non-Gaussian random forcing, due to its multiplicative structure.

Resseguier et al. (2017b) show numerical simulations of a LU version of a 2-dimensional geophysical model referred to as the Surface Quasi-Geostrophic (SQG) model (Blumen, 1978; Held et al., 1995; Lapeyre, 2017). For this purpose, a modification of (53) is considered. Instead of involving the operator  $\mathcal{P} = \mathbb{I}_d - \Delta^{-1} \nabla \nabla^T$ , Resseguier et al. (2017b) work on the streamfunction  $\psi_\sigma d\mathbf{B}_t$ . As such no additional divergence-free constraint is needed. The streamfunction kernel,  $\check{\psi}_\sigma$  is defined such as:

$$\boldsymbol{\sigma}(\mathbf{x}) d\mathbf{B}_t = \nabla^\perp \psi_\sigma(\mathbf{x}) d\mathbf{B}_t = \left( \nabla^\perp \check{\psi}_\sigma * d\mathbf{B}_t \right) (\mathbf{x}). \quad (57)$$

with  $\nabla^\perp = (-\partial_y, \partial_x)^T$ . Accordingly, a single cylindrical Wiener process,  $B_t$ , is sufficient to sample the Gaussian process. This is specific to two-dimensional domains. In 3D, a vector of 3 independent  $\mathbb{I}_d$ -cylindrical Wiener processes, and a curl must be considered to simulate an isotropic small-scale velocity. Thus, the small-scale velocity can be conveniently specified from its omnidirectional spectrum:

$$\mathbf{k} \mapsto \frac{1}{\mu(\Omega)} \mathbb{E} \oint_{[0, 2\pi]} d\theta_{\mathbf{k}} \|\mathbf{k}\| \left\| \widehat{\boldsymbol{\sigma} \mathbf{B}}(\mathbf{k}) \right\|^2 = \frac{2\pi}{\Delta t} \|\mathbf{k}\|^3 \left| \widehat{\check{\psi}_\sigma}(\|\mathbf{k}\|) \right|^2, \quad (58)$$

where  $\mu(\Omega)$  is the surface of the spatial domain  $\Omega$ ,  $\theta_{\mathbf{k}}$  is the angle of the wave-vector  $\mathbf{k}$  and  $\Delta t$  the simulation time-step. Consistent with SQG turbulence, the omni-directional spectrum slope, denoted  $s$ , is fixed to  $-5/3$ . For 2D Euler equations, the slope would be set to  $-3$ . The unresolved velocity should be energetic only where the dynamics cannot be properly resolved. Consequently, the authors apply to the spectrum a smooth band-pass filter,  $f_{BP}$ , which has non-zero values between two wavenumbers  $\kappa_m$  and  $\kappa_M$ . The parameter  $\kappa_m$  is inversely related to the spatial correlation length of the unresolved component.  $\kappa_M$  is set to the theoretical resolution,  $\pi/\Delta x$ , and  $\kappa_m$  to the estimated effective

resolution ( $\kappa_m = \kappa_M/2$  in Resseguier et al. (2017b)). The small scales' energy is specified by the diffusion coefficient  $a$  and the simulation time step:

$$\mathbb{E} \left( \boldsymbol{\sigma} \dot{\mathbf{B}} \right) \left( \boldsymbol{\sigma} \dot{\mathbf{B}} \right)^T = \frac{1}{\Delta t} \mathbf{a} = \frac{1}{\Delta t} \begin{pmatrix} a_0 & 0 \\ 0 & a_0 \end{pmatrix}. \quad (59)$$

The diagonal structure of the variance tensor is due both to incompressibility and isotropy. The scalar variance tensor,  $a_0$ , is similar to an eddy viscosity coefficient. So, a typical value of eddy viscosity used in practice is a good proxy to setup this parameter. Otherwise, this parameter can be tuned. The time step depends itself, through the CFL conditions, on both the spatial resolution and the maximum magnitude of the resolved velocity. Finally, similarly to the Kraichnan model, the Fourier transform of the small-scale velocity writes:

$$\widehat{\boldsymbol{\sigma} \dot{\mathbf{B}}}(\mathbf{k}) \triangleq \frac{A}{\sqrt{\Delta t}} i\mathbf{k}^\perp \text{f}_{\text{BP}}(\|\mathbf{k}\|) \|\mathbf{k}\|^{-\alpha} \widehat{\frac{dB_t}{\sqrt{\Delta t}}}(\mathbf{k}) \text{ with } s = 3 - 2\alpha = -\frac{5}{3}, \quad (60)$$

where  $A$  is a constant to ensure  $\mathbb{E} \|\boldsymbol{\sigma} \dot{\mathbf{B}}\|^2 = \text{tr}(\mathbf{a})/(\Delta t) = 2a_0/(\Delta t)$  (see (59) above),  $\widehat{dB_t}$  is the spatial Fourier transform of  $dB_t$ , with  $dB_t/\sqrt{\Delta t}$ , a discrete scalar white noise process of unit variance in space and time. To sample the small-scale velocity, we first sample  $dB_t/\sqrt{\Delta t}$ , to get  $\widehat{dB_t}/\sqrt{\Delta t}$ , and finally  $\widehat{\boldsymbol{\sigma} \dot{\mathbf{B}}}(\mathbf{k})$  with the above equation.

A MATLAB code simulating the model under location uncertainty with this parameterization for a SQG flow is available online (<http://vresseguier.github.io/sqgm>). Resseguier et al. (2017b) have numerically demonstrated the good UQ skills of this method for SQG flows.

### 6.3.3 The homogeneous non-stationary model

A main drawback of the previous choice of  $\boldsymbol{\sigma}$  is the needed tuning of the parameters. Resseguier et al. (2019) propose a more general homogeneous parameterization where no tuning is needed. A new  $\boldsymbol{\sigma}$  will be defined at each time step from the resolved velocity kinetic energy spectrum.

In the previous parameterization, the absolute diffusivity (i.e. KE times correlation time (Falkovich et al., 2001; Penland, 2003b; Klyatskin, 2005; Vallis, 2006; Keating et al., 2011)) of the unresolved velocity is twice the variance tensor trace  $\tau(\mathbf{a})$  whereas the unresolved kinetic energy is  $\tau(\mathbf{a})/dt$ . Clearly, this kinetic energy has no physical meaning. Indeed, it depends on the simulation time step and one should have the possibility to choose the time step as close as possible to zero. Thus, the unresolved velocity amplitude is specified through an absolute diffusivity rather than through a KE. In the mathematics literature of homogenization, Kubo-type formulas may be seen as what physicists call absolute diffusivities. More generally, since the variance of a time-continuous white noise is infinite, it is more relevant to deal with absolute diffusivity rather than kinetic energy in order to describe the statistics of the time-uncorrelated velocity. Thus, keeping a spectral approach, we define – for any spatio-temporal field – an Absolute Diffusivity Spectral Density

(ADSD) denoted  $A(\kappa)$  at the wave-number  $\kappa$ . We will rely on this ADSD rather than on the KE spectrum,  $E(\kappa)$ . Since the absolute diffusivity is the variance multiplied by the correlation time, it is naturally to defined the ADSD as follows:

$$A(\kappa) \triangleq E(\kappa)\tau(\kappa) \text{ where } \tau(\kappa) = \frac{1/\kappa}{v_\kappa} = \frac{1/\kappa}{\sqrt{\kappa E(\kappa)}}, \quad (61)$$

is the eddy turnover time at the scale  $1/\kappa$  and  $v_\kappa$  is at characteristic velocity a this scale. Accordingly, we have:

$$A(\kappa) = \kappa^{-3/2} E^{1/2}(\kappa). \quad (62)$$

If in addition we assume a KE self-similar distribution,

$$E(\kappa) = C^2 \kappa^{-s}, \quad (63)$$

we obtain:

$$A(\kappa) = C \kappa^{-r}, \quad (64)$$

where  $r = (s + 3)/2$ .

We aim at defining the unresolved velocity ADSD from the large-scale velocity. For this purpose, we will assume the self similar model (64) is valid at all spatial scales. At each time step, we compute the ADSD of the large-scale velocity,  $A_{\mathbf{w}}$ , from formula (62). Then, we fit the coefficients  $C$  and  $r$  of equation (64). Let us denote with  $C_{\mathbf{w}}$  and  $r_{\mathbf{w}}$  these coefficients. Note that they are time-dependent because they depend on  $\mathbf{w}$  which is. More precisely, we estimate the coefficients  $C_{\mathbf{w}}$  and  $r_{\mathbf{w}}$  in a wavenumber interval which approximately represents a inertial range of fully-resolved scales (i.e. before the spectrum roll-off).

From there, we can define the operator  $\sigma$  in such a way that the total velocity – resolved plus unresolved – meets (64) at small spatial scales with the same coefficients ( $C_{\mathbf{w}}, r_{\mathbf{w}}$ ). Since the two velocity components are not correlated, the total ADSD is the sum of the ADSD of each velocity component. Therefore,  $\sigma$  is chosen such as the unresolved ADSD compensates the resolved ADSD roll off – introduced by the deterministic subgrid parameterization – at small scales. Specifically, the unresolved ADSD is set to:

$$A_{\sigma \hat{B}}(\kappa) = \max(0, C_{\mathbf{w}} \kappa^{-r_{\mathbf{w}}} - A_{\mathbf{w}}(\kappa)) f_{BP}^2(\kappa). \quad (65)$$

As previously,  $f_{BP}^2$  is a band-pass filter between  $\kappa_m$  and  $\kappa_M$ . In practice, we set  $\kappa_M$  to the theoretical resolution:

$$\kappa_M = \frac{\pi}{\Delta x}, \quad (66)$$

and  $\kappa_m$  to the effective resolution which is estimated as follows:

$$\kappa_m = \left( \frac{\ln(0.95)}{\ln(0.1)} \right)^{1/p} \kappa_M, \quad (67)$$

where  $p$  is the order of the Laplacian used as deterministic subgrid tensor (i.e.  $D_t b = -\nu(-\Delta)^p b$   $dt$ ). The justification of the above formula is left in the Appendix C. Compared to the work of Resseguier et al. (2017b), the value of  $\kappa_m$  is less critical. Indeed, equation (65) implies a weaker unresolved ADSD at larger scales where the resolved ADSD,  $A_w$  is stronger. This softens the threshold effect introduced by the band-pass filter  $f_{BP}$ .

In practice, we set an upper-bound for the estimation of  $r_w$ . Without this upper bound, a concentration of energy at relatively large wave-numbers – scales smaller than  $\kappa_m$  – in the resolved fields can become unstable. Indeed, this localized energy concentration would decrease the  $r_w$  estimation, and hence increase the unresolved ADSD  $A_{\sigma\dot{B}}$  through (65) at large wave-numbers – larger than  $\kappa_m$ . This implies a larger noise intake, which can induce a larger concentration of energy at relatively large wave-numbers in the resolved fields, resulting in a positive feedback loop. To prevent these unphysical instabilities, the slope  $r_w$  is bounded.

In order to link the unresolved ADSD to the kernel  $\check{\sigma}$  which defines the unresolved velocity, we note that

$$A_{\sigma\dot{B}}(\kappa) = E_{\sigma\dot{B}}(\kappa)dt = \frac{1}{\mu(\Omega)} \oint_{[0,2\pi]} d\theta_{\mathbf{k}\kappa} \left\| \widehat{\check{\sigma}}(t, \kappa) \right\|^2 = 2\pi\kappa^3 \left| \widehat{\check{\psi}}_{\sigma}(t, \kappa) \right|^2. \quad (68)$$

From formulas (65)-(68), we can finally express the unresolved velocity as follow:

$$\widehat{\sigma\dot{B}}(t, \mathbf{k}) = \frac{1}{\sqrt{\Delta t}} i\mathbf{k}^{\perp} \widehat{\check{\psi}}_{\sigma}(t, \|\mathbf{k}\|) \frac{d\widehat{B}_t}{\sqrt{\Delta t}}(\mathbf{k}), \quad (69)$$

$$= \frac{1}{\sqrt{\Delta t}} i\mathbf{k}^{\perp} \sqrt{\frac{\max(0, C_w \|\mathbf{k}\|^{-r_w} - A_w(\|\mathbf{k}\|))}{2\pi \|\mathbf{k}\|^3}} f_{BP}(\|\mathbf{k}\|) \frac{d\widehat{B}_t}{\sqrt{\Delta t}}(\mathbf{k}). \quad (70)$$

Again the simulated unresolved velocity ADSD is physically relevant while the KE spectrum is not. Indeed, the simulated unresolved velocity ADSD is expected to match the true (time-correlated) unresolved velocity ADSD, whereas the KE spectra of the simulated and true unresolved velocities differ. Indeed, the true unresolved velocity correlation time spectral distribution  $\tau(\kappa)$  is not restricted to the time step  $dt$ .

Resseguier et al. (2019) have successfully applied this parameterization for UQ purpose in a SQG flow.

#### 6.3.4 Heterogeneous modulation of homogeneous models

The SALT-LU stochastic parameterization randomly folds tracer isolines. This process is often desirable. For instance, it can trigger physically-relevant instabilities, such as the filament instabilities highlighted by Resseguier et al. (2017b). After these instabilities have been randomly triggered, eddies are formed by non-linear processes. However, a homogeneous small-scale velocity may also perturb the tracer isolines which should remain still (i.e. which remain

still in high-resolution deterministic simulations), e.g. sharp, straight, coherent fronts. A typical application of this problem in more realistic flow simulations is the simulation of jets like the Gulf Stream and regions of the Antarctic Circumpolar Current. These real-world jets are associated with diffusivity suppression (Ferrari and Nikurashin, 2010), and this effect is not present in our formulation so far. If we seek to preferentially perturb some tracer gradients, a heterogeneous small-scale velocity is required. Note that the heterogeneity discussed here needs to be non-stationary and thus cannot be represented by the stationary EOF presented later in this paper. Besides, in a small ensemble of realizations, relevant heterogeneity of the small scales may make the spreading more accurate for UQ and enable comparable ensemble forecast accuracy with fewer members. We here propose a possible heterogeneous version of the previous method.

In order to obtain a heterogeneous model of the unresolved velocity, we need a heterogeneous version of the ADSD (64). Since the wave-number  $\kappa$  cannot depend of the position  $\mathbf{x}$ , the constant  $C_w$  and/or the spectrum slope  $r$  should do. A spatially varying spectrum slope is probably difficult to estimate. Hence, we restrict ourselves to a spatially-varying constant  $C_w$  and a spatially-homogeneous spectrum slope. The constant may also vary with the time and the wave number. According to the Kolmogorov theory (e.g. Frisch, 1995) and (63):

$$C_w(\mathbf{x}, t, \kappa) = \text{cst} \cdot \epsilon_F^p(\mathbf{x}, \kappa), \quad (71)$$

where  $\epsilon_F$  is the energy flux through the spatial scales and  $p = 1/3$  for a SQG flow. More specifically, the energy flux describes the energy moving from scales larger than  $1/\kappa$  toward smaller scales and can be computed as follows:

$$\epsilon_F(\mathbf{x}, t, \kappa) \triangleq \overline{q_\kappa^<((\mathbf{w} \cdot \nabla)q)_\kappa^<}}, \quad (72)$$

where  $q$  is the transported (up to possible source terms) quantity,  $g_\kappa^<$  is the low-pass filtered version of  $g$  (setting to zero the Fourier modes of  $g$  which have frequencies larger than  $\kappa$ ) and  $\overline{\bullet}$  stands for the spatial average (Frisch, 1995). For a SQG flow,  $q$  corresponds to the buoyancy normalized by the stratification:  $q = b/N$ . The energy flux is essentially a third-order moment. It is very important because it describes the cascade of the flow by non-linear energy transfers (Frisch, 1995).

If the energy flux through scale is understood locally in space (as indeed Smagorinsky (1963) also assumes), the formula (71) provides a natural parameterization of the unresolved velocity heterogeneities. We simply modulates the unresolved ADSD (65) by the heterogeneous ratio  $\epsilon_F^p/\overline{\epsilon_F^p}$  – averaged over the

resolved inertial range wave-numbers:

$$\widehat{\sigma \dot{\mathbf{B}}}(t, \mathbf{k}) = \frac{1}{\sqrt{\Delta t}} i\mathbf{k}^\perp \sqrt{\frac{\max(0, C_w \|\mathbf{k}\|^{-r_w} - A_w(\|\mathbf{k}\|))}{2\pi \|\mathbf{k}\|^3}} f_{BP}(\|\mathbf{k}\|) \frac{d\widehat{B}_t}{\sqrt{\Delta t}}(\mathbf{k}), \quad (73)$$

$$\sigma \dot{\mathbf{B}}(t, \mathbf{x}) = \mathcal{P} \left\{ \underbrace{\sqrt{\frac{\epsilon_F^p(t, \mathbf{x})}{\epsilon_F^p(t)}}}_{\text{Heterogeneous modulation}} \underbrace{\widetilde{\sigma \dot{\mathbf{B}}}(t, \mathbf{x})}_{\text{Homogeneous velocity}} \right\}, \quad (74)$$

where  $\mathcal{P} = \mathbb{I}_d - \Delta^{-1} \nabla \nabla^T$  is the projector onto the space of free-divergence functions. This parameterization is physically meaningful since, locally in space, a stronger direct cascade at large scales (larger  $\epsilon_F$  and thus larger  $C_w$ ) suggests that the unresolved velocity (large) should maintain this cascade by folding smaller-scale tracer structures. Furthermore, considering that the energy flux is a third-order structure makes this parameterization relevant to differentiate between strait fronts and curved structures (e.g. eddies). Indeed, at least three points are needed to define a curvature and differentiate between these structures.

In order to keep a divergence-free velocity and the ensuing properties (e.g. energy conservation), the modulated velocity is projected onto the space of free-divergence functions, using the operator  $\mathcal{P}$ . Because of that we do not consider the advection correction  $\mathbf{w}^* - \mathbf{w} = -\frac{1}{2}(\nabla \cdot \mathbf{a})^T$  of the LU formalism. Indeed, here the variance tensor has the simple form  $\mathbf{a} = \frac{1}{d} \text{tr}(\mathbf{a}) \mathbb{I}_d$ . As such, the advection correction is a gradient field and is hence removed by the projection onto the space of free-divergence functions.

Resseguier et al. (2019) have numerically shown that this parameterization indeed keeps the sharp straight fronts while perturbing the eddies and the meanders in SQG flows.

### 6.3.5 The heterogeneous spatially-uncorrelated model

Yang and Mémin (2018) have considered a spatially heterogeneous subgrid velocity. The variance tensor  $\mathbf{a}$  is estimated on-line by data assimilation. To simplify, they neglect the spatial correlation of that velocity.

### 6.3.6 Off-line learning of the EOF from velocity

It is always possible to write the unresolved velocity as follows (see e.g. Mémin, 2014):

$$\sigma(\mathbf{x}, t) \dot{\mathbf{B}} = \sum_{k=1}^{n_{KL}} \boldsymbol{\xi}_k(\mathbf{x}, t) \dot{W}_k, \quad (75)$$

with  $W_k$  are independent one-dimensional Brownian motions and  $n_{KL}$  possibly infinite. The  $(\boldsymbol{\xi}_k)_k$  are so-called (weighted) Empirical Orthogonal Functions



(EOFs). If we denotes by  $\lambda_k$  the  $L^2$  norm of the EOFs, their normalized versions  $(\tilde{\boldsymbol{\xi}}_k = \boldsymbol{\xi}_k / \lambda_k)_k$  are the eigenvectors of the auto-adjoint operator defined by the kernel  $\boldsymbol{\sigma}(\mathbf{x}, t)\boldsymbol{\sigma}^T(\mathbf{y}, t)$ . The  $(\lambda_k)_k$  are the corresponding eigenvalues. That kernel corresponds – up to the factor  $\frac{1}{dt}$  – to the one-time two-point covariance of the unresolved velocity. In the case of stationary EOFs, the type of decomposition (75) is supported by the mathematical theory of  $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ -Wiener processes (Da Prato and Zabczyk, 1992; Prévôt and Röckner, 2007).

First, we will consider the off-line learning of the EOFs  $(\boldsymbol{\xi}_k)_k$  from a velocity field set. Specifically, we will assume that we have a history of “small-scale” velocity fields  $(\mathbf{v}'(\mathbf{x}_{ij}, m))_{1 \leq m \leq n_O}$  at  $n_O$  different times  $t_m$  (or  $n_O$  different “realizations”) sampled on a spatial grid of  $M$  points  $(\mathbf{x}_{ij})_{ij} = ((i\Delta x, j\Delta y))_{ij}$ . Typically, the velocity snapshots are high-resolution simulation outputs or high-pass-filtered versions of it. We also need to assume that the EOFs are stationary in time. For the sake of clarity, we restrict the methods presentation to  $d = 2$ . Notwithstanding, the generalizations to three-dimensional flow are straightforward.

#### *Candidate for the increment realizations*

In order to estimate the EOFs,  $(\boldsymbol{\xi}_i)_i$ , we assume that we can observe increments

$$\int_{t_m}^{t_m + \Delta T} \boldsymbol{\sigma} d\mathbf{B}_{t'} = \sum_{k=1}^{n_{KL}} \boldsymbol{\xi}_k(\mathbf{x}_{ij})(W_k(t_m + \Delta T) - W_k(t_m)). \quad (76)$$

at several times  $t_m$ . Here,  $\Delta T$  is the time step of a low-resolution simulation. We will interpret the following snapshots:

$$\widetilde{\Delta \mathbf{X}}_{ij}^m \triangleq \mathbf{v}'(\mathbf{x}_{ij}, m)\Delta T, \quad (77)$$

as a realizations of the flow increments (76). Since Brownian time increments (76) are independent, by the above interpretation, we implicitly assume that  $\mathbf{v}'(\mathbf{x}_{ij}, m)$  are independent for different values of  $m$ . In this case, it is equivalent to consider them as a set of independent “realizations” or as a set of snapshots at several times  $t_m$ .

#### *Preprocessing*

The increments are supposed to be centered and divergence free ( $\nabla \cdot \boldsymbol{\xi}_i = 0$ ). Therefore, after computing the residual flow increments  $(\widetilde{\Delta \mathbf{X}}_{ij}^m)_m$ , they are centered:

$$\Delta \mathbf{X}'_{ij}{}^m \triangleq \widetilde{\Delta \mathbf{X}}_{ij}^m - \hat{\mathbb{E}}\{\widetilde{\Delta \mathbf{X}}_{ij}^m\}, \quad (78)$$

with the estimator

$$\hat{\mathbb{E}}\{\widetilde{\Delta\mathbf{X}}_{ij}^m\} \triangleq \frac{1}{n_O} \sum_{m=1}^{n_O} \widetilde{\Delta\mathbf{X}}_{ij}^m, \quad (79)$$

and projected onto the space of divergence-free functions:

$$\Delta\mathbf{X}_{ij}^m \triangleq \mathcal{P}\{\Delta\mathbf{X}_{ij}^m\} \text{ with } \mathcal{P} = \mathbb{I}_d - \Delta^{-1}\nabla\nabla^T. \quad (80)$$

Note that the projection can also be applied directly on the final EOFs  $\boldsymbol{\xi}_k^m$  – at the end of the estimation process – rather than on each increments  $\Delta\mathbf{X}_{ij}^m$ . Besides, if  $\nabla \cdot v' = 0$  then the centred flow increments are already divergence-free and the step (80) can be skipped.

### Covariance and EOF

Then, we can define the EOFs from an estimate of the spatial covariance of the residual flow increments (averaging over the time index  $m$ ) :

$$\gamma_\sigma(\mathbf{x}_{ij}, \mathbf{x}_{pq}) \approx \frac{1}{(n_O - 1)\Delta T} \sum_{m=1}^{n_O} \Delta\mathbf{X}_{ij}^m (\Delta\mathbf{X}_{pq}^m)^T. \quad (81)$$

Besides, we have:

$$\gamma_\sigma(\mathbf{x}_{ij}, \mathbf{x}_{pq}) = \sum_{k=1}^{n_{KL}} \boldsymbol{\xi}_k(\mathbf{x}_{ij}) \boldsymbol{\xi}_k^T(\mathbf{x}_{pq}). \quad (82)$$

In order to properly define the EOFs, we must add the following orthogonal constraint:

$$\sum_{pg} \boldsymbol{\xi}_i(\mathbf{x}_{pg}) \cdot \boldsymbol{\xi}_j(\mathbf{x}_{pg}) = 0 \text{ if } i \neq j. \quad (83)$$

According to equations (82) and (83), the EOFs can be obtained from a diagonalization of the covariance  $(\gamma_\sigma(\mathbf{x}_{ij}, \mathbf{x}_{pq}))_{(ij),(pq)}$ .

### Dual problem for the EOF estimation

Since the number of grid points  $M$  is often larger than the number of increments,  $n_O$ , the covariance (81) can be a very large matrix ( $Md \times Md$ ). It is easier to consider the dual problem to estimate the EOFs. As such, the temporal covariance is first computed:

$$\tilde{\gamma}_\sigma(t_p, t_q) \triangleq \frac{1}{M} \sum_{ij} \frac{1}{\Delta T} \Delta\mathbf{X}_{ij}^p \cdot \Delta\mathbf{X}_{ij}^q. \quad (84)$$

After diagonalizing this  $n_O \times n_O$  matrix, the EOFs are obtained by projecting the eigenvectors onto the normalized increments  $\frac{1}{\sqrt{\Delta T}} \Delta\mathbf{X}_{ij}^k$ .

Finally, after estimating the  $(\boldsymbol{\xi}_k)_k$  off-line, the ensemble forecast can be generated on-line with the formula (75). This techniques has recently been applied for the stochastic simulation of turbulent channel flow. These stochastic simulation show how the inhomogeneity of the small-scales velocity components allows structuring the large-scale component in terms of streaks Chandramouli et al. (2019).

### 6.3.7 Off-line learning of EOF from flows

Now, we detail a procedure proposed by Cotter et al. (2018c,a,b) for the estimation of the EOFs,  $(\boldsymbol{\xi}_k)_k$ , involved in the unresolved velocity definition (75). Here again, the EOFs are assumed stationary in time. The main difference with the off-line learning of EOF from velocity 6.3.6 is the candidates which are used for the increments  $\widetilde{\Delta\mathbf{X}}_{ij}^m$ .

The data-driven method of Cotter et al. (2018c,a,b) relies on Lagrangian paths defined at two “resolutions”. The first paragraph of this section defines these two types of Lagrangian paths. Then, we explain how to obtain the candidates for the flow increment realizations from these Lagrangian paths.

#### Preliminary definitions

We introduce two types of velocity field:

- a high-resolution velocity  $\mathbf{v}$  on a fine spatial mesh-grid,
- a low-resolution velocity  $\bar{\mathbf{v}}$  on a coarse spatial mesh-grid. This velocity field is a spatially-low-pass-filtered version of  $\mathbf{v}$ .

Then, two types of Lagrangian path are defined:

- a “high-resolution flow”,  $\mathbf{X}_{ij}(t_0, t)$ , defined by the high-resolution velocity  $\mathbf{v}$ :

$$\frac{d\mathbf{X}_{ij}}{dt}(t) = \mathbf{v}(t, \mathbf{X}_{ij}(t_0, t)) \text{ and } \mathbf{X}_{ij}(t_0, t_0) = \mathbf{x}_{ij}. \quad (85)$$

- a “low-resolution flow”,  $\bar{\mathbf{X}}_{ij}(t_0, t)$ , defined by the low-resolution velocity  $\bar{\mathbf{v}}$ :

$$\frac{d\bar{\mathbf{X}}_{ij}}{dt}(t_0, t) = \bar{\mathbf{v}}(t, \bar{\mathbf{X}}_{ij}(t_0, t)) \text{ and } \bar{\mathbf{X}}_{ij}(t_0, t_0) = \mathbf{x}_{ij}. \quad (86)$$

#### Candidate for the increment realizations

In order to estimate the EOFs, we will interpret the following residual flow increments as realizations of the unresolved stochastic flow increments (76):

$$\widetilde{\Delta\mathbf{X}}_{ij}^m \triangleq \mathbf{X}_{ij}(t_m, t_m + \Delta T) - \bar{\mathbf{X}}_{ij}(t_m, t_m + \Delta T). \quad (87)$$

Then, the following steps are the same than in 6.3.6 (preprocessing and covariance diagonalization).

This method (Cotter et al., 2018c,a) may seem more natural than the one of 6.3.6, since the increments (87) are really flow increments. Nevertheless, the time step  $\Delta t$  being small,  $\widetilde{\Delta \mathbf{X}}_{ij}^m$  is close to  $(\mathbf{v} - \bar{\mathbf{v}})(t_m, \mathbf{x}_{ij}) \Delta T$ . Therefore, the main difference between the two methods is probably in the definition of the “residual velocities”:  $\mathbf{v}'$  on one hand and  $(\mathbf{v} - \bar{\mathbf{v}})$  on the other hand.

The method 6.3.7 has been successfully tested on a 2D Euler flow (Cotter et al., 2018c), a quasi-geostrophic flow (Cotter et al., 2018a) and a SQG flow (Resseguier et al., 2019). Resseguier et al. (2019) also compare the data-driven method 6.3.7 and the homogeneous non-data driven method 6.3.3 for forced homogeneous SQG turbulence. Similarly good UQ results are obtained.

### 6.3.8 Off-line learning of the stochastic advection operator

The accuracy of the unresolved velocity Karhunen-Loeve decomposition (75) is influenced by the number of kept modes EOFs,  $n_{KL}$ . In fluid mechanics, the state-space dimension is often huge. Therefore, that kind of spectral decomposition of the covariance often involves a number of modes much smaller than the state-space dimension. Hence, the decomposition (75) strongly reduces the sampling computational cost since there is no need to manipulate the full spatial covariance.

Nevertheless, for specific applications like dimensionally-reduced models, the state space can be reduced to a small well-chosen vector space (e.g. Holmes et al., 1996). By definition, the resolved velocity component lies on this subspace, whereas the unresolved component does not (Mémin, 2014; Resseguier et al., 2015, 2017c). If the state-space dimension – says  $n_r$  – is small enough, it can be more interesting to directly work with the statistics of the random operator:

$$f \mapsto \sigma d\mathbf{B}_t \cdot \nabla f, \quad (88)$$

expressed in the reduced basis, rather than dealing with the unresolved velocity statistics. Indeed, the reduced representation of the stochastic operator is a (conditionally-) Gaussian matrix of size  $n_r \times n_r$ . It is centered and time-uncorrelated. Therefore, the probability law of this matrix is exactly and entirely determined by its covariance, which have  $n_r^2 \times n_r^2$  coefficients.

In comparison, the approximate unresolved velocity Karhunen-Loeve decomposition would lead to stochastic operator

$$f \mapsto \sum_{k=1}^{n_{KL}} (\boldsymbol{\xi}_k \cdot \nabla f) \dot{W}_k, \quad (89)$$

characterized by  $n_r^2 \times n_{KL}$  coefficients. For  $n_{KL} < n_r^2$ , this last representation is necessary an approximation (since the covariance of the exact representation of (89) in the reduced subspace is of full rank, i.e. of rank  $n_r^2$ ). For small  $n_r$ , one can rely on the random operator reduced representation instead. Rigorous and low-cost estimators exist for this representation covariance matrix coefficients and provide good UQ skills (Resseguier, 2017, chapter 8).

Nevertheless, for large state-space dimension  $n_r$ , this  $n_r^2 \times n_r^2$  covariance matrix of the random operator is too large to handle. Hence, additional assumptions on the unresolved small-scale velocity (e.g. homogeneity, parametric model, transport) or a Karhunen-Loeve decomposition become necessary.

### 6.3.9 Transport of the unresolved velocity

Methods 6.3.6 and 6.3.7 learn the unresolved small-scale structure from high-resolution simulations. However, such simulations are not always available. Moreover, methods 6.3.6 and 6.3.7 assume heterogeneity without time dependence. While some turbulence heterogeneities could be fixed in time (e.g. heterogeneity due to boundaries conditions), we can expect that a large part of it moves with the flow.

Accordingly, Gay-Balmaz and Holm (2018) assume a “transport equation” – in the sense of differential geometry – for the EOFs:

$$D_t^S \xi_k = (\xi_k \cdot \nabla)(\mathbf{w}^* dt + \sigma \circ d\mathbf{B}_t). \quad (90)$$

As such, the unresolved velocity is both heterogeneous and non-stationary in time, without requiring any learning. However, as far as we know, this method has not been tested in numerical simulations yet.

### 6.3.10 On-line learning of EOF from velocity

We propose here a new approach – based on the EOF decomposition (75) – but where the EOFs are time-dependent. They are estimated on-line from a coarse simulation. For this purpose, we propose to generate pseudo-observations of the small scales directly from the resolved velocity. Then, we will compute an EOF representation as in 6.3.6. Finally, we will adapt the noise variance to smaller scales using a turbulence power-law scaling (originally proposed by Harouna and Mémin, 2017).

#### *Pseudo-observations*

The approach proposed in this section defines  $n_O$  pseudo-observations  $\mathbf{v}'$  at each simulation grid point before computing the singular value decomposition (SVD).

For a given time  $t$  and a given resolved low-resolution velocity component,  $\mathbf{w}$ , we build pseudo-observations by sliding a  $n_w \times n_w$  ( $n_w$  odd) window over the spatial grid. We denote by  $L = n_w l$  the spatial scale of the window, where  $l$  is the smallest scale of the simulation. At every grid point  $\mathbf{x}_{ij}$ , we list the  $n_w^2$  velocity values contained in the window centered at that point – with appropriate boundary conditions (replication, periodicity, etc. ) when looking at a point on the border:

$$I(\mathbf{x}_{ij}, t) \triangleq \left\{ \mathbf{w}(\mathbf{x}_{pq}, t) \mid |p - i| \leq \frac{n_w - 1}{2}, |q - j| \leq \frac{n_w - 1}{2} \right\}. \quad (91)$$

Then, for each  $m \in \{1, \dots, n_O\}$ , for each point  $\mathbf{x}_{ij}$  independently, we set the value of the pseudo-observation  $\mathbf{v}'(\mathbf{x}_{ij}, t, m)$  by randomly choosing a value in the set  $I(\mathbf{x}_{ij}, t)$ .

After this, we proceed as in 6.3.6 to obtain the EOFs, but averaging over the pseudo-realization index  $m$  instead of averaging over the time steps  $t_m$ .

### Rescaling

From the SVD, we obtain a set of EOFs  $\xi_k^{(L)}$  and hence a model for the unresolved velocity  $\sigma^{(L)}(\mathbf{x}, t)\dot{\mathbf{B}}$ :

$$\sigma^{(L)}\dot{\mathbf{B}} = \sum_{k=1}^{n_{KL}} \xi_k^{(L)} \dot{W}_k. \quad (92)$$

The pseudo-realizations  $\mathbf{v}' = \mathbf{v}'_L$  have been generated at a spatial scale  $L = nl$ . These fluctuations correspond to a virtual observation scale  $L$  ( $n \times n$  window) and must be scaled down to the ‘‘simulation scale’’  $l$ . Therefore the unresolved velocity variance tensor,  $\mathbf{a}$ , is rescaled by a coefficient proposed by Harouna and Mémin (2017), and which in 2D reads:

$$\mathbf{a}_l = \left(\frac{l}{L}\right)^{2/3} \mathbf{a}_L, \quad (93)$$

where  $\mathbf{a}_L$  and  $\mathbf{a}_l$  are the variance tensors at the scales  $L$  and  $l$  respectively. This scaling relies on the Kolmogorov-Richardson cascade assumption with the velocity fluctuations at scale  $\ell$  as  $u_\ell \propto \epsilon^{1/3} \ell^{1/3}$ . The unresolved velocity,  $\sigma\dot{\mathbf{B}}$ , can be finally simulated – at the ‘‘simulation scale’’  $l$  – as:

$$\sigma\dot{\mathbf{B}} = n^{-1/3} \sigma^{(L)}\dot{\mathbf{B}}. \quad (94)$$

Let us note that in 3D the scaling must be adapted as the power exponent is 4/3 Harouna and Mémin (2017).

The on-line SVD method 6.3.10 will be numerically tested in section 8. There, the LU parameterizations 6.3.2, 6.3.3 and 6.3.10 will be compared to the new scheme 5.3.2 and to ensembles generated by initial condition perturbations.

## 7 Metrics for UQ

In this section, we detail some important metrics to quantify ensemble forecasts prediction skills.

This section is a short summary about existing verification ensemble tools. More detailed review could be found in (Hamill and Colucci, 1997; Gneiting and Raftery, 2007; Scheuerer and Hamill, 2015; Thorarinsdottir, 2017).

### 7.1 Talagrand diagram

A Talagrand diagram (or rank histogram) is a technique used to check the reliability of an ensemble forecast or a set of quantile. The idea is considers the availability of  $N$  observation  $q_1^o, \dots, q_N^o$  in a set of rank  $r$ . The sets of rank are built from rank statistic which is a sorted ensemble  $\{(q_1^{(i_1)}, \dots, q_1^{(i_{n_e})}), \dots, (q_N^{(i_1)}, \dots, q_N^{(i_{n_e})})\}$  of  $n_e$  members (Anderson, 1996; Talagrand, 1999; Hamill and Colucci, 1997):

$$\forall (j, k) \in \{1, \dots, n_e\} \times \{1, \dots, N\}, r_j = \frac{1}{N} \sum_{k=1}^N \widehat{\mathbb{P}}(q_k^{(i_{j-1})} \leq q_k^o < q_k^{(i_j)}), \quad (95)$$

where  $\widehat{\mathbb{P}}$  is the estimated probability of ranking an observation between two sorted ensemble members. In the context of (95),  $N$  represents the spatial points or times steps of the observation.

A calibrated ensemble should result in a flat histogram. However, a flat histogram does not guarantee a calibrated ensemble (Hamill, 2001). A flat histogram mostly indicates that the ensemble and observation are sampled from a common distribution. A U-shaped rank histogram corresponds to an underdispersion or to a conditional bias (Hamill, 2001) of the ensemble members. A dome-shaped shows an overdispersion of the ensemble. A nonsymmetric histogram is the footprint of bias.

### 7.2 Mean absolute and mean squared error

The mean absolute error (MAE) and mean squared error (MSE) are the state-of-the-art verification and selection tools specially used in regression problems as decision function. In an ensemble verification problem, the MAE is defined by the following formulation:

$$\forall (j, k) \in \{1, \dots, n_e\} \times \{1, \dots, N\}, \text{MAE} = \frac{1}{Nn_e} \sum_{k=1}^N \sum_{j=1}^{n_e} |q_k^o - q_k^{(j)}|, \quad (96)$$

and the MSE by this one:

$$\forall (j, k) \in \{1, \dots, n_e\} \times \{1, \dots, N\}, \text{MSE} = \frac{1}{Nn_e} \sum_{k=1}^N \sum_{j=1}^{n_e} (q_k^o - q_k^{(j)})^2. \quad (97)$$

The goal behind these two equations is to calculate the total error between a sample of observation and ensemble of size  $N$ . So, the empirical mean of all error between an observation  $q_k^o$  and each member  $q_k^{(j)}$  of the ensemble of size  $n_e$  is calculated. Both metrics are proper scores. A proper score is negatively oriented, such that a lower score indicated a better ensemble forecast. A proper score converging to zero means that the observation cannot be distinguished from the ensemble members. There are several proper scoring rules.

### 7.3 Continuous ranked proper score

Another proper score used for ensemble model evaluation is the continuous ranked proper score (CRPS). This kind of score is characterized by the so-called predictive distribution  $F$  – the distribution represented by the ensemble forecast  $\{q^{(1)}, \dots, q^{(n_e)}\}$  realizations and the observation  $q^o$ . The CRPS is defined by the following equation (Matheson and Winkler, 1976; Hersbach, 2000):

$$\text{CRPS}(F, q^o) = \int_{\mathbb{R}} (F(\mathcal{B}) - H(\mathcal{B} - q^o))^2 d\mathcal{B}, \quad (98)$$

where  $H$  is the heaviside function (equals to one if  $q^o \leq \mathcal{B}$  and zero otherwise),  $F$  is the cumulative distribution function (CDF) of the random process  $q$  (represented by the ensemble) at **one spatial point  $x$  and time  $t$** . The CRPS is the distance between the heaviside function and the distribution of the random variable  $q$ . This heaviside function represents the inequality between an ensemble member and the observation. If  $F$  has a finite first-order moment, there is another representation of the CRPS (Gneiting and Raftery, 2007):

$$\text{CRPS}(F, q^o) = \mathbb{E}_F |q - q^o| - \frac{1}{2} \mathbb{E}_F |q - q'|, \quad (99)$$

where  $q$  and  $q'$  are independent realizations of  $F$ . If a priori assumption are made on the behaviour of the random variable  $q$ . It is possible to deduce an analytic expression of the CRPS thanks to the equation (98) and (99).

In the case where  $F$  is issued from a normal distribution  $N(\mu, \sigma^2)$ , the CRPS formula is (Gneiting et al., 2005):

$$\text{CRPS}(F, q^o) = \sigma \left( \omega(2\Phi(\omega) - 1) + 2\phi(\omega) - \frac{1}{\sqrt{\pi}} \right), \quad (100)$$

where  $\omega = \frac{(q^o - \mu)}{\sigma}$ ,  $\phi$  is probability density function (PDF) of the standard normal distribution and  $\Phi$  is the CDF of the standard normal distribution.

Also following (Grimt et al., 2006), for an ensemble forecast  $\{q^{(1)}, \dots, q^{(n_e)}\}$ , a natural fair estimator of the CRPS is given by:

$$\widehat{\text{CRPS}}(F, q^o) = \frac{1}{n_e} \sum_{i=1}^{n_e} |q^{(i)} - q^o| - \frac{1}{2n_e^2} \sum_{i=1}^{n_e} \sum_{j=1}^{n_e} |q^{(i)} - q^{(j)}|. \quad (101)$$

The expression (101) allows to compute the CRPS without making a priori assumptions on our random variable. However, this expression has a  $\mathcal{O}(n_e^2)$  computational complexity.

An algebraically equivalent representation of the CRPS based on the generalized quantile function is proposed by (Laio and Tamea, 2007)

$$\widehat{\text{CRPS}}(\hat{F}, q^o) = \frac{2}{n_e^2} \sum_{j=1}^{n_e} \left( (q^{(i_j)} - q^o) (n_e H(q^{(i_j)} - q^o) - j + \frac{1}{2}) \right), \quad (102)$$



where  $\{q^{(i_1)}, \dots, q^{(i_{n_e})}\}$  is the sorted ensemble forecast. The computational complexity of the expression (102) is  $\mathcal{O}(n_e \log(n_e))$ . The interested reader can refer to (Jordan, 2016) for mathematical details on the discrete form of the CRPS.

#### 7.4 Energy score

In a CFD problem, observations and ensembles evolve in space and time. The CRPS is a pointwise score, hence cannot summarize the whole uncertainty of the ensemble in one value in a multivariate case. Therefore, to perform the evaluation of our ensemble in a multivariate case, the CRPS will not be enough. Of course, the means of the CRPS can be use. However, we would miss some information. Instead, the energy score can be considered.

There, our observation  $q^o$  and our random variable  $q$  is multivariate. The corresponding ensemble  $\{(q_1^{(1)}, \dots, q_N^{(1)}), \dots, (q_1^{(n_e)}, \dots, q_N^{(n_e)})\}$  contains  $n_e$  realizations. Each of them represents the field of interest in  $N$  time steps or spatial points.  $F$  is now the multivariate CDF associated to  $q$ . We need a generalized form of the CRPS to work in this case. Following (Gneiting and Raftery, 2007; Scheuerer and Hamill, 2015), a possible generalization of the CRPS is the energy score (ES):

$$\text{ES}(F, q^o) = \mathbb{E}_F \|q - q^o\| - \frac{1}{2} \mathbb{E}_F \|q - q'\|, \quad (103)$$

where  $\|\cdot\|$  is the euclidean norm. So, an estimator of the ES would be:

$$\widehat{\text{ES}}(\hat{F}, q^o) = \frac{1}{n_e} \sum_{i=1}^{n_e} \|q^{(i)} - q^o\| - \frac{1}{2n_e^2} \sum_{i=1}^{n_e} \sum_{j=1}^{n_e} \|q^{(i)} - q^{(j)}\| \quad (104)$$

#### 7.5 Variogram score - p

The energy score can provide an evaluation of the uncertainty of a multivariate ensemble. Nevertheless, this score is often not sufficiently sensitive to misspecification of the correlations structures of the multivariate observation and the multivariate ensemble. Those correlations cannot be study by univariate proper score unlike the means and variance. Thus, it does not evaluate the capacity of the ensemble model to reproduce the correlations structures in time and/or space of the observation.

One tool used to study these correlations is the variogram of order  $p$  (also called structure function). This tool consists in studying the pairwise differences of the component of the multivariate ensemble forecast and observation by considering the statistics of our multivariate observation. Following (Scheuerer and Hamill, 2015), the p-variogram score (VS-p) is an alternative

solution which overcomes this drawback. We express it using the same multivariate random variable  $q$  and observation  $q^o$ :

$$VS^p(F, q^o) = \sum_{i=1}^N \sum_{j=1}^N \frac{1}{w_{ij}} (|q_i^o - q_j^o|^p - \mathbb{E}_F |q_i - q_j|^p)^2, \quad (105)$$

where the  $w_{ij}$  are non negative weights that allows one to emphasize or down-weight pairs of component combinations based on subjective expert decisions.  $p$  is the order of the variogram score.  $N$  represents the number of spatial points or time steps.  $F$  is the multivariate cumulative distribution function of  $q$ . According to (Scheuerer and Hamill, 2015), a typical choice of weight to discriminate time or space dependencies is the inverse distance weight:  $w_{ij} := \frac{1}{\sqrt{|i-j|}}$  and  $p := 0.5$ . The VS- $p$  score measures the dissimilarity between approximations of the variograms of order  $p$  of the observations and forecast over all pairs of the components of our variable of interest.

## 8 Numerical results : Ensemble forecasts verification for SQG dynamics

In this section, we will present new numerical UQ results. As a test case, a simple geophysical fluid dynamics – the stochastic surface quasi-geostrophic (SQG) model – will be considered. After presenting it, several stochastic sub-grid parameterizations will be compared for short-term and for long-term ensemble forecasts. We will focus on SALT-LU schemes, the new random forcing derived from numerical dissipation of section 5.3.2 and random perturbation of initial condition.

### 8.1 Test flows

#### 8.1.1 Deterministic and stochastic surface quasi-geostrophic (SQG) dynamics

From here, the field  $q = b$  will denote the buoyancy. It is proportional to the density anomaly  $\rho'$

$$b \triangleq -g \frac{\rho'}{\rho_0} \text{ with } \rho(x, y, z, t) = \rho_0 \left( 1 - \frac{N^2}{g} z \right) + \rho'(x, y, z, t), \quad (106)$$

where  $\rho$  is the density and  $N$  the stratification. In the ocean, the density anomaly is small compared to the total density (Boussinesq approximation) and the flow is approximately isochoric. The conservation of salinity and temperature, with a linearized equation of state, provides the transport of buoyancy (Vallis, 2006). Then, considering rapid rotation, strong stratification and uniform potential vorticity leads to the so-called SQG model (Blumen, 1978; Held et al., 1995; Lapeyre, 2017).

A random version of this model (denoted  $SQG_{MU}$ ) can be derived from the location uncertainty principle (Resseguier et al., 2017b). It will keep the same structure except that the buoyancy is now transported in the stochastic sense (14). The horizontal velocity  $\mathbf{u} = \mathbf{w}$  is related to the buoyancy  $b$  in Fourier space through the usual SQG relation:

$$\hat{\mathbf{u}} = i\mathbf{k}^\perp \frac{\hat{b}}{N\|\mathbf{k}\|}, \quad (107)$$

where  $\mathbf{k}$  is the horizontal wave-vector. The unresolved velocity,  $\sigma\hat{\mathbf{B}}$ , is also horizontal. Consequently, the variance tensor,  $\mathbf{a}$ , is a  $2 \times 2$  matrix and the  $SQG_{MU}$  model is two-dimensional.

### 8.1.2 Our SQG simulations

High-resolution deterministic SQG simulations of test flows will provide references to which we will compare random simulations performed at a lower resolution. For this purpose, the high-resolution ( $512^2$ ) simulation outputs will be projected onto the space of low-resolution ( $128^2$ ) fields, i.e. it will be adequately filtered and subsampled to a low-resolution. We will refer to this projected fields as "observations" and denote them  $b^o$ . For all simulations, a standard hyperviscosity (HV) scheme has been introduced (Held et al., 1995):

$$D_t b = \alpha \Delta^4 b \, dt, \quad (108)$$

with a positive coefficient  $\alpha$  proportional to  $M_x^{-8}$  where  $M_x$  denotes the grid size (i.e. 128 or 512). The domain size is a square box  $L_x \times L_y = 1000 \text{ km} \times 1000 \text{ km}$  and the boundary conditions are doubly periodic.

Several simulations of these models have been performed based on two types of initial fields and on various parameterisations of uncertainty. A first type of initial buoyancy field is shown in figure 2 (a). As in Resseguier et al. (2017b), this field consists of a spatially smooth buoyancy field with two warm elliptical anticyclones (positive buoyancy) and two cold elliptical cyclones (negative buoyancy) given by:

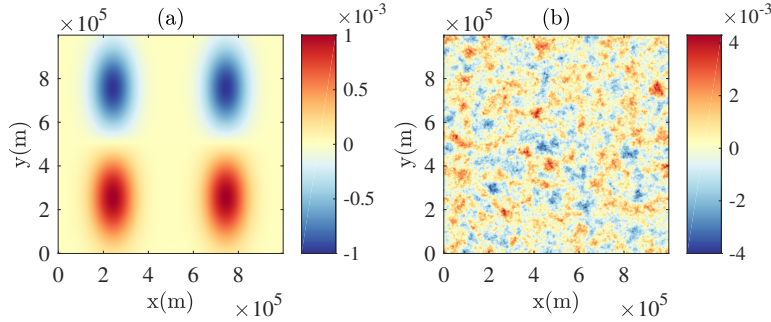
$$b_0(\mathbf{x}) \triangleq F\left(\mathbf{x} - \begin{pmatrix} 250 \text{ km} \\ 250 \text{ km} \end{pmatrix}\right) + F\left(\mathbf{x} - \begin{pmatrix} 750 \text{ km} \\ 250 \text{ km} \end{pmatrix}\right) \\ - F\left(\mathbf{x} - \begin{pmatrix} 250 \text{ km} \\ 750 \text{ km} \end{pmatrix}\right) - F\left(\mathbf{x} - \begin{pmatrix} 750 \text{ km} \\ 750 \text{ km} \end{pmatrix}\right), \quad (109)$$

with

$$F(\mathbf{x}) \triangleq B_0 \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \text{ and } \begin{cases} \sigma_x = 67 \text{ km}, \\ \sigma_y = 133 \text{ km}. \end{cases} \quad (110)$$

The amplitude of the buoyancy and the stratification are set to:  $B_0 = 10^{-3} m \cdot s^{-2}$  and  $N = 3f_0$ . The Coriolis frequency  $f_0$  is fixed to  $1.028 \times 10^{-4} s^{-1}$ , which corresponds to a latitude of  $45^\circ$ .

Another type of initial buoyancy field is shown in figure 2(b), which is a homogeneous Gaussian random field generated from a spectrum of buoyancy with a prescribed slope equals to  $-5/3$ . This slope corresponds to the power law of a developed SQG turbulence.



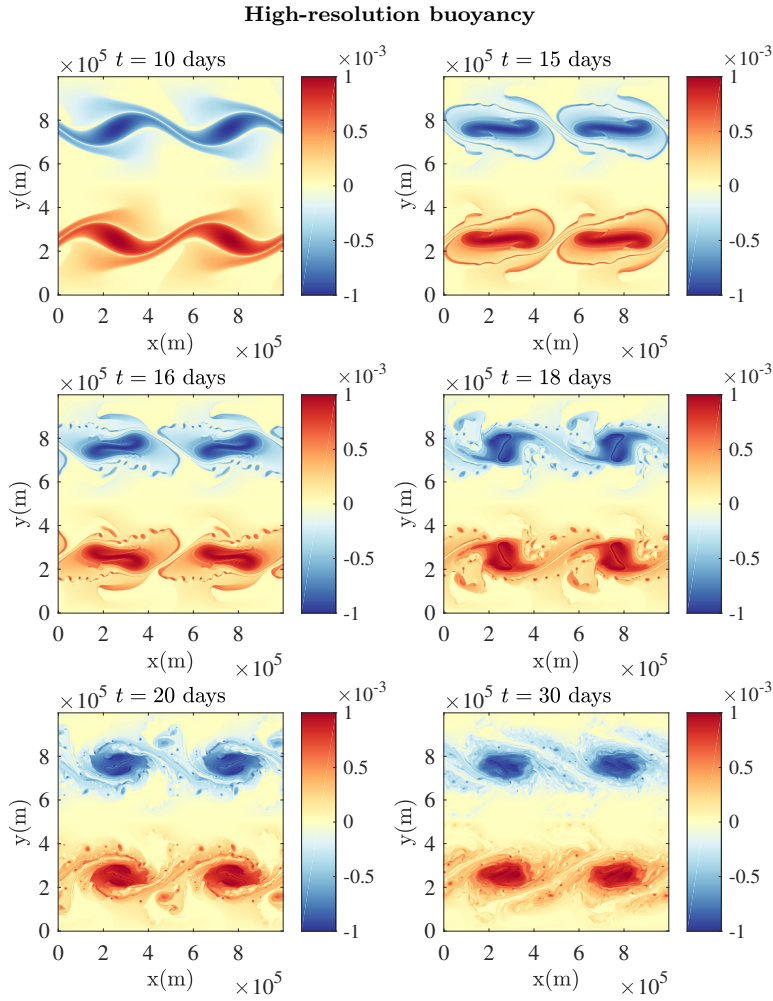
**Fig. 2** Two initial buoyancy ( $m.s^{-2}$ ) fields : (a) – A smooth field with four given vortices (cyclones in blue and anti-cyclones in red); (b) – A homogeneous Gaussian field generated from a  $-5/3$ -spectrum.

We will now forecast the initial buoyancy fields of figure 2 with several ensembles of simulations and assess the performance of each of these ensembles. Short term and long term forecasts will be treated separately because they are very different in nature.

## 8.2 Short-term ensemble forecast

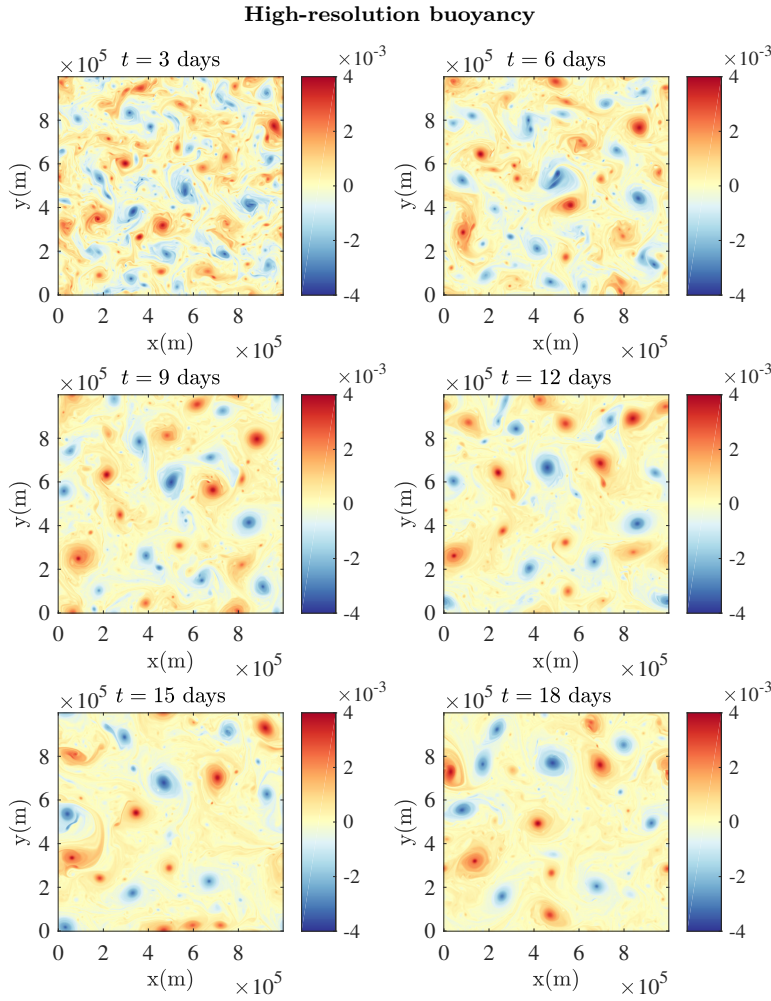
As in the study Resseguier et al. (2017b), we first focus on the first-month forecast of the smooth initial field (figure 2(a)). Figure 3 shows the reference high-resolution simulation for this first month. After two weeks, filament instabilities (Lapeyre, 2017) create developed turbulence. Here, we also study the free-decaying SQG turbulence flow initiated by the rough field (figure 2(b)). The free-decaying turbulence can be seen in figure 4. A part of the initial energy is dissipated by the deterministic subgrid tensor. But, a part of this initial energy cascades to the larger scales by creating larger vortices from the merging of small vortices. Each ensemble contains 30 realizations.

Several low-resolution ensembles have been forecast. Two ensembles rely on initial conditions random perturbations, three on dynamics under location uncertainty and five on the dissipation-adapted noise (see section 5.3.2). Specifically, the small-scale velocity in the dynamics under location uncertainty and the random perturbation of initial condition in  $SQG_{PIC}$  have been both generated using three approaches: a homogeneous stationary model (denoted “*Spectral*”), a homogeneous non-stationary model (denoted “*ADSD*”) and a heterogeneous non-stationary model (denoted “*SVDpseudo*”). For the



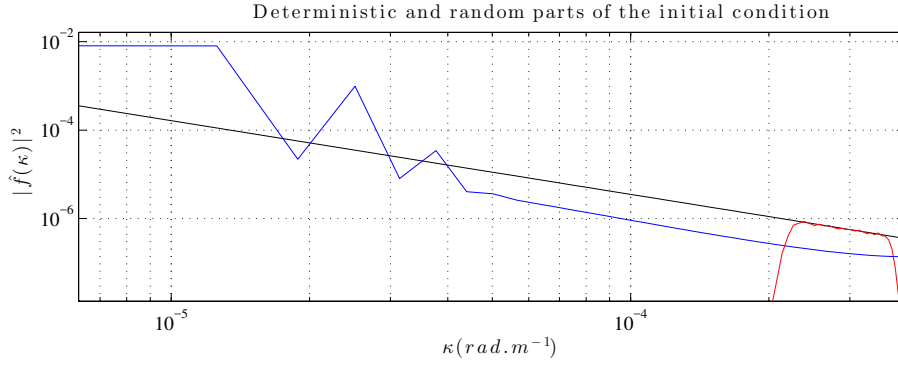
**Fig. 3** Buoyancy ( $m \cdot s^{-2}$ ) at  $t = 10, 15, 16, 18, 20, 30$  days of advection for the usual SQG model at resolution  $512^2$ , based on the smooth initial field – figure 2(a).

model under location uncertainty, it corresponds to parameterizations 6.3.2, 6.3.3 and 6.3.10 respectively, of section 6.3. For the “*Spectral*” method, the small-scale energy is specified by the diffusion coefficient  $\frac{\alpha}{2} = 9 m^2 \cdot s^{-1}$ . We also need to choose the spectrum slope and the minimum and maximum wavelengths of the unresolved velocity. For more, information on the values of these parameters, the reader can refer to Resseguier et al. (2017b). In contrast, for the improved “*ADSD*” method, we do not need to choose any parameter. The parameterization must adapt itself using large-scale informations at each time step. Despite the difference of flow and resolution, the very same code – without any modification – is used here and in Resseguier et al. (2019) to



**Fig. 4** Buoyancy ( $m.s^{-2}$ ) at  $t = 3, 6, 9, 12, 15, 18$  days of advection for the usual SQG model at resolution  $512^2$ , based on the random initial field – figure 2(b).

define the “ADSD” small-scale velocity. For the ensembles with perturbed initial conditions, similar sampling methods are adopted. But the obtained “small-scale” random fields are used only at the initial time. As an example, figure 5 shows the initial perturbation spectrum used in the “Spectral” method for the first initial condition (109). Then, it is added to the initial condition  $b(\mathbf{x}, t = 0)$ . Besides, the dissipation-adapted noise 5.3.2 has been implemented in *SQG<sub>MU</sub>MATLAB* code. For this method – denoted “WavHypervis” – we use a wavelet basis  $(e_k)_{k \in \mathbb{Z}}$ . We have forecast 5 ensembles with 5 different values of the scaling factor  $\zeta$ . For instance, figure 6 shows us that under the noise the same low-resolution, when the scaling factor  $\zeta$  increases (i.e. when the noise



**Fig. 5** Kinetic energy spectrum ( $m^2.s^{-4}/(rad.m^{-1})$ ), at the initial time, of the mean buoyancy, in blue, spectrum of its random perturbation, in red, and slope  $-\frac{5}{3}$  in black. The initial perturbation is restricted to a narrow spectral band. This random initial condition has been used to simulate an ensemble with the deterministic SQG model.

variance contains a larger part of the numerical dissipation), more and more small-scale structures are presented in physical field.

Once the ensembles have been produced by the previous models, we try to measure the quality of ensemble forecasts by some easy-to-implement criteria. We have  $N_e$  ensemble members for each forecast with  $b^{(i)}$  denotes the  $i$ -th ensemble member of the forecast. A first necessary condition for reliability is that the mean squared bias (MSB) of the ensemble (i.e. the MSE of the ensemble mean) is close to the mean intra-ensemble variance (MEV), up to an ensemble size-dependant scaling factor :

$$MSB \approx \frac{N_e + 1}{N_e} MEV, \quad (111)$$

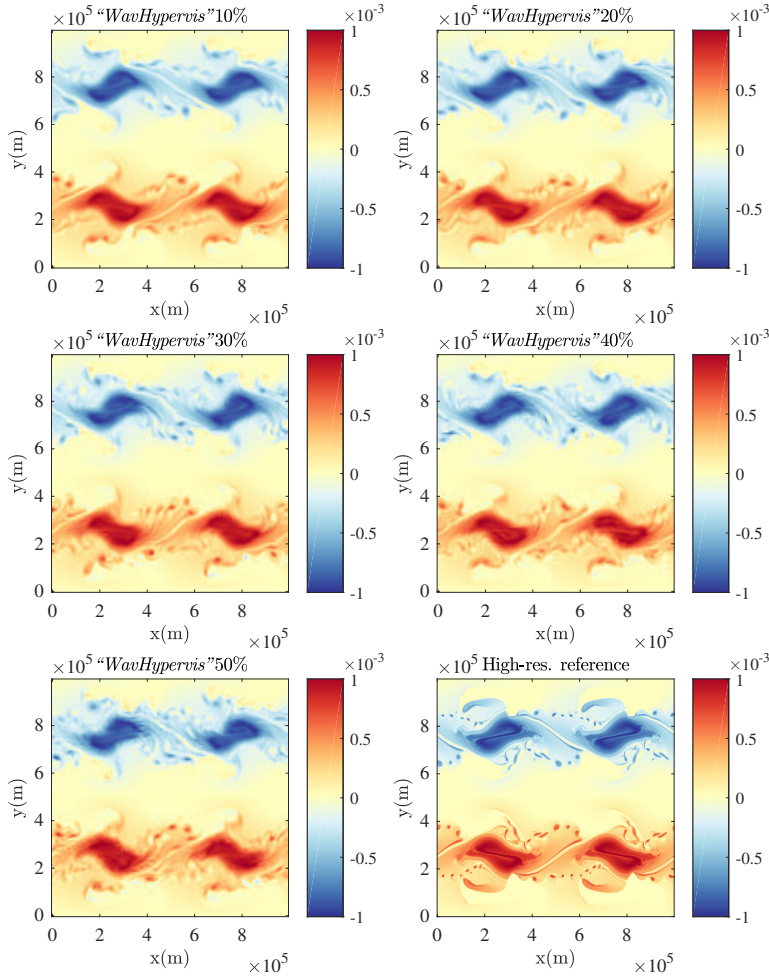
where

$$MSB \triangleq \frac{1}{M} \sum_{j=1}^M (b^o - \widehat{\mathbb{E}}b)^2(t, \mathbf{x}_j), \quad (112)$$

$$MEV \triangleq \frac{1}{M} \sum_{j=1}^M \widehat{Var}(b)(t, \mathbf{x}_j), \quad (113)$$

$M$  the number of grid points  $\mathbf{x}_j$ ,  $\widehat{\mathbb{E}}\{f\}$  and  $\widehat{Var}(f)$  the empirical mean and the empirical variance of  $f$ , computed from the ensemble, respectively. At the same time, a classical error metric – the ensemble mean square error (MSE) is also considered :

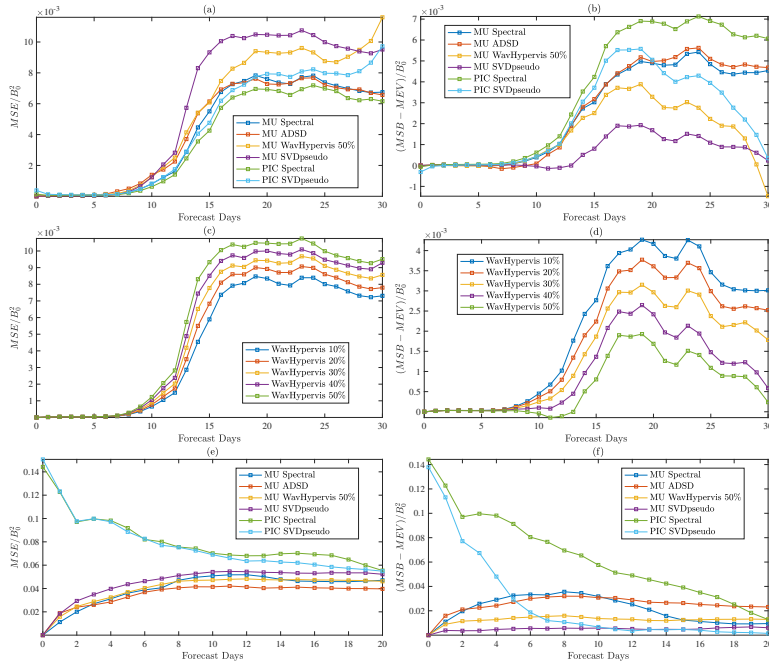
$$MSE(t) = \widehat{\mathbb{E}} \left( \frac{1}{M} \sum_{j=1}^M (b^o - b)^2(t, \mathbf{x}_j) \right). \quad (114)$$



**Fig. 6** Comparison of one realisation of the buoyancy field after 17 days of advection, according to various energy scaling factor in “*WavHypervis*”.

In figures 7(a) and 7(b), we compare the criterions (111) and (114) produced by our models. For the 4 vortices field 2(a), it seems that for all models produce slight errors during the first 10 days. Their forecast skills tend to a stationary state after about 17 days’ advectons. In this case, *WavHypervis* and *SVDpseudo* under location uncertainty have a better ensemble spread, yet conversely the homogeneous models – *Spectral* and *ADSD* have lower MSE. As illustrated in figures 7(c) and 7(d), in the *WavHypervis* method, when the factor  $\zeta$  increases, both the spreading and the ensemble errors increase. In figures 7(e) and 7(f), we compare the results with the spectral initial field 2(b), in which the ensembles models tend to a stationary state much more faster. In this case, *WavHypervis* and *SVDpseudo* under location uncertainty have still



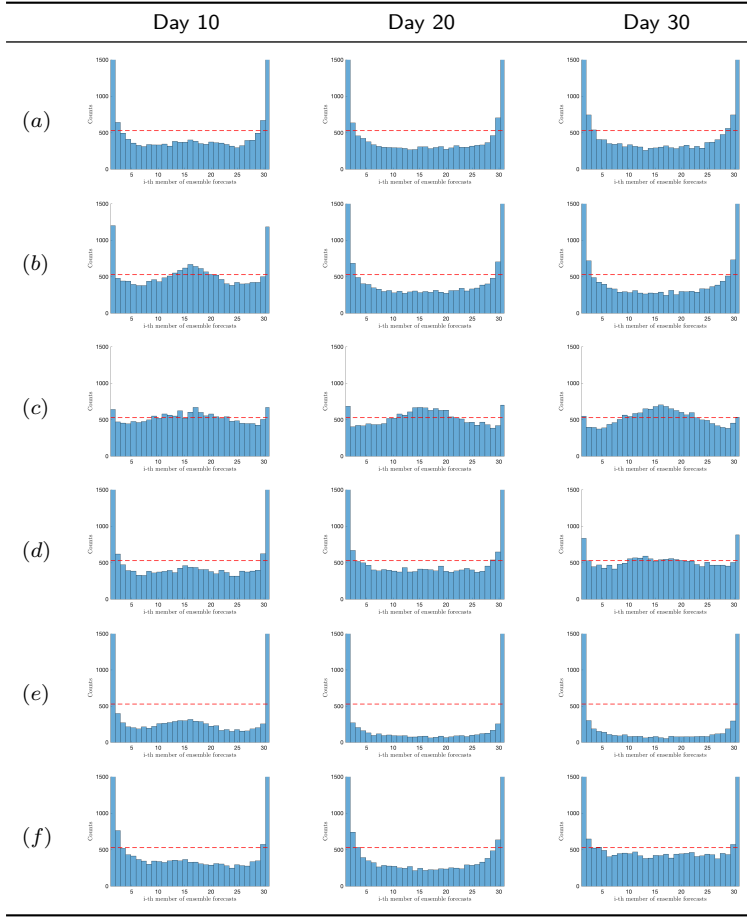


**Fig. 7** Comparisons of ensemble forecasts for different stochastic models based on the two tested initial conditions : (a) – The mean square error produced by different models under the background of 4 vortices; (b) – The corresponding difference between mean square bias and mean ensemble variance; (c), (d) – Using the empirical noise (5.3.2) with wavelet basis and various energy scaling factors, under the background of 4 vortices; (e), (f) – Results under the spectral background. Notice that all these results are normalized by the amplitude  $-B_0$  of the referent initial buoyancy fields.

better ensemble spread. The forecasts by *Spectral* and *ADSD* consist of less errors in ensemble.

Another intuitive estimation of ensembles dispersions have been performed by Talagrand diagram. As shown in table 1, under the 4 vortices background, the small perturbations method applied on the initial condition is too underdispersive. Instead, *Spectral* and *ADSD* under location uncertainty formed a U-shape obtaining underdispersion and so higher but still underestimated variance. *SVDpseudo* under location uncertainty estimates high uncertainty. Instead, *WavHypervis* with 50% energy allocated from the numerical dissipation provides a slight overdispersion at last. This is probably due to a premature bifurcation in this ensemble (see section 8.3 for more details about the bifurcation). From table 2, with the spectral initial field, *WavHypervis*, and *SVDpseudo* under initial perturbations provide almost perfect ensemble spread. The ensemble skills converge toward a stationary state much faster for this flow.

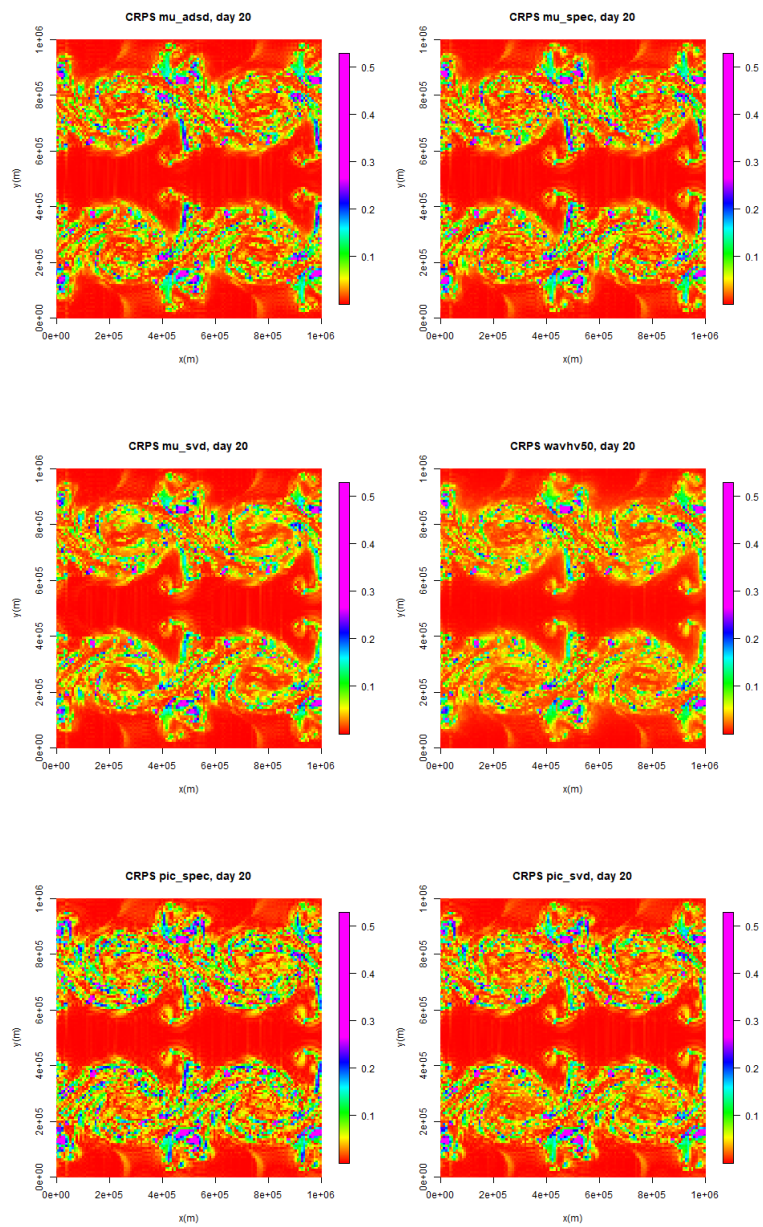
As explained above, distinct models yield distinct spreading based criterions (111), (114) and Talagrand diagram. Nonetheless, these two criterions



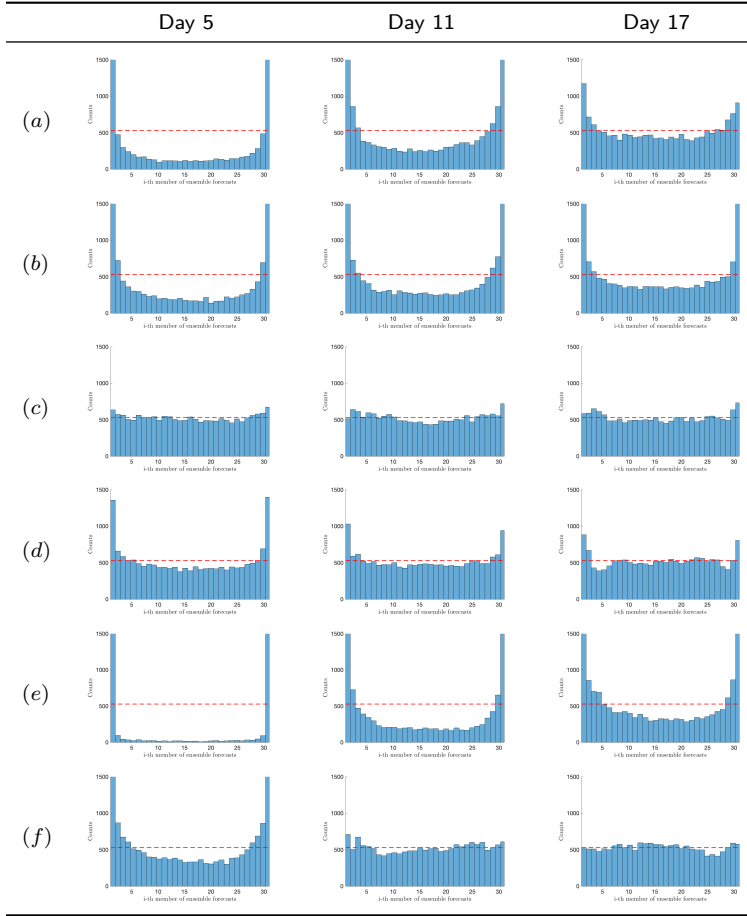
**Table 1** Talagrand diagram on some forecast days under the background of 4 vortices : (a) – MU Spectral; (b) – MU ADSD; (c) – WavHypervis 50%; (d) – MU SVDpseudo; (e) – PIC Spectral; (f) – PIC SVDpseudo.

type provide opposite conclusions. An objective analysis need hence others metrics like proper score (see section 7).

To begin, we will compare the numerical results with the proper score CRPS (see the section 7.3). The CRPS is defined as a pointwise score. Applied on ensemble of spatio-temporal fields, maps of CRPS can be represented at each fixed time step as in the figure 8. As expected on all maps of the figure 8, the normalized CRPS is relatively high on small scale structures and low on the center of each vortices. Indeed, turbulent structures are the hardest things to reproduced. Moreover, the figure 8 suggests that the model *Spectral* in  $SQG_{PIC}$  is the worst model since it has the largest CRPS values. Model *WavHypervis* in  $SQG_{MU}$  shows smaller CRPS. Thus reveals that this model makes fewer local errors at this advection time.



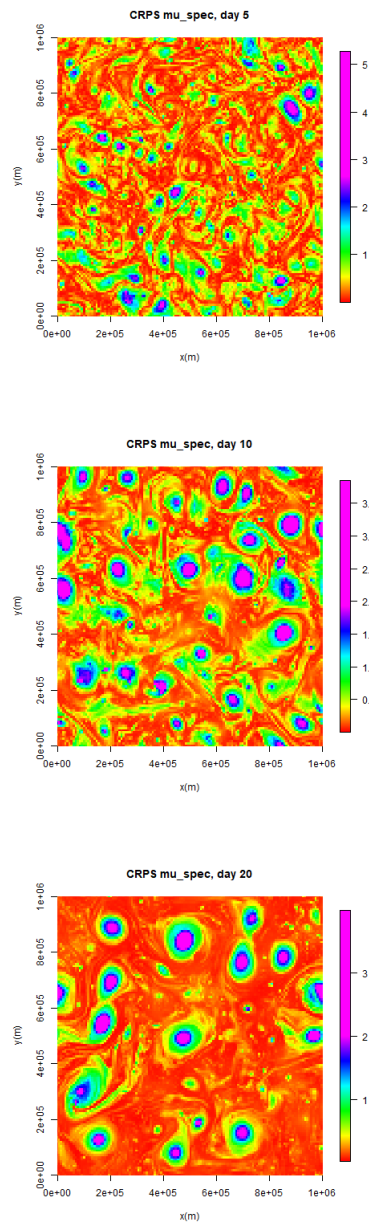
**Fig. 8** CRPS of each model after 20 days of advection with an initial condition based on a smooth field with four given vortices



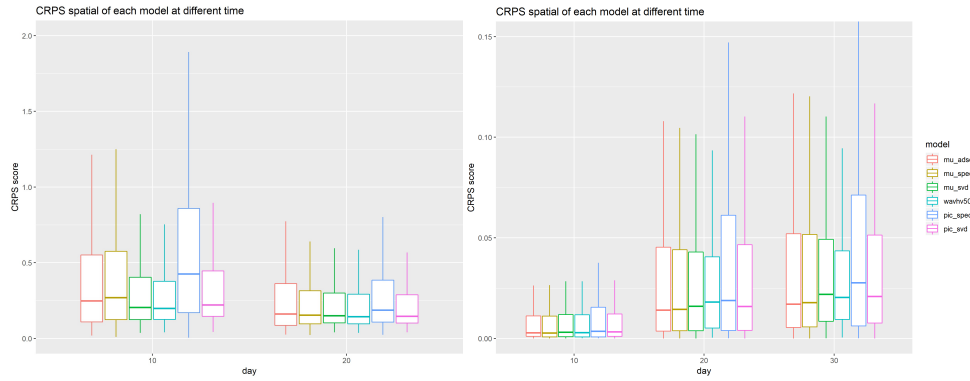
**Table 2** Talagrand diagram on some forecast days under the spectral background : (a) – MU Spectral; (b) – MU ADSD; (c) – WavHypervis 50%; (d) – MU SVDpseudo; (e) – PIC Spectral; (f) – PIC SVDpseudo.

The CRPS has only been computed at one fixed time. So, again, it is difficult to conclude on which model is better.

To push further the analysis, we maps at different times. The figure 9 reveals a part of the spatial evolution of those CRPS maps for the model *Spectral* in  $SQG_{MU}$ . With the random gaussian field as initial condition, we can see on the figure 9 that the CRPS structures are first small and spread over the spatial domain. Then, these small scale CRPS structures merge and create larger structures of larger intensity. The merging is due to the inverse energy cascade of SQG turbulence (Lapeyre, 2017) which merges the badly resolved small-scale turbulence structures. On the top of that, the new large vortices have chaotic trajectories. The difficulty for the models to track these trajectories yield large CRPS values in the centers of those vortices.



**Fig. 9** Normalized CRPS of the model mu spectrum during 5,10 and 20 days of advection with an initial condition based on an initial random gaussian field.



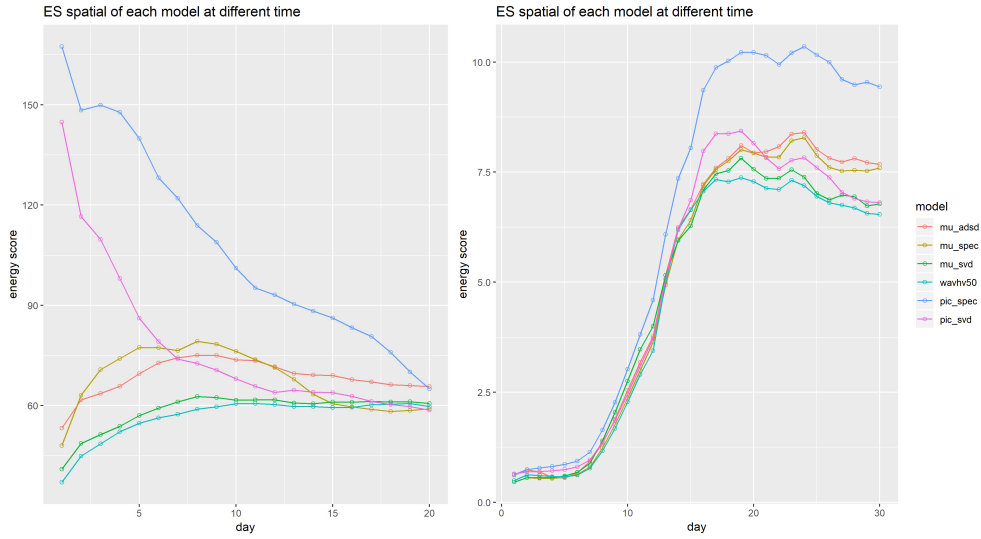
**Fig. 10** Box plot of the normalized CRPS computed on the ensemble at each spatial points of each model at different time step of advection for two different initial condition (left: Random gaussian field, right: Four given vortices).

Our analysis describes the variability of the CRPS in time and space. But, the difficulty to select a model remains to summarize the CRPS information.

The figure 10 summarizes model performances by representing each CRPS by one box plot. As such, we can directly see the variability of CRPS values over one CRPS map. For the four vortices initial condition, the model *Spectral* in *SQG<sub>PIC</sub>* has the most spread CRPS. Moreover, at 30 days of advection, it has the largest CRPS mean. Therefore, the figure 10 confirms that the model *Spectral* in *SQG<sub>PIC</sub>* has the worst uncertainties quantification skill. The models *Spectral* in *SQG<sub>MU</sub>* and *ADSD* in *SQG<sub>MU</sub>* obtain the lowest CRPS mean and spread at 20 and 30 days of advection. However, it is difficult to choose between these two models. Their CRPS distributions are too close. On the left panel of the figure 10 spectral background, it is easier to select a model. Indeed, at 10 days of advection, the model *Spectral* in *SQG<sub>PIC</sub>* gets the lowest CRPS mean and at 20 days the model *ADSD* in *SQG<sub>MU</sub>* has the lowest mean. Nevertheless, it is still difficult to conclude about the most efficient model. Boxplots of CRPS give statistical description about this proper score. But, it does not analyze the multivariate structure of the error between the ensemble and the reference.

The figure 11 plots the normalized energy scores (see 7.4) of advection times. There, the analysed multivariate structure is the spatial structure of the random fields. The figure 11 shows that the model *Spectral* in *SQG<sub>PIC</sub>* has the largest energy score for both initial conditions. So, this model provides the worst uncertainties quantification with multivariate spatially ensemble.

The model *SVDpseudo* in *SQG<sub>PIC</sub>* has a low energy score after two weeks of advection for the spectral background and after three weeks for the four-vortice initial condition. With the spectral background, after two weeks large-scale structures has formed and the model *SVDpseudo* in *SQG<sub>PIC</sub>* makes less error with the reference. Nevertheless, during the first week of advection,



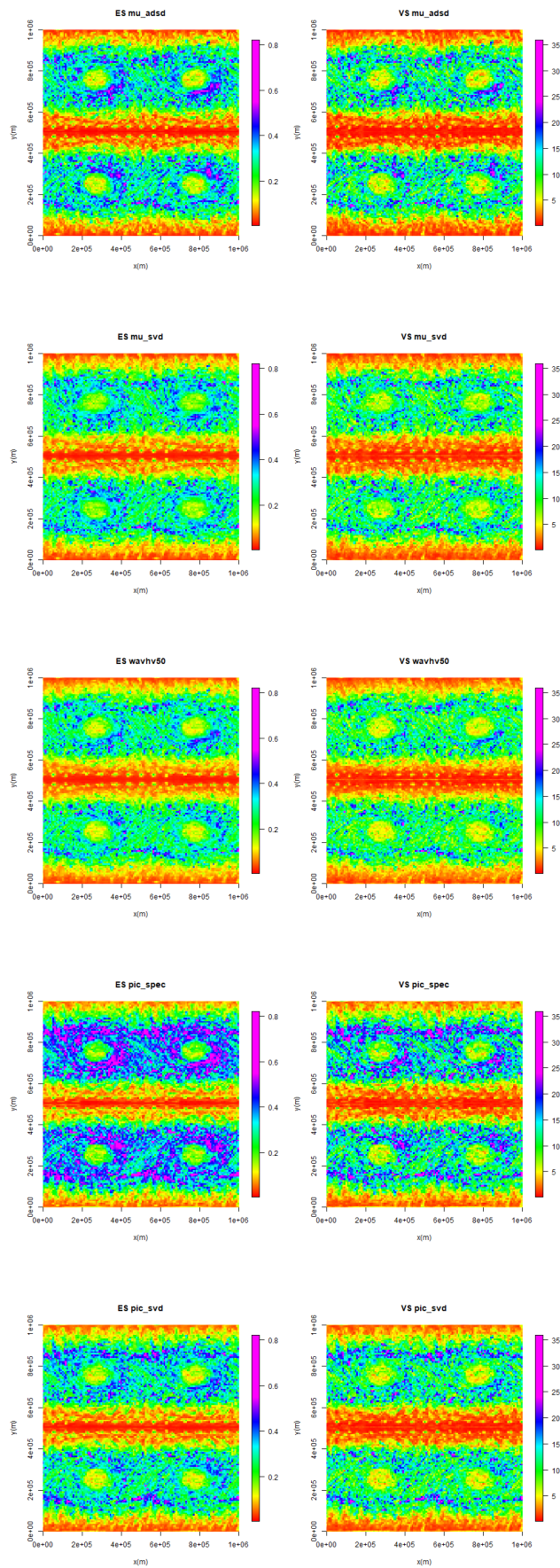
**Fig. 11** Normalized energy score for each model and spatial multivariate ensemble on all time steps and both initial condition (left: Random gaussian field, right: Four given vortices).

the model  $SVDpseudo$  in  $SQG_{PIC}$  has a really large energy score because, small-scale structures are present and are not well reproduced by  $SVDpseudo$  in  $SQG_{PIC}$ . So, in this case, the models  $SQG_{PIC}$  are less efficient than the models  $SQG_{MU}$  and  $WavHypervis$ .

For the four-vortice initial condition, the model  $SVDpseudo$  in  $SQG_{PIC}$  has a larger energy score at 16 to 20 days of advection. During this period of advection, some small-scale structures are produced and well resolved by the simulations. After this period, the energy score of  $SVDpseudo$  in  $SQG_{PIC}$  decreases probably due to the bifurcation phenomenon described in section 8.3.

The figure 11 suggests that models  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  make the least error. Indeed, during the most important times steps and for both initial condition, these models show the lowest energy scores. In particular, the energy score of the  $WavHypervis$  method reveals that it is the most efficient model according to this metric.

We will now consider proper scores and variograms where the temporal structures are considered instead of the spatial ones. The figure 12 reflects that the model  $Spectral$  in  $SQG_{PIC}$  has the largest normalized energy score and variogram-0.5 score (see section 7.5). On the edges of the centers of the vortices, this model encompasses the worst temporal structures. The models  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  seem to perform better. Nonetheless, in the model  $Spectral$  in  $SQG_{MU}$ , minimum values are located in the center of the vortices which indicates that the temporal evolution of the vortices is well reproduced by this model.



**Fig. 12** Normalized energy score and variogram-0.5 score of the temporal multivariate ensemble of each models at each spatial points with an initial condition based on the four vortices (left column: normalized energy score of each model, right: normalized variogram-0.5 score of each model).



Model	MSE	MSB	CRPS	ES	VS
$SQG_{MU_{SPEC}}$	+	+	+	+	+
$SQG_{MU_{ADSD}}$	++	+	+	+	+
$SQG_{MU_{SVD}}$	-	++	++	++	++
$WavHypervis_{50}$	-	+	++	++	++
$SQG_{PIC_{SPEC}}$	-	--	--	--	-
$SQG_{PIC_{SVD}}$	--	-	+	+	+

**Table 3** Model performance by score, MSE: Mean squared error (pointwise), MSB: Mean squared bias (pointwise), CRPS: Continuous ranked proper score (pointwise), ES: Energy score (spatial and time).

### 8.2.1 Conclusion of the short-term forecasts analysis

For this numerical study, ensemble forecast skills have been assessed through a set of verification tools. The table 3 summarizes validation scores estimated for each UQ model. Talagrand histogram is an evaluation of the calibration behavior of an ensemble, while proper scores focus on dispersion and errors between univariate or multivariate ensembles and references. Each verification score has his own specificity, it is essential to be aware of their properties to select the most suitable measure (see section 7). Also, to avoid misselection of ensemble methods, the assessment of a number of scoring rules is advised.

In the table 3,  $SQG_{PIC}$  methods obtained the lowest overall verification scores. In contrast,  $ADSD$  in  $SQG_{MU}$ ,  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  models present the best performances. Therefore, we recommend one of this method for UQ tasks, and we strongly advice to avoid relying only on initial conditions randomization.

The model  $ADSD$  in  $SQG_{MU}$  has the best overall MSE, meaning that it introduces less errors than other UQ methods. The model  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  get lower MSE performances. Yet, their CRPS, ES and VS are the best among all the methods. This illustrates the accurate univariate and multivariate UQ produced by  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  algorithms.  $WavHypervis$  model reveals a weaker MSB score analysis than  $SVDpseudo$  in  $SQG_{MU}$  induced by a lack in the ensemble spread. For this reason, if one UQ method has to be chosen among  $ADSD$  in  $SQG_{MU}$ ,  $SVDpseudo$  in  $SQG_{MU}$  and  $WavHypervis$  models, it would probably be  $SVDpseudo$  in  $SQG_{MU}$ . Nevertheless, the higher UQ skills of this method also come with a slightly higher CPU time.

This numerical study have characterized UQ skills of short-term ensemble forecasts. To complete the analysis, we now focus on long-term forecast.

### 8.3 Long-term ensemble forecast

After the first month of advection, the flow initiated with (109) (figure 3) breaks its symmetries. This leads to a chaotic behavior and a complete loss

of (deterministic) predictability. In this section, we will focus on this regime. As demonstrated here, the bifurcation associated with the symmetry breaking is efficiently tracked using the model under location uncertainty. Using few realizations, the probability density functions of each subsequent scenarios is well characterized, whereas methods based on random initial conditions do not converge. The identification of several scenarios, done at each time step in a reduced subspace, is obtained by Principal Component Analysis (PCA), also termed Empirical Orthogonal Functions method (EOF).

Finally, a diagnosis of bifurcation is performed and discussed from both the model under location uncertainty and a method based on randomized initial conditions.

### 8.3.1 Chaotic test flow and resolution issues

The boundaries conditions of the simulation are doubly periodic and, for the initial condition (109), there is a meridional line of symmetry at  $x = 500$  km. Therefore, the zonal period of the initial condition is  $L_x/2 = 500$  km. This periodicity is relatively stable and holds during the first month. Nevertheless, the SQG dynamics, is subject to an inverse cascade of energy (Capet et al., 2008), and vortices of the same sign tend to merge. When this merging occurs, this affects the global shape of the flow. In particular, the periodicity that remains in the first month eventually disappears. This symmetry is hence metastable rather than stable. The symmetry breaking corresponds to a transition from one “state” to another. By “state”, we mean a relatively “compact” and connected subspace of the state space. Warm vortices can merge at  $x = 0$  or at  $x = 500$  km. In the following, we will refer to the first case as “scenario 1”, and to the second case as “scenario 2”. Because of the periodic boundary conditions, these two possible transitions are likely to occur. In a deterministic numerical simulation, the appearance of one transition or the other is determined by an infinitesimal asymmetry in the initial condition or possibly by a numerical error. This is a bifurcation. The bifurcation related to the merging of cold vortices is similar. With those two simultaneous bifurcations, there are thus  $2 \times 2 = 4$  likely transitions.

To trigger a particular transition, we introduce two infinitesimal modifications in the initial condition. For the sake of simplicity, we focus on the bifurcation associated with the warm vortices. To do so, the merging of cold vortices in  $x = 500$  km will be forced, by adding an infinitesimal small-scale cold eddy in  $(x, y) = (480, 750)$  (in km). The barycentre of northern structures becomes slightly closer to  $x = 500$  km. This gives rise to the desired transition, as shown in Figures 13 and 14. To trigger the bifurcation associated with the southern warm vortices, an infinitesimal small-scale eddy has been added in  $(x, y) = (20, 250)$  (in km). If the eddy is warm, the southern barycentre is moved closer to  $x = 0$ , and the two warm vortices merge near  $x = 0$  (scenario 1), as shown in Figure 13. If this eddy is cold, the southern barycentre is moved closer to  $x = 500$  km, and the two warm vortices merge near  $x = 500$  km (scenario 2), as shown in Figure 14. The exact expression of the initial

condition is the following:

$$b(\mathbf{x}, t = 0) = b_0(\mathbf{x}) + 0.3s_w F \left( \frac{1}{40} \left( \mathbf{x} - \begin{pmatrix} 20 \text{ km} \\ 250 \text{ km} \end{pmatrix} \right) \right) - 0.3F \left( \frac{1}{40} \left( \mathbf{x} - \begin{pmatrix} 480 \text{ km} \\ 750 \text{ km} \end{pmatrix} \right) \right), \quad (115)$$

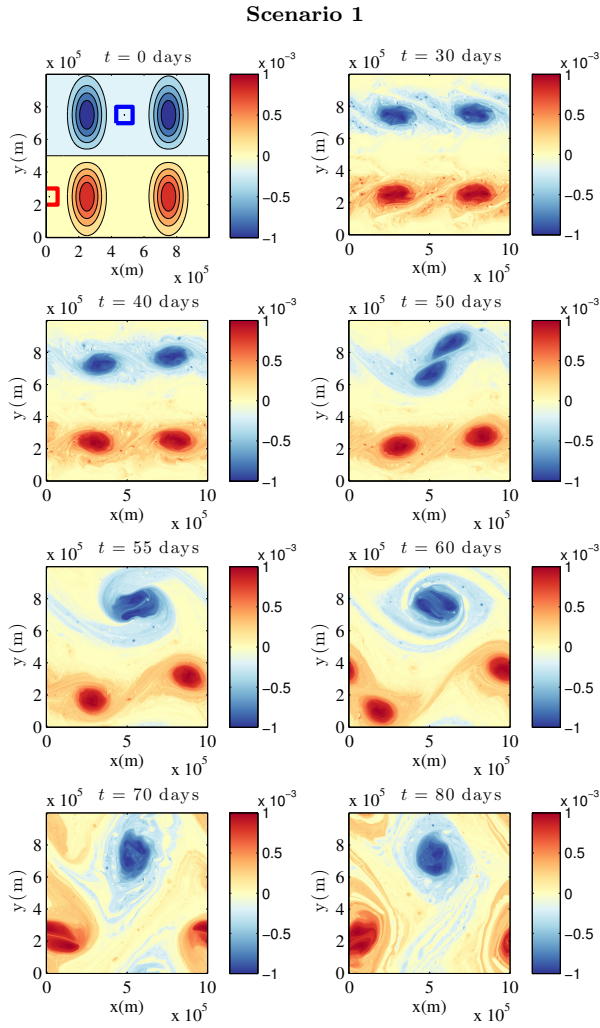
where the large-scale field,  $b_0$ , and the two-dimensional Gaussian function,  $F$ , remain defined by (109) and (110). The factor  $s_w$  is set to 1 (respectively  $-1$ ) if one wants to force the scenario 1 (respectively the scenario 2). The size of the large vortices of  $b_0$  is of the order of the Rossby radius  $L_d$ , whereas the small-scale eddies spread only over few kilometers.

The high-resolution simulations corresponding to scenario 1 and 2 are displayed in Figures 13 and 14, respectively. At this resolution, the evolution toward scenario 1 or scenario 2 is determined by the value of the parameter  $s_w$  (*i.e.*  $+/-1$ ). The associated variations are hardly taken into account by the low-resolved SQG model, as shown in Figure 15. The two SQG simulations correspond to 70 days of advection at high and low resolution, respectively. Both simulations have been initialized in the same way ( $s_w = 1$ ). The low-resolution field differs from the high-resolution field, as it apparently followed the wrong transition. As understood, whatever the random or deterministic nature of the tracer evolution law, a dissipation or a filtering at small scales is necessary to remove aliasing effects. In present simulations, the dissipation is created by an hyperviscosity scheme. At low resolution, the initial perturbation is rapidly diffused (few days). When the symmetry breaking occurs, after 40 days of advection, this initial perturbation has been completely forgotten. Moreover, another infinitesimal asymmetry triggers the other likely transition. According to the expression of  $b_0$  (equations (109) and (110)), the large scale of the initial condition is not exactly zonally periodic with period 500 km. The southern part is slightly warmer in the middle. Indeed, the value of  $b_0$  on  $(x, y) = (500, 250)$  (in km) is about  $1.8 \times 10^{-5} m.s^{-2}$ , and on  $(x, y) = (0, 250)$  (in km) is about  $8.8 \times 10^{-6} m.s^{-2}$ . The initial barycentre of the southern structures is thus closer to  $x = 500$  km. This asymmetry has a very weak amplitude but a large spatial length scale which prevents its diffusion. This explains the merging in the wrong location.

In the next section, we will show that low-resolution simulations of the dynamics under location uncertainty can retrieve the right scenario.

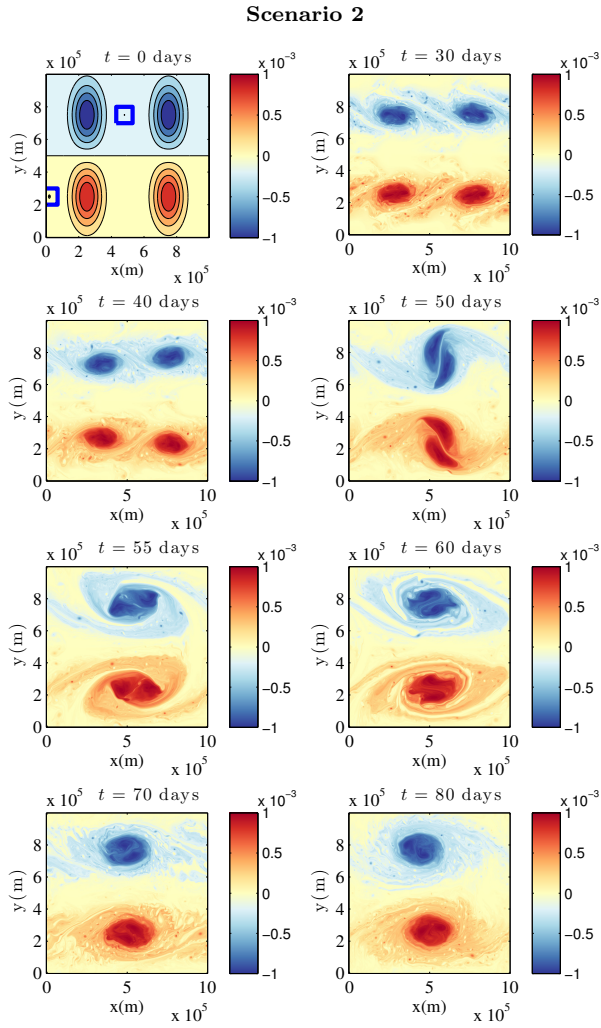
### 8.3.2 Stochastic analysis

Unlike the deterministic SQG model, the *SQG* dynamics under location uncertainty, with the exact same initialization  $s_w = 1$ , yields several likely transitions. Here, for simplicity, only the variant *Spectral* of  $SQG_{MU}$  is considered (neither *ADSD* nor *SVDpseudo*). The calibration *Spectral* of  $SQG_{MU}$  is detailed in section 6.3.2. In Figure 16, we show two realizations of the  $SQG_{MU}$  dynamics. One of those realizations corresponds to the reference scenario (scenario 1), the other does not. The model encodes several likely transitions, and



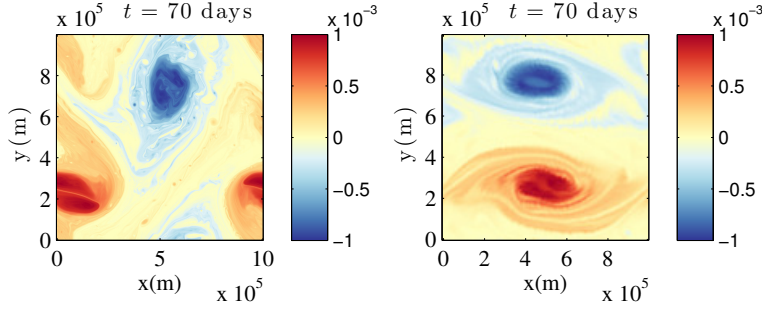
**Fig. 13** Buoyancy ( $m.s^{-2}$ ) at  $t = 0, 30, 40, 50, 55, 60, 70$  and  $80$  days of advection for the SQG model at resolution  $512^2$ . Here, a cold and a warm very small eddies are added respectively in the top and the bottom of the initial condition. These eddies are highlighted by respectively a blue square and a red square. This small difference in the initial condition does not modify the flow until the onset of the symmetry breaking, the  $40^{th}$  day. Since the flow is chaotic, the small perturbation at  $t = 0$  determines how the symmetry breaking occurs a month and a half later.

thus several potential scenarios. Indeed, the random forcing provides various small-scale perturbations that may trigger these transitions. As this triggering is random, the large-scale changes are also random. In other words, there is a backscattering of uncertainty toward the large scales, as illustrated in Figure 17. We decomposed the mean omni-directional spectrum, *i.e.* the mean energy at a given scale,  $\hat{\mathbb{E}}\{\Gamma_b\}$ , into the spectrum of the mean tracer,  $\Gamma_{\hat{\mathbb{E}}\{b\}}$ , (blue



**Fig. 14** Buoyancy ( $m.s^{-2}$ ) at  $t = 0, 30, 40, 50, 55, 60, 70$  and  $80$  days of advection for the usual SQG model at resolution  $512^2$ . Here, two very small cold eddies are added in the top and the bottom of the initial condition. They are highlighted by two blue squares. This small difference in the initial condition does not modify the flow until the onset of the symmetry breaking, the  $40^{th}$  day. Since the flow is chaotic, the small perturbation at  $t = 0$  determines how the symmetry breaking occurs a month and a half later.

line) and the mean spectrum of the tracer random component,  $\hat{\mathbb{E}} \left\{ \Gamma_{b-\hat{\mathbb{E}}\{b\}} \right\}$ ,



**Fig. 15** Buoyancy ( $m.s^{-2}$ ) after 70 days of advection for the SQG model at resolution  $512^2$  (left) and at resolution  $128^2$  (right). The small-scale perturbation in the initial condition activating the symmetry breaking in the reference simulation ( $512^2$ ), is dissipated in few days in the low-resolution simulation. This makes this single low-resolution simulation erroneous.

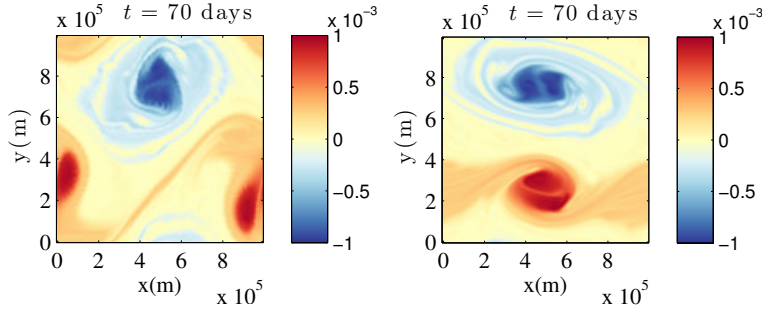
(shaded grey):

$$\hat{\mathbb{E}}\{\Gamma_b\} = \hat{\mathbb{E}} \left\{ \frac{1}{\mu(\Omega)} \oint \|\mathbf{k}\|_2 |\hat{b}|^2 d\theta_{\mathbf{k}} \right\}, \quad (116)$$

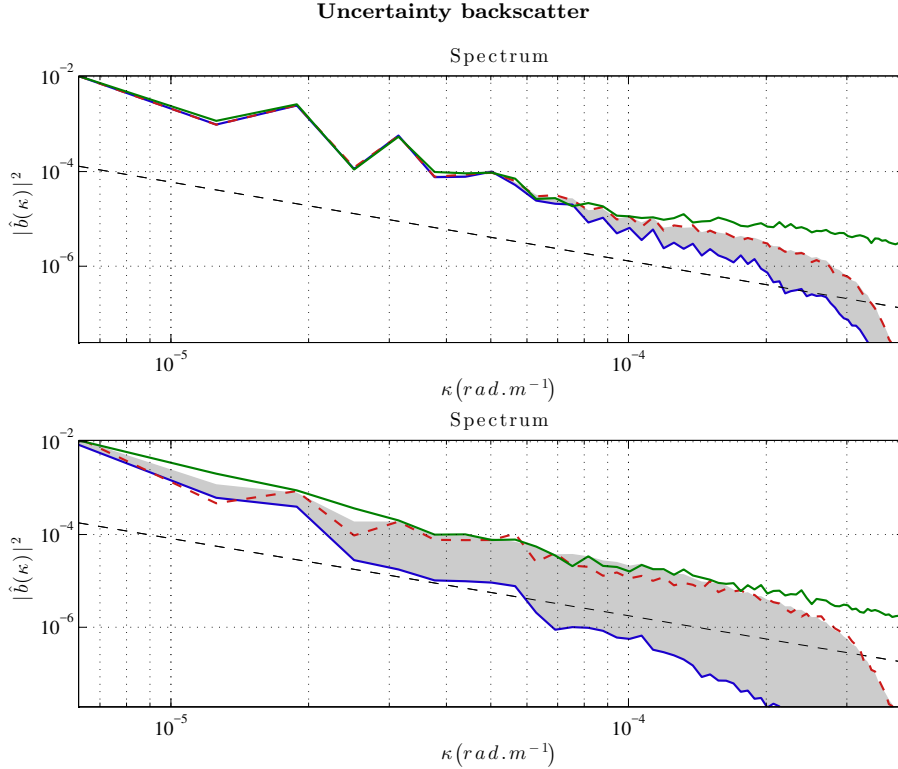
$$= \frac{1}{\mu(\Omega)} \oint \|\mathbf{k}\|_2 \left( \underbrace{|\hat{\mathbb{E}}\{\hat{b}\}|^2}_{\text{Energy of the mean}} + \underbrace{\hat{\mathbb{E}}\{|\hat{b} - \hat{\mathbb{E}}\{\hat{b}\}|^2\}}_{\text{Variance}} \right) d\theta_{\mathbf{k}}, \quad (117)$$

$$= \Gamma_{\hat{\mathbb{E}}\{b\}} + \hat{\mathbb{E}} \left\{ \Gamma_{b - \hat{\mathbb{E}}\{b\}} \right\}, \quad (118)$$

where  $\mu(\Omega)$  denotes the area of the domain  $\Omega$ ,  $\hat{\mathbb{E}}\{f\}$  the empirical mean of  $f$ , computed from the ensemble, and  $\hat{f}$  the Fourier transform of  $f$ .



**Fig. 16** Two realizations of buoyancy ( $m.s^{-2}$ ) after 70 days of advection for the SQG under location uncertainty at resolution  $128^2$ . Event though the small-scale perturbation in the initial condition is dissipated in few days, the small-scale component of the dynamics under location uncertainty randomly triggers the symmetry breaking. Therefore, some realizations follow the right transitions (left) and some do not (right).



**Fig. 17** Tracer spectrum ( $m^2.s^{-4}/(\text{rad}.m^{-1})$ ) after 30 days (top) and 70 days (bottom) of advection for SQG model at resolution  $512^2$  (green), one realization of SQG<sub>MU</sub> model,  $\Gamma_{b^{(1)}}$ , at resolution  $128^2$  (red dashed line), the spectrum of the mean,  $\hat{\mathbb{E}}\{\Gamma_b\}$ , (blue line) and the mean spectrum of the tracer random component,  $\hat{\mathbb{E}}\{\Gamma_{b-\hat{\mathbb{E}}\{b}\}$ , (shaded grey). After being stacked, the two last plots represent the mean spectrum:  $\hat{\mathbb{E}}\{\Gamma_b\} = \Gamma_{\hat{\mathbb{E}}\{b\}} + \hat{\mathbb{E}}\{\Gamma_{b-\hat{\mathbb{E}}\{b}\}$ . The more thick the shaded grey area is, the more variance is contained at this scale. At  $t = 30$  days, the variance remains at small scales but this small-scale uncertainty activates the symmetry breaking. This results in a variance backscattering with a thickening of the spectrum of the random component at large scales (visualized at  $t = 70$ ). Since large scales influence strongly the small scales the small-scale variance is also enforced.

One can then wonder about the number of realizations following the right scenario. This necessitates the analysis of the 4-dimensional spatio-temporal random field  $(b^{(i)}(x_j, y_k, t_l))_{ijkl}$ . The superscript  $(i)$  designates the  $i$ -th realization of the ensemble. To reduce the associated dimension, a Principal Component Analysis (PCA) – also termed Empirical Orthogonal Function (EOF) representation – is performed over the realizations, at a fixed time  $t$ . Within this analysis, unlike usual EOF representation, the time axis is replaced by the realization index. At a given time, it helps to represent the whole ensemble (200 realizations of the random field) by the ensemble mean field and few other

EOF spatial modes. In the present case, the buoyancy is approximated as:

$$b^{(i)}(\mathbf{x}, t) = \hat{\mathbb{E}}(b)(\mathbf{x}, t) + \sum_{n=1}^{N_e} c_n^{(i)}(t) \Psi(\mathbf{x}, t), \quad (119)$$

$$\approx \hat{\mathbb{E}}(b)(\mathbf{x}, t) + \sum_{n=1}^{N_{EOF}} c_n^{(i)}(t) \Psi_n(\mathbf{x}, t), \quad (120)$$

where  $N_e$  is the size of the ensemble,  $N_{EOF} \ll N_e$  is the number of EOF modes chosen to describe the whole ensemble and the  $(\Psi_n)_{1 \leq n \leq N_{EOF}}$  denote the EOF spatial modes. Those spatial fields are orthogonal:

$$\frac{1}{M} \sum_{j=1}^M \Psi_n(\mathbf{x}_j, t) \Psi_m(\mathbf{x}_j, t) = \delta_{nm}, \quad (121)$$

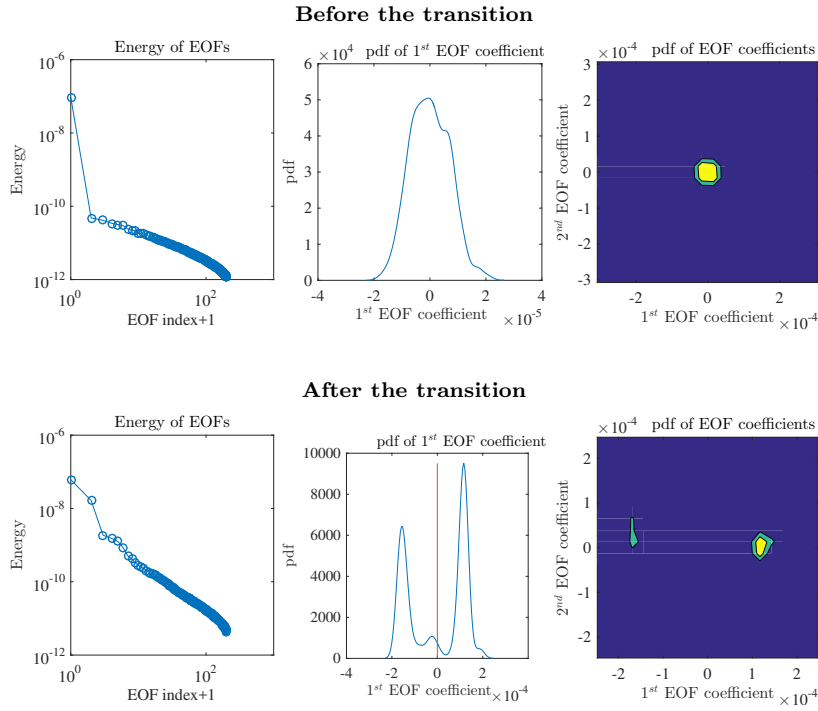
with  $M$  the number of grid points. The mean energies – or variances – of EOF coefficients correspond to the eigenvalues of the two-points correlation matrix; they are ordered in decreasing order and represent the energies associated with each spatial mode. To describe the ensemble with respect to a maximal variance point of view, only the EOF coefficients  $c_n$  concentrating the largest part of the buoyancy mean energy are kept. This energy, which differs from the energy of the mean, reads:

$$\underbrace{\hat{\mathbb{E}} \left\{ \frac{1}{M} \sum_{j=1}^M b^2(\mathbf{x}_j, t) \right\}}_{\text{Mean energy}} = \frac{1}{M} \sum_{j=1}^M \left( \hat{\mathbb{E}}(b) \right)^2(\mathbf{x}_j, t) + \frac{1}{M} \sum_{j=1}^M \widehat{Var}(b)(\mathbf{x}_j, t), \quad (122)$$

$$= \underbrace{\frac{1}{M} \sum_{j=1}^M \left( \hat{\mathbb{E}}(b) \right)^2(\mathbf{x}_j, t)}_{\text{Energy of the mean}} + \sum_{n=1}^{N_e} \underbrace{\hat{\mathbb{E}} \{ c_n^2(t) \}}_{\text{Mean energy of the } n\text{-th EOF}}. \quad (123)$$

In Figure 18, left part, the energy associated with the different spatial modes is displayed. At  $t = 30$  days, the energy of the mean field, denoted as a 0-th order EOF (index+1 = 1), is much larger than the variance field, described by the other EOFs. On the contrary, at  $t = 70$  days, after the breaking symmetry, the energy of the mean and the variance have the same order of magnitude. The variance is mainly explained by the first EOF. Thus, at the first order, the randomness of the tracer is approximately encoded by this first EOF coefficient,  $c_1$ . Its probability density function and the joint probability density function for the two first EOF coefficients presented in Figure 18 are unimodal at  $t = 30$  days and bimodal at  $t = 70$  days. Note that the tracer is clearly non-Gaussian. The symmetry breaking has created two likely scenarios in the ensemble. The scenario corresponding to negative values of the first EOF coefficient (probability of 47%) is called scenario A and the scenario corresponding to positive values (probability of 53%) is called scenario





**Fig. 18** Energy of the EOFs ( $m^2 \cdot s^{-4}$ ) (left), probability density function of the first EOF coefficient (middle) and joint probability density function for the two first EOF coefficients (right) after 30 days (top) and 70 days (bottom). The ensemble is simulated according to the SQG model under moderate uncertainty at resolution  $128^2$ . At  $t = 30$  days, the energy of the mean field, denoted as a 0-th order EOF (index+1 = 1), is much higher than the variance field, described by the other EOFs. On the contrary, at  $t = 70$  days (after the symmetry breaking), the energy of the mean and variance have the same order of magnitude. The variance is mainly explained by the first EOF. Therefore, at the first order, the random component of the tracer can be approximated by this first EOF. The probability density function of the first EOF coefficient and the joint probability density function for the two first EOF coefficients are unimodal at  $t = 30$  days and bimodal at  $t = 70$  days. The breaking symmetry has created two likely scenarios, which are very different from one another. The scenario A corresponds to negative values of the first EOF coefficient (probability of 47%) whereas the scenario B corresponds to positive value (probability of 53%). The red line separates the probability density function between the two scenarios.

B. The red line at zero separates the probability density function between the two scenarios. In Figure 19, the same probability density function along time is plotted. The bifurcation is clearly visible. Also shown, the mean buoyancy ( $m \cdot s^{-2}$ ) of the two likely scenarios are represented after 70 days of advection.

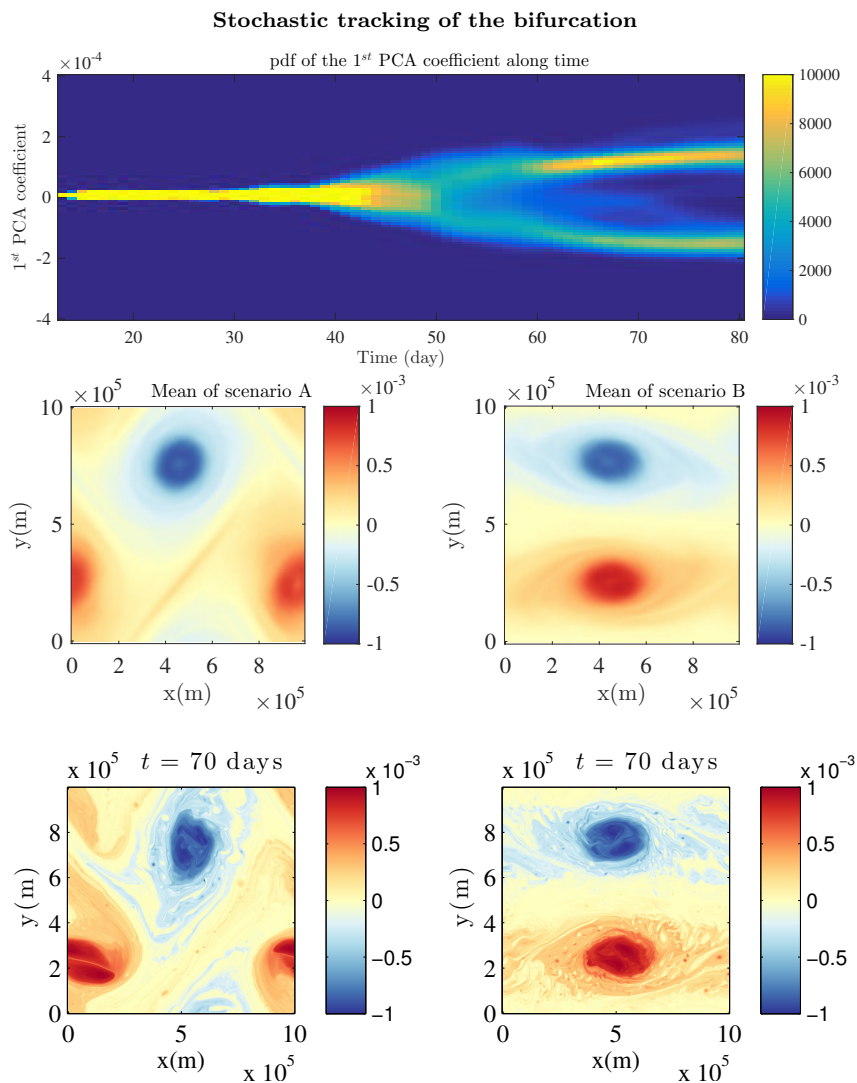
The two mean fields,  $\hat{\mathbb{E}}(b|1)$  and  $\hat{\mathbb{E}}(b|2)$ , are calculated as:

$$\hat{\mathbb{E}}(b|1) = \hat{\mathbb{E}}(b) + \hat{\mathbb{E}}(c_1|1)\Psi_1 = \hat{\mathbb{E}}(b) + \left( \frac{1}{\#\{i|c_1^{(i)} < 0\}} \sum_{c_1^{(i)} < 0} c_1^{(i)} \right) \Psi_1, \quad (124)$$

$$\hat{\mathbb{E}}(b|2) = \hat{\mathbb{E}}(b) + \hat{\mathbb{E}}(c_1|2)\Psi_1 = \hat{\mathbb{E}}(b) + \left( \frac{1}{\#\{i|c_1^{(i)} \geq 0\}} \sum_{c_1^{(i)} \geq 0} c_1^{(i)} \right) \Psi_1, \quad (125)$$

where  $\#$  stands for the cardinality of a set. The scenario A is quite close to the scenario 1, which is the reference with this initial condition, whereas scenario B is close to scenario 2. The stochastic model has enabled the ensemble to track both scenarios and to describe them statistically. Let us point out that the shape of the isotropic small-scale velocity expression has been loosely fixed by an *a priori* form of the spectrum. Some learning procedures of the noise topology from past data could lead to express more informative heterogeneous random fields, and to statistically favor the most likely transition. Moreover, since the two scenarios are very different, the introduction of few observations, through an assimilation procedure, could very easily help to select the right scenario.

For sake of comparison, we also show results obtained using the deterministic SQG model, with the “*Spectral*” perturbation of the initial conditions. This method was already used in the previous section for short-term forecast. We first briefly recall how these perturbations are generated. Initial small-scale buoyancy perturbations are assumed Gaussian and sampled from a  $(-\frac{5}{3})$  spectrum, as shown in Figure 5. These perturbations should not change the large-scale flow before the predictability time (about one month). Accordingly, we require those perturbations to be of small amplitudes and restrict them to small scales. The same analysis is performed, including the EOF decomposition and the distinction between two likely scenarios. Figure 20 gathers the results. After 70 days of advection, the scenario A is hardly visible in the probability density function of the first EOF. The ensemble estimates a probability of only 39% for the reference scenario. The  $SQG_{MU}$  ensemble estimated a probability of 47%. Moreover, the probability density function is very noisy. This suggests that the ensemble may not be converged, *i.e.* the empirical statistics of the ensemble will change if the ensemble size grows. The scenario A completely disappears in the joint probability density function for the two first EOF coefficients. The probability density function of the first EOF along time, before and after the symmetry breaking, exhibits very narrow branches associated with high probabilities, compared to Figure 19. It indicates that randomized initial conditions may lead to underdispersive ensemble. Furthermore, in Figure 20, trajectories of some realizations are still visible after the bifurcation. This confirms that the probability density function did not converge. Indeed, to estimate this density, we use the well-known Parzen-Rosenblatt estimator (Rosenblatt, 1956; Parzen, 1962): each realization is associated with a kernel



**Fig. 19** Probability density function of the first EOF coefficient along time (top), buoyancy ( $m.s^{-2}$ ) after 70 days of advection for the mean of the two likely scenarios A and B of the SQGMU model at resolution  $128^2$  (respectively middle left and middle right) and the reference scenarios 1 and 2 at high resolution  $512^2$  (respectively bottom left and bottom right). The bifurcation is clearly visible on the top plot. The two likely scenarios differ from the sign of the first EOF coefficient. They are almost associated with an equal probability of occurrence: a probability of 47% for scenario A and a probability of 53% for scenario B. The scenario A (respectively B) is similar to the scenario 1 (respectively 2).

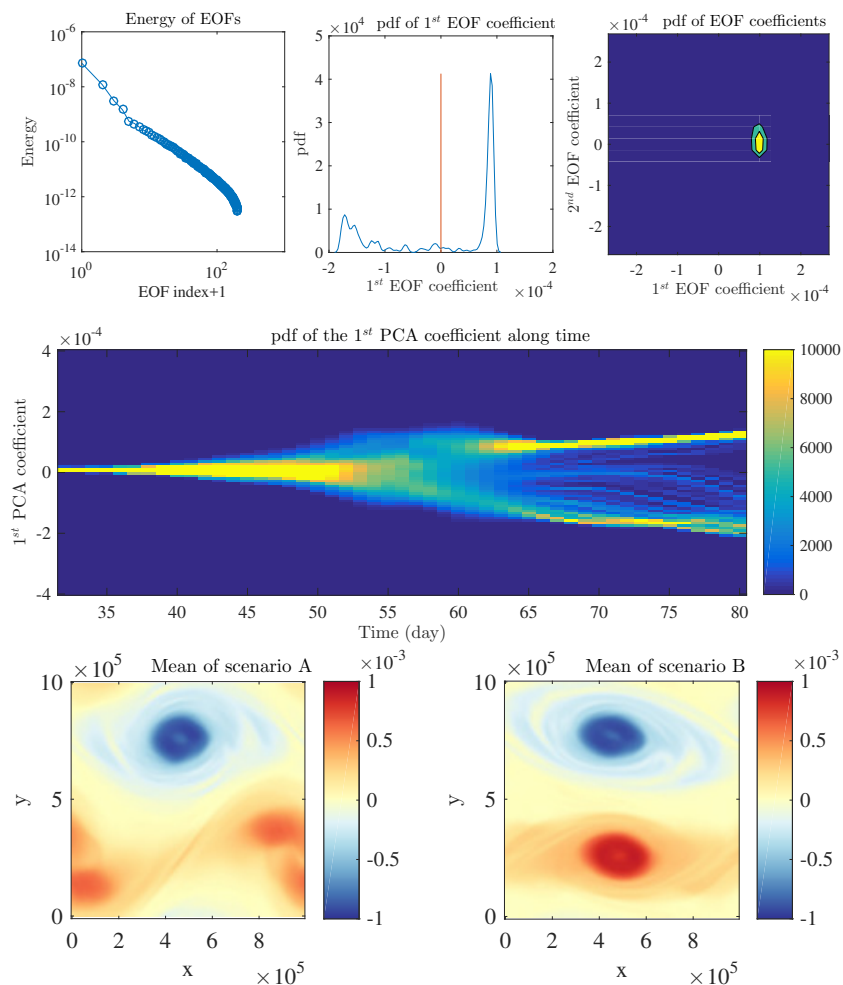
and the estimator is the sum of those kernels. Here, some realizations or set of few realizations are isolated and create spikes in the estimator. More realizations would be needed to have almost continuously distributed realizations. In other words, the ensemble is not converged. This drawback could be expected for at least two reasons. First, the structure of the initial noise contains little physical information, while the dimension of the state space is huge. Without phase information, covering all the possibilities requires a very large number of realizations. Furthermore, the subgrid tensor diffuses the small-scales components of the tracer where the ensemble variability is encoded. This is a known feature of ensemble forecasts: ensemble members tend to align with most unstable directions of the dynamics (Le Dimet and Talagrand, 1986; Ng et al., 2011; Gottwald and Harlim, 2013). Since small scales are stabilized by the subgrid tensor, the ensemble shrinks to span a smaller large-scale unstable subspace (Sapsis, 2013). On the contrary, the stochastic model associates phase and intermittency with the noise and continuously injects it into the dynamics. The phase information or inhomogeneity as well as the non-Gaussianity come from the multiplicative structure. Even though the uncorrelated velocity is only prescribed by a spectrum, the tracer gradients have phase and dynamically constraint the regions of application of the noise. This process makes the stochastic forcing much more efficient. Hence, a smaller number of realizations are needed. In Figure 20, the convergence of the probability density function of scenario B (positive values of the first EOF coefficient) seems slightly better than the density of scenario A. Unfortunately, the bottom Figures shows that the scenario B is not the one followed by the high-resolution simulation. Let us note that the reference is deterministic. Accordingly, the reference probability density is a dirac measure. Indeed, the deterministic reference initial condition is assumed to be known and is used in all large-scale simulations. Moreover, the reference dynamic defined by the high-resolution SQG model is deterministic as the real ocean dynamics is. The bad description of the scenario A tends to confirm that the SQG model with randomized initial conditions fails to describe the bifurcation.

Another argument is the distance between the reference and each ensemble. In Figure 21 reports the error corresponding to the realization closest to the reference, *i.e.* with minimal error, for each ensemble. In geophysical data assimilation, large confidence is often given to observations. Thus, the maximum a posteriori estimator is almost equal to the minimal error realization. Figure 21 shows that the ensemble from the stochastic model is closer to the reference than the ensemble with randomized initializations. It suggests that the stochastic method should lead to a better maximum a posteriori estimator.

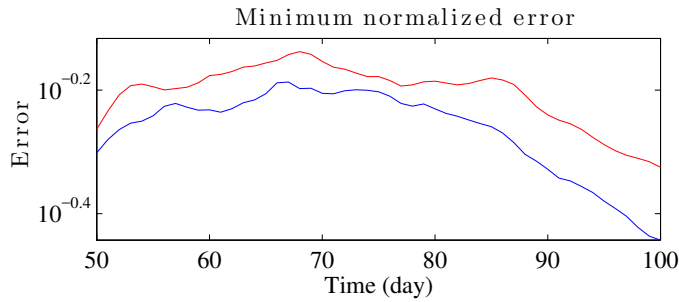
### 8.3.3 Conclusion of the long-term forecasts analysis

In this study, long-time forecasts of two different SQG models have been compared. The first one corresponds to the classical SQG equations with a random initial condition. The second one is the SALT-LU SQG, derived from a

## Same diagnostic based on a deterministic model with random initial conditions



**Fig. 20** Energy of the EOFs ( $m^2 \cdot s^{-4}$ ) (top left), probability density function of the first EOF coefficient (top middle), joint probability density function for the two first EOF coefficients (top right) after 70 days, probability density function of the first EOF coefficient along time (middle), buoyancy ( $m \cdot s^{-2}$ ) after 70 days of advection for the means of the two likely scenarios A and B (bottom left and bottom right respectively). The ensemble is simulated according to the usual deterministic SQG model with random initial conditions at resolution  $128^2$ . The joint probability density function for the two first EOF coefficients suggests only one likely scenario. The central Figure confirms that this model fails to correctly describe the bifurcation. Indeed, the probability density function appears to be not converged in this case. The bottom Figures show that the worst resolved scenario (the scenario A) is the one similar to the true reference scenario (the scenario 1).



**Fig. 21** Minimum normalized error along time in the ensemble with random initial conditions (red) and in the ensemble with random dynamics (blue). The square error was integrated over the space and divided by the energy of the reference.

stochastic expression of the transport equations (see section 6). Both models are compared to a high-resolution simulation reference.

The chosen high-resolution reference is subject to a bifurcation after 40 days of advection. An infinitesimal modification of the initial condition determines the global shape of the flow two months later. Depending on the value of this initial modification, two different scenarios are isolated. For the same initial condition, the deterministic high-resolution and the low-resolution simulations do not follow the same transition. Indeed, the sub-grid tensor associated with the coarser resolution diffuses the crucial initial perturbation before the transition. This makes the deterministic forecast useless. This result questions the classical definition of predictability and associated error which only rely on initial perturbations (Lorenz, 1969). Note that before the symmetry breaking, the large-scale errors induced by slight modifications of the initial conditions are negligible in front of large-scale model errors. To recover, the true scenario at low resolution, a possible solution could be to randomize the initial condition. As shown, this solution would require a large number of realizations. At the opposite, the proposed stochastic model tracked both scenarios. Its efficiency is mainly explained by the continuous injection of multiplicative noise. This structure dynamically constrains both the phase and the intermittency of the noise. The model achieved to predict the likelihood and the point-wise tracer probability density in each case.

To identify and separate the scenarios, a simple threshold is sufficient regarding the high energy distributed along the first PCA axis. For more realistic flow, the number of scenarios is likely to be larger. Hence, this classification may be inadequate and more advanced clustering methods (e.g. k-means and expectation-maximization algorithms) are probably required. In this paper, this ensemble aggregation was used as a diagnosis. But, it could also improve filtering methods. Indeed, for geophysical flows, the state-space dimension being large, the ensemble size is generally too small to encode all possibilities. So, when an observation is assimilated, only few realizations of the ensemble are close enough, and all the others are considered useless. This often leads to

filter degeneracy. Considering distance to likely scenarios rather than distance to realizations may help preventing this deficiency.

While the numerical analysis of this section 8 is applied to a randomized version of the SQG model with toy initial conditions, the multiplicative structure of the noise and the balance with diffusion hold for any fluid dynamics models under location uncertainty. This suggests that similar conclusions could be expected in more complex problems. Hence, this SALT-LU method opens for new ensemble forecasts methods, for both short-term forecasts and climate projections where uncertainty quantification is a main issue (Allen and Stainforth, 2002).

## 9 Conclusion

We have reviewed existing methods to generate ensemble forecasts quantifying modes errors related to coarse resolution in computational fluid dynamics. The accuracy of this quantification being a main issue in data assimilation, the paper began by recalling the principles of that simulation-measurement coupling framework. We have categorized UQ methods based on their relations to coarse-graining. Though they have not been proposed in UQ context, we have also covered stochastic backscattering models, since they have inspired several UQ methods. Many UQ methods deal with random parameters, forcing, boundary or initial conditions. The latter type of methods has been widely misused to quantify resolution-induced error. While covariance inflation mitigates this issue, most of the data assimilation community now reckons that randomizing initial conditions underestimates resolution-induced errors. To tackle this issue, many authors introduces Gaussian and non-Gaussian noises in the dynamical equations. Most of these stochastic models are empirical, but more and more are based on energy budgets or derived from physical principles. Many of the latter assume a time scale separation in the dynamic.

After presenting a new UQ method (*WavHypervis*) adaptable to most deterministic subgrid dissipation, we have focused on a recent family of stochastic subgrid models: the dynamics under location uncertainty (LU) and the stochastic advection by Lie transport (SALT). These frameworks rely on a time scale separation of the velocity field, and a stochastic Navier-Stokes model. In the latter, the large-scale velocity component is transported – up to some forcings – by the small-scale and the large-scale random velocity components. Rigorously derived from stochastic calculus theory, this stochastic transport can be decomposed into a skew-symmetric multiplicative noise and an eddy-viscosity-like diffusion term. Being a transport, it naturally ensures the conservation of many physical invariants. LU and SALT models differ in the interpretations of "transport" (classical fluid dynamics VS geometric interpretation) and of "large-scale velocity". The first interpretation difference implies in particular that the dynamics under location uncertainty conserves kinetic energy whereas stochastic advection by Lie transport conserves helicity

and circulation. The second interpretation difference leads to a modification of the large-scale advection in LU models. After an extended theoretical description, we have reviewed existing parameterization choices – i.e. subgrid velocity statistics modeling – for the SALT-LU framework.

Then, after briefly presenting the state-of-art metrics to assess UQ skills, new numerical results have been presented. We have compared the *WavHypervis* method, the dynamics under location uncertainty and the randomization of initial conditions. Test cases were short-term and long-term predictions of free decaying turbulence. The chosen dynamics was a two-dimensional meteorologic and oceanic model called the Surface Quasi-Geostrophic dynamics. For this flow, SALT and LU models mostly coincide (only the large-scale advection correction differs). Our results confirm that initial conditions randomization is not adapted to resolution-induced UQ, and that the dynamics under location uncertainty accurately spreads ensemble members along time. LU and *WavHypervis* short-term forecasts show very good calibrations and UQ scores. In particular, the recent LU parameterizations *ADSD* and *SVDpseudo* obtain the best UQ scores. At long term, even after a bifurcation of the reference fluid dynamics, the model under location uncertainty accurately and smoothly describes the likely scenarios.

SALT and LU dynamics are now mature frameworks. Yet, many questions remain.

First, it is still unclear whether SALT or LU is more appropriate for UQ purpose. Many numerical studies with appropriate UQ metrics and/or with data assimilation procedures would probably be necessary to try to answer this question.

Another possible research focus is the improvement of SALT-LU parameterizations. *ADSD* method is a promising one since it is parameter-free, fast, accurate and could in theory adapt itself to any turbulent dynamics. Nevertheless, this method is currently defined in the Fourier space. Developing a physical-domain-based *ADSD* implementation would be very useful. Scaling estimations of relative diffusion (Keating et al., 2011) or velocity structure function (Gawędzki and Kupiainen, 1995) and Matérn covariances (Williams and Rasmussen, 2006; Lim and Teo, 2009; Lilly et al., 2017; Resseguier et al., 2019) could probably help in this task. An adaptation of this method to complex boundary conditions would also be needed. The Gaussian unresolved velocity moments conditionally to the unresolved velocity boundary conditions may give a path toward this direction. New spatial anisotropic and heterogeneous SALT-LU parameterizations – like *SVDpseudo* – would also be helpful. Current works based on data-driven methods (related e.g. machine learning, Koopman operator, Girsanov-based maximum likelihood estimations) are on going.

LU and SALT long-term purpose concerns data assimilation. Preliminary studies have been published towards that direction. Using the pseudo-stochastic Navier-Stokes model under location uncertainty (see section 6.2.3) and a dynamics error model, Yang and Mémin (2017) have assimilated high-



resolution observations into simple shallow-water models. Yang and Mémin (2018) have applied the Navier-Stokes model under location uncertainty (see section 6.2.1) to ensemble filters. They have proposed estimations of both the covariance model error and the variance tensor. Using SALT and particle filters, Cotter et al. (2018b) assimilate data in a 2D Euler dynamics. Works on similar assimilation procedures with a quasi-geostrophic dynamics are on going. Yet, SALT-LU-based data assimilation is still in its infancy. Those first promising results need to be extended and assessed on more realistic fluid dynamic model. For flow control purposes, very fast data assimilation procedures could also be developed in stochastic reduced order frameworks (Resseguier, 2017, chapter 8).

## Acknowledgments

The authors acknowledge the support of the ESA DUE GlobCurrent project, the “Laboratoires d’Excellence” CominLabs, Lebesgue and Mer through the SEACS project. The authors also acknowledge Darryl D. Holm, Francois Gay-Balmaz, Dan Crisan, Wei Pan and Igor Shevchenko for helpful discussions.

## A Spectral energy flux in models under location uncertainty

In the homogeneous case,  $\sigma d\mathbf{B}_t = \check{\sigma} * d\mathbf{B}_t$  and  $\mathbf{a} = a_0 \mathbb{I}_d$  does not depend on  $x$ . Therefore, a – possibly active – tracer  $q$  is solution of:

$$0 = D_t q = d_t q + (\mathbf{w} dt + \sigma d\mathbf{B}_t) \cdot \nabla q - \frac{a_0}{2} \Delta q dt. \quad (126)$$

For sake of simplicity, we do not explicitly show the time dependence in  $q$  and its Fourier transform, in this appendix. In Fourier space, we get an infinite dimensional SDE (not an SPDE):

$$\begin{aligned} d_t \hat{q}(\mathbf{k}) = & -\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left( \hat{\mathbf{w}}(\mathbf{k}') dt + \hat{\check{\sigma}}(\mathbf{k}') d\hat{\mathbf{B}}_t(\mathbf{k}') \right) \cdot i(\mathbf{k} - \mathbf{k}') \hat{q}(\mathbf{k} - \mathbf{k}') d\mathbf{k}' \\ & - \frac{a_0}{2} \|\mathbf{k}\|_2^2 \hat{q}(\mathbf{k}) dt. \end{aligned} \quad (127)$$

To find a PDE associated to the Fourier modes of the tracer (Fokker-Planck equation), we need an SDE in  $\mathbb{R}$ . The previous one is in  $\mathbb{C}$ . Either we consider the closed coupled equations on the real part,  $\Re(\hat{q}(\mathbf{k}))$ , and imaginary part,  $\Im(\hat{q}(\mathbf{k}))$ , either we decompose through modulus and phase. The equation on the modulus is not closed but still very instructive. From now on, the dimension of the spatial space,  $d$ , will be 2. However, we expect that the following results are still correct for  $d = 3$ . Since  $\sigma d\mathbf{B}_t$  is divergence free, we set  $\check{\sigma} = \nabla^\perp \psi_\sigma$ . We can notice that:

$$d \langle \mathbf{B}_t(\mathbf{x}_1), \mathbf{B}_t(\mathbf{x}_2) \rangle = \delta(\mathbf{x}_1 - \mathbf{x}_2) dt, \quad (128)$$

and then,

$$\Re \left\{ d \langle \hat{\mathbf{B}}_t(\mathbf{k}_1), \overline{\hat{\mathbf{B}}_t(\mathbf{k}_2)} \rangle \right\} = \Re \left\{ \iint_{(\mathbb{R}^d)^2} \delta(\mathbf{x} - \mathbf{y}) e^{-i\mathbf{k}_1 \cdot \mathbf{x}} e^{+i\mathbf{k}_2 \cdot \mathbf{y}} d\mathbf{x} d\mathbf{y} \right\} dt \quad (129)$$

$$= (2\pi)^d \delta(\mathbf{k}_1 - \mathbf{k}_2) dt. \quad (130)$$

The classical Ito formula in  $\mathbb{C}$ , equation (127) and the above result lead to:

$$d_t|\hat{q}(\mathbf{k})|^2 = 2\Re\left\{\hat{q}(\mathbf{k})d\overline{\hat{q}(\mathbf{k})}\right\} + \Re\left\{d\langle\hat{q}(\mathbf{k}),\overline{\hat{q}(\mathbf{k})}\rangle\right\}, \quad (131)$$

$$\begin{aligned} &= -a_0\|\mathbf{k}\|_2^2\hat{q}(\mathbf{k})dt + \underbrace{\frac{2}{(2\pi)^d}\Re\int_{\mathbb{R}^d}\hat{\mathbf{w}}(\mathbf{k}')\cdot i(\mathbf{k}-\mathbf{k}')\hat{q}(\mathbf{k}-\mathbf{k}')\overline{\hat{q}(\mathbf{k})}d\mathbf{k}'}_{=F_1}dt \\ &+ \underbrace{\frac{2}{(2\pi)^d}\Re\int_{\mathbb{R}^d}\hat{\sigma}(\mathbf{k}')d\hat{\mathbf{B}}_t(\mathbf{k}')\cdot i(\mathbf{k}-\mathbf{k}')\hat{q}(\mathbf{k}-\mathbf{k}')\overline{\hat{q}(\mathbf{k})}d\mathbf{k}'}_{=dF_2} \end{aligned} \quad (132)$$

$$+ \frac{(2\pi)^d}{(2\pi)^{2d}}\int_{\mathbb{R}^d}\left|(\mathbf{k}-\mathbf{k}')^T\hat{\sigma}(\mathbf{k}')\hat{q}(\mathbf{k}-\mathbf{k}')\right|^2d\mathbf{k}'dt. \quad (133)$$

$dF_2$  does not modify the mean spectral energy budget since  $\mathbb{E}(dF_2) = 0$ , even if it modifies the law of  $|\hat{q}(\mathbf{k})|^2$ . Note that the mean spectral energy budget is the study of the evolution of  $\mathbb{E}|\hat{q}(\mathbf{k})|^2$  whereas the spectral energy budget of the mean is the much more trivial study of  $|\mathbb{E}\hat{q}(\mathbf{k})|^2$ . We can explicit the Laplacian term using the following expression of the variance tensor, based on the homogeneity assumption:

$$\mathbf{a}_{ij} = \mathbf{a}_{ij}(0) = \int_{\mathbb{R}^d}\check{\sigma}_i(\mathbf{z})\check{\sigma}_j(\mathbf{z})d\mathbf{z} = \int_{\mathbb{R}^d}(\nabla^\perp\psi_\sigma(\mathbf{z}))_i(\nabla^\perp\psi_\sigma(\mathbf{z}))_jd\mathbf{z}, \quad (134)$$

$$= \frac{1}{(2\pi)^d}\Re\int_{\mathbb{R}^d}\widehat{(\nabla^\perp\psi_\sigma(\mathbf{k}'))}_i\widehat{(\nabla^\perp\psi_\sigma(\mathbf{k}'))}_jd\mathbf{k}', \quad (135)$$

$$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\underbrace{(\mathbf{k}'^\perp)_i(\mathbf{k}'^\perp)_j}_{\substack{\text{even function if } i=j, \\ \text{odd function if } i\neq j}}\left|\widehat{\psi_\sigma(\|\mathbf{k}'\|)}\right|^2d\mathbf{k}', \quad (136)$$

$$= \frac{1}{d}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\underbrace{\left|\widehat{\psi_\sigma(\|\mathbf{k}'\|)}\right|^2}_{=a_0}d\mathbf{k}'\delta_{ij}. \quad (137)$$

Injecting this into the spectral energy budget yields:

$$\begin{aligned} d_t|\hat{q}(\mathbf{k})|^2 &= F_1dt + dF_2 - \frac{1}{(2\pi)^dd}\int_{\mathbb{R}^d}\left|\|\mathbf{k}\|\|\mathbf{k}'\|\widehat{\psi_\sigma(\mathbf{k}')}\hat{q}(\mathbf{k})\right|^2d\mathbf{k}'dt \\ &+ \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left|\mathbf{k}\cdot\mathbf{k}'^\perp\widehat{\psi_\sigma(\mathbf{k}')}\hat{q}(\mathbf{k}-\mathbf{k}')\right|^2d\mathbf{k}'dt, \end{aligned} \quad (138)$$

$$\begin{aligned} &= F_1dt + dF_2 \\ &+ \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left|\|\mathbf{k}\|\|\mathbf{k}'\|\widehat{\psi_\sigma(\mathbf{k}')}\right|^2\left(\sin^2(\widehat{(\mathbf{k},\mathbf{k}')} )|\hat{q}(\mathbf{k}-\mathbf{k}')|^2 - \frac{1}{d}|\hat{q}(\mathbf{k})|^2\right)d\mathbf{k}'dt. \end{aligned} \quad (139)$$

Taking the expectation of the previous equation highlights the terms which contribute to the mean spectral energy budget.

$$\begin{aligned} \partial_t\mathbb{E}|\hat{q}(\mathbf{k})|^2 &= \mathbb{E}(F_1) \\ &+ \frac{1}{(2\pi)^d}\mathbb{E}\int_{\mathbb{R}^d}\left|\|\mathbf{k}\|\|\mathbf{k}'\|\widehat{\psi_\sigma(\mathbf{k}')}\right|^2\left(\sin^2(\widehat{(\mathbf{k},\mathbf{k}')} )|\hat{q}(\mathbf{k}-\mathbf{k}')|^2 - \frac{1}{d}|\hat{q}(\mathbf{k})|^2\right)d\mathbf{k}'. \end{aligned} \quad (140)$$

## B Stratonovich material derivative

Here, we express the Stratonovich material derivative:

$$D_t^S q(\mathbf{x}, t) \triangleq \left( q \left( \mathbf{X}_{t+\frac{dt}{2}}, t + \frac{dt}{2} \right) - q \left( \mathbf{X}_{t-\frac{dt}{2}}, t - \frac{dt}{2} \right) \right) \Big|_{\mathbf{X}_t = \mathbf{x}}, \quad (141)$$

as a function of the Ito material derivative

$$D_t q(\mathbf{x}, t) \triangleq (q(\mathbf{X}_{t+dt}, t + dt) - q(\mathbf{X}_t, t)) \Big|_{\mathbf{X}_t = \mathbf{x}} \quad (142)$$

and its forcing

$$D_t q = F dt + \mathbf{H}^T d\mathbf{B}_t. \quad (143)$$

To do derive these results, we will apply the Stratonovich-Ito-notation-change formula:

$$r \circ ds = r ds + \frac{1}{2} d \langle r, s \rangle, \quad (144)$$

which can be found in Kunita (1997), theorem 3.2.5 page 60. First, let us rewrite equation (143) in Lagrangian coordinates:

$$d(q(\mathbf{X}_t, t)) = q(\mathbf{X}_{t+dt}, t + dt) - q(\mathbf{X}_t, t) = F(\mathbf{X}_t, t) dt + \mathbf{H}^T(\mathbf{X}_t, t) d\mathbf{B}_t. \quad (145)$$

Then, we apply the formula (144) to obtain the corresponding Stratonovich equation:

$$d(q(\mathbf{X}_t, t)) = q \left( \mathbf{X}_{t+\frac{dt}{2}}, t + \frac{dt}{2} \right) - q \left( \mathbf{X}_{t-\frac{dt}{2}}, t - \frac{dt}{2} \right), \quad (146)$$

$$= F(\mathbf{X}_t, t) dt - d \langle \mathbf{H}^T(\mathbf{X}_t, t), \mathbf{B}_t \rangle + \mathbf{H}^T(\mathbf{X}_t, t) \circ d\mathbf{B}_t, \quad (147)$$

$$= \left( F - \frac{1}{2} \frac{d}{dt} \langle \mathbf{H}^T, \mathbf{B}_t \rangle - \frac{1}{2} \text{tr}((\sigma^T \nabla) \mathbf{H}^T) \right) (\mathbf{X}_t, t) dt + \mathbf{H}^T(\mathbf{X}_t, t) \circ d\mathbf{B}_t, \quad (148)$$

since, by application of the Ito-Wentzell-formula,

$$d(\mathbf{H}^T(\mathbf{X}_t, t)) = (d\mathbf{H})^T(\mathbf{X}_t, t) + (\sigma d\mathbf{B}_t \cdot \nabla) \mathbf{H}^T + (\text{others terms in } dt). \quad (149)$$

Finally, rewriting everything in the Eulerian grid  $\mathbf{X}_t = \mathbf{x}$  gives the result:

$$D_t^S q(\mathbf{x}, t) = \left( F - \frac{1}{2} \frac{d}{dt} \langle \mathbf{H}^T, \mathbf{B}_t \rangle - \frac{1}{2} \text{tr}((\sigma^T \nabla) \mathbf{H}^T) \right) (\mathbf{x}, t) dt + \mathbf{H}^T(\mathbf{x}, t) \circ d\mathbf{B}_t. \quad (150)$$

After this, we can reapply formula (144) in the Eulerian grid to get:

$$D_t^S q(\mathbf{x}, t) = \left( F - \frac{1}{2} \text{tr}((\sigma^T \nabla) \mathbf{H}^T) \right) (\mathbf{x}, t) dt + \mathbf{H}^T(\mathbf{x}, t) d\mathbf{B}_t, \quad (151)$$

$$= D_t q(\mathbf{x}, t) - \frac{1}{2} \text{tr}((\sigma^T \nabla) \mathbf{H}^T) (\mathbf{x}, t) dt. \quad (152)$$

by identification with the balance (142).

## C Effective resolution and inertial range

Let us assume the simulated evolution law is  $D_t q = -\nu(-\Delta)^p q dt$ . The deterministic subgrid model  $-\nu(-\Delta)^p q$  acts, in a finite time  $t$ , as a low-pass filter. In Fourier space, this filter is:

$$F(\|\mathbf{k}\|) = \exp(-\nu t \|\mathbf{k}\|^{2p}). \quad (153)$$

If the hyperviscosity  $\nu$  is well chosen, we may expect that at the Shanon resolution  $\pi/\Delta x = \kappa_M$ , only 10% of the energy is left by the filter, i.e.

$$F(\kappa_M) = 1/10. \quad (154)$$

A ratio smaller than 10% may lead to an over-damped simulation. Moreover, the precise value of this ratio does not influence much our final estimate.

We may define the effective resolution as the scale  $\kappa = \kappa_m$  where the deterministic subgrid model influence is negligible. There, we may expect the filter to be equal to 95%, i.e.:

$$F(\kappa_m) = 95/100. \quad (155)$$

The ratio  $\kappa_m/\kappa_M$  can then be derived from formulas (153), (154) and (155).

## References

- Allen M, Stainforth D (2002) Towards objective probabilistic climate forecasting. *Nature* 419(6903):228–228
- Anderson J, Anderson S (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* 127(12):2741–2758
- Anderson JL (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* 9(7):1518–1530
- Berloff P (2005) Random-forcing model of the mesoscale oceanic eddies. *Journal of Fluid Mechanics* 529:71–95
- Berloff P, McWilliams J (2002) Material transport in oceanic gyres. part II: Hierarchy of stochastic models. *Journal of Physical Oceanography* 32(3):797–830
- Berner J, Shutts G, Leutbecher M, Palmer T (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences* 66(3):603–626
- Berner J, Ha SY, Hacker J, Fournier A, Snyder C (2011) Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review* 139(6):1972–1995
- Berner J, Achatz U, Batte L, Camara ADL, Crommelin D, Christensen H, Colangeli M, Dolaptchiev S, Franzke C, Friederichs P, Imkeller P, Jarvinen H, Juricke S, Kitsios V, Lott F, Lucarini V, Mahajan S, Palmer T, Penland C, Storch JSV, Sakradzija M, Weniger M, Weisheimer A, Williams P, Yano JI (2015) Stochastic parameterization: towards a new view of weather and climate models. Tech. rep., arXiv:1510.08682 [physics.ao-ph]
- Blumen W (1978) Uniform potential vorticity flow: part I. theory of wave interactions and two-dimensional turbulence. *Journal of the Atmospheric Sciences* 35(5):774–783
- Bocquet M, Sakov P (2014) An iterative ensemble kalman smoother. *Quarterly Journal of the Royal Meteorological Society* 140(682):1521–1535
- Boussinesq J (1877) Essai sur la théorie des eaux courantes. *Mémoires présentés par divers savants à l'Académie des Sciences*, 23 (1): 1–680
- Brankart J (2013) Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modeling. *Ocean Modelling* 66:64–76
- Brzeźniak Z, Capiński M, Flandoli F (1991) Stochastic partial differential equations and turbulence. *Mathematical Models and Methods in Applied Sciences* 1(01):41–59
- Buehner M (2005) Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational nwp setting. *Quarterly Journal of the Royal Meteorological Society* 131(607):1013–1043
- Buizza R (2016a) Representing model uncertainty for climate forecasts. In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading
- Buizza R (2016b) Weather prediction in a world of uncertainties: should ensembles simulate the effect of model approximations? In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading

- Buizza R, Miller M, Palmer T (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal Royal Meteorological Society* 125:2887–2908
- Cai S, Mémin E, Dérian P, Xu C (2018) Motion estimation under location uncertainty for turbulent fluid flows. *Experiments in Fluids* 59(1):8
- Candy J (2011) Bayesian signal processing: Classical, modern and particle filtering methods, vol 54. John Wiley & Sons
- Capet X, Klein P, Hua B, Lapeyre G, McWilliams J (2008) Surface kinetic energy transfer in surface quasi-geostrophic flows. *J Fluid Mech* 604:165–174
- Chandramouli P, Heitz D, Laizet S, Mémin E (2018) Coarse large-eddy simulations in a transitional wake flow with flow models under location uncertainty. *Computers & Fluids* 168:170–189
- Chandramouli P, Mémin E, Chapron B, Heitz D (2019) Deciphering the role of small-scale inhomogeneity on flow structuration: a stochastic approach, submitted
- Chapron B, Dérian P, Mémin E, Resseguier V (2018) Large-scale flows under location uncertainty: a consistent stochastic framework. *Quarterly Journal of the Royal Meteorological Society* 144(710):251–260
- Chasnov J (1991) Simulation of the Kolmogorov inertial subrange using an improved subgrid model. *Physics of Fluids A: Fluid Dynamics* (1989-1993) 3(1):188–200
- Chekroun MD, Kondrashov D, Ghil M (2011) Predicting stochastic systems by noise sampling, and application to the el niño-southern oscillation. *Proceedings of the National Academy of Sciences* 108(29):11766–11771
- Chekroun MD, Liu H, McWilliams JC (2017) The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories. *Computers & Fluids* 151:3–22
- Chow PL (2014) Stochastic partial differential equations. CRC Press
- Cintolesi C, Mémin E (2019a) Pseudo-stochastic simulation of turbulent channel flows with near-wall modelling, URL <https://hal.archives-ouvertes.fr/hal-02044818>, working paper or preprint
- Cintolesi C, Mémin E (2019b) Stochastic modelling of turbulent flows for numerical simulations, URL <https://hal.archives-ouvertes.fr/hal-02044809>
- Cotter C, Gottwald G, Holm D (2017) Stochastic partial differential fluid equations as a diffusive limit of deterministic lagrangian multi-time dynamics. arXiv preprint arXiv:170600287
- Cotter C, Crisan D, Holm DD, Pan W, Shevchenko I (2018a) Modelling uncertainty using circulation-preserving stochastic transport noise in a 2-layer quasi-geostrophic model. arXiv preprint arXiv:180205711
- Cotter C, Crisan D, Holm DD, Pan W, Shevchenko I (2018b) Sequential monte carlo for stochastic advection by lie transport. In preparation
- Cotter CJ, Crisan D, Holm DD, Pan W, Shevchenko I (2018c) Numerically modelling stochastic lie transport in fluid dynamics. arXiv preprint arXiv:180109729
- Craig G, Forbes R, Abdalla S, Balsamo G, Bechtold P, Berner J, Buizza R, Pallares AC, Meutner PD, Dueben P, Frognier IL, Gagnon N, Hodyss D, Holm D, Lock SJ, Machulskaya E, Najafi H, Ollinaho P, Selz T, Separovic L, Subramanian A, Tseng J, Weisheimer A (2016) Reports from working group 1 : What are the sources of model error and how can we improve the physical basis of model uncertainty representation? In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading
- Crisan D, Flandoli F, Holm DD (2017) Solution properties of a 3d stochastic euler fluid equation. arXiv preprint arXiv:170406989
- Da Prato G, Zabczyk J (1992) Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and its Applications, Cambridge University Press
- Desroziers G, Berre L, Chapnik B, Poli P (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 131(613):3385–3396
- Doucet A, Johansen A (2009) A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering* 12:656–704

- Doucet A, De Freitas N, Gordon N (2001) Sequential Monte Carlo methods in practice. Springer
- Dukowicz JK, Smith RD (1997) Stochastic theory of compressible turbulent fluid transport. *Physics of Fluids* 9(11):3523–3529
- Dwivedi S, Franzke CL, Lunkeit F (2019) Energetically consistent scale adaptive stochastic and deterministic energy backscatter schemes for an atmospheric model. *Quarterly Journal of the Royal Meteorological Society*
- Ehrendorfer M (1997) Predicting the uncertainty of numerical weather forecasts: A review. *METEOROLOGISCHE ZEITSCHRIFT-BERLIN-* 6:147–183
- Falkovich G, Gawędzki K, Vergassola M (2001) Particles and fields in fluid turbulence. *Reviews of modern Physics* 73(4):913
- Farrell B, Ioannou P (2014) Statistical state dynamics: a new perspective on turbulence in shear flow. arXiv preprint arXiv:14128290
- Ferrari R, Nikurashin M (2010) Suppression of eddy diffusivity across jets in the Southern Ocean. *Journal of Physical Oceanography* 40(7):1501–1519
- Flandoli F (2011) The interaction between noise and transport mechanisms in PDEs. *Milan Journal of Mathematics* 79(2):543–560
- Frank JE, Gottwald GA (2013) Stochastic homogenization for an energy conserving multi-scale toy model of the atmosphere. *Physica D: Nonlinear Phenomena* 254:46–56
- Franzke C, Majda A, Vanden-Eijnden E (2005) Low-order stochastic mode reduction for a realistic barotropic model climate. *Journal of the atmospheric sciences* 62(6):1722–1745
- Franzke C, O’Kane T, Berner J, Williams P, Lucarini V (2015) Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change* 6(1):63–78
- Franzke CL (2017) Extremes in dynamic-stochastic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27(1):012101
- Frederiksen JS (1999) Subgrid-scale parameterizations of eddy-topographic force, eddy viscosity, and stochastic backscatter for flow over topography. *Journal of the atmospheric sciences* 56(11):1481–1494
- Frisch U (1995) Turbulence: the legacy of AN Kolmogorov. Cambridge university press
- Gawędzki K, Kupiainen A (1995) Anomalous scaling of the passive scalar. *Physical review letters* 75(21):3834
- Gay-Balmaz F, Holm DD (2018) Stochastic geometric models with non-stationary spatial correlations in lagrangian fluid flows. *Journal of nonlinear science* 28(3):873–904
- Gent P, McWilliams J (1990) Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography* 20(1):150–155
- Germano M, Piomelli U, Moin P, Cabot W (1991) A dynamic subgrid-scale eddy viscosity model. *Phys of Fluids* 3:1760–1765
- Givon D, Kupferman R, Stuart A (2004) Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* 17(6):R55
- Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477):359–378
- Gneiting T, Raftery AE, Westveld III AH, Goldman T (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review* 133(5):1098–1118
- Gottwald G, Harlim J (2013) The role of additive and multiplicative noise in filtering complex dynamical systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 469(2155):20130096
- Gottwald G, Melbourne I (2013) Homogenization for deterministic maps and multiplicative noise. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 469(2156)
- Gottwald G, Crommelin D, Franzke C (2015) Stochastic climate theory. In: *Nonlinear and Stochastic Climate Dynamics*, Cambridge University Press
- Grimit EP, Gneiting T, Berrocal V, Johnson NA (2006) The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 132(621C):2925–2942
- Grooms I, Majda A (2014) Stochastic superparameterization in quasigeostrophic turbulence. *Journal of Computational Physics* 271:78–98

- Gugole F, Franzke C (219) Numerical development and evaluation of an energy conserving conceptual stochastic climate model. *Mathematics of climate and weather forecasting* In press
- Gula J, Molemaker J, McWilliams J (2015) Gulf stream dynamics along the southeastern us seaboard. *Journal of Physical Oceanography* 45(3):690–715
- Hamill TM (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review* 129(3):550–560
- Hamill TM, Colucci SJ (1997) Verification of eta-rsm short-range ensemble forecasts. *Monthly Weather Review* 125(6):1312–1327
- Harouna SK, Mémin E (2017) Stochastic representation of the reynolds transport theorem: revisiting large-scale modeling. *Computers & Fluids* 156:456–469
- Hasselmann K (1976) Stochastic climate models. part I: theory. *Tellus* 28:473–485
- Held I, Pierrehumbert R, Garner S, Swanson K (1995) Surface quasi-geostrophic dynamics. *Journal of Fluid Mechanics* 282:1–20
- Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* 15(5):559–570
- Holm D (2015) Variational principles for stochastic fluid dynamics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 471(2176)
- Holmes P, Lumley J, Berkooz G (1996) *Turbulence, coherence structures, dynamical systems and symmetry*. Cambridge university press
- Jain A, Timofeyev I, Vanden-Eijnden E (2014) Stochastic mode-reduction in models with conservative fast sub-systems. arXiv preprint arXiv:14103004
- Jansen MF, Held IM (2014) Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling* 80:36–48
- Jordan A (2016) Facets of forecast evaluation
- Keating S, Smith S, Kramer P (2011) Diagnosing lateral mixing in the upper ocean with virtual tracers: Spatial and temporal resolution dependence. *Journal of Physical Oceanography* 41(8):1512–1534
- Keating S, Majda A, Smith S (2012) New methods for estimating ocean eddy heat transport using satellite altimetry. *Monthly Weather Review* 140(5):1703–1722
- Klein P, Hua B, Lapeyre G, Capet X, Le Gentil S, Sasaki H (2008) Upper ocean turbulence from high-resolution 3D simulations. *Journal of Physical Oceanography* 38(8):1748–1763
- Klyatskin V (2005) *Stochastic equations through the eye of the physicist: Basic concepts, exact results and asymptotic approximations*. Elsevier
- Klyatskin V, Woyczynski W, Gurarie D (1996) Short-time correlation approximations for diffusing tracers in random velocity fields: A functional approach. In: *Stochastic modelling in physical oceanography*, Springer, pp 221–269
- Kondrashov D, Berloff P (2015) Stochastic modeling of decadal variability in ocean gyres. *Geophysical Research Letters* 42(5):1543–1553
- Kondrashov D, Chekroun MD, Ghil M (2015) Data-driven non-markovian closure models. *Physica D: Nonlinear Phenomena* 297:33–55
- Kraichnan R (1968) Small-scale structure of a scalar field convected by turbulence. *Physics of Fluids* (1958-1988) 11(5):945–953
- Kraichnan R (1994) Anomalous scaling of a randomly advected passive scalar. *Physical review letters* 72(7):1016
- Kunita H (1997) *Stochastic flows and stochastic differential equations*, vol 24. Cambridge university press
- Kurtz T (1973) A limit theorem for perturbed operator semigroups with applications to random evolutions. *Journal of Functional Analysis* 12(1):55–67
- Laio F, Tamea S (2007) Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences* 11(4):1267–1277
- Lapeyre G (2017) Surface quasi-geostrophy. *Fluids* 2(1):7
- Le Dimet FX, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* 38(2):97–110
- Le Maitre O, Reagan M, Najm H, Ghanem R, Knio O (2002) A stochastic projection method for fluid flow. II. random process. *Journal of Computational Physics* 181:9–44
- Leith C (1971) Atmospheric predictability and two-dimensional turbulence. *Journal of the Atmospheric Sciences* 28(2):145–161

- Leith C (1990) Stochastic backscatter in a subgrid-scale model: Plane shear mixing layer. *Physics of Fluids A: Fluid Dynamics* (1989-1993) 2(3):297–299
- Leutbechner M, Ollinaha P, Lock SJ, Lang S, Bechtold P, Beljaars A, Bozzo A, Forbes R, Haiden R T Hogan, Sandu I (2016) Stochastic representations of model uncertainties in the ifs. In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading
- Lilly D (1966) On the application of the eddy viscosity concept in the inertial subrange of turbulence. Tech. Rep. 123, NCAR
- Lilly J, Sykulski A, Early J, Olhede S (2017) Fractional brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Processes in Geophysics* 24(1):481–514
- Lim S, Teo L (2009) Generalized whittle–matérn random field as a model of correlated fluctuations. *Journal of Physics A: Mathematical and Theoretical* 42(10):105202
- Liu C, Xiao Q, Wang B (2009) An ensemble-based four-dimensional variational data assimilation scheme. part ii: Observing system simulation experiments with advanced research wrf (arw). *Monthly Weather Review* 137(5):1687–1704
- Lorenz E (1969) The predictability of a flow which possesses many scales of motion. *Tellus* 21:289–307
- Lu F, Lin KK, Chorin AJ (2017) Data-based stochastic model reduction for the kuramoto–sivashinsky equation. *Physica D: Nonlinear Phenomena* 340:46–57
- Lucarini V, Blender R, Herbert C, Ragone F, Pascale S, Wouters J (2014) Mathematical and physical ideas for climate science. *Reviews of Geophysics* 52(4):809–859
- Majda A (2015) Statistical energy conservation principle for inhomogeneous turbulent dynamical systems. *Proceedings of the National Academy of Sciences* 112(29):8937–8941
- Majda A, Timofeyev I, Eijnden E (1999) Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences* 96(26):14687–14691
- Majda A, Timofeyev I, Vanden Eijnden E (2001) A mathematical framework for stochastic climate models. *Communications on Pure and Applied Mathematics* 54(8):891–974
- Majda A, Franzke C, Khouider B (2008) An applied mathematics perspective on stochastic modelling for climate. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 366(1875):2427–2453
- Mana P, Zanna L (2014) Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling* 79:1–20
- Matheson JE, Winkler RL (1976) Scoring rules for continuous probability distributions. *Management science* 22(10):1087–1096
- McCabe A, Swinbank R, Tennant W, Lock A (2016) Representing model error in the met office convection permitting ensemble prediction system. In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading
- Mémin E (2014) Fluid flow dynamics under location uncertainty. *Geophysical & Astrophysical Fluid Dynamics* 108(2):119–146, DOI 10.1080/03091929.2013.836190
- Métais O, Lesieur M (1986) Statistical predictability of decaying turbulence. *Journal of the atmospheric sciences* 43(9):857–870
- Michel Y (2013a) Estimating deformations of random processes for correlation modelling in a limited area model. *Quarterly Journal of the Royal Meteorological Society* 139(671):534–547
- Michel Y (2013b) Estimating deformations of random processes for correlation modelling: methodology and the one-dimensional case. *Quarterly Journal of the Royal Meteorological Society* 139(672):771–783
- Mikulevicius R, Rozovskii B (2004) Stochastic Navier–Stokes equations for turbulent flows. *SIAM Journal on Mathematical Analysis* 35(5):1250–1310
- Mitchell L, Gottwald G (2012) Data assimilation in slow-fast systems using homogenized climate models. *Journal of the atmospheric sciences* 69(4):1359–1377
- Neves W, Olivera C (2015) Wellposedness for stochastic continuity equations with ladyzhenskaya–prodi–serrin condition. *Nonlinear Differential Equations and Applications NoDEA* 22(5):1247–1258
- Ng GH, McLaughlin D, Entekhabi D, Ahanin A (2011) The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A* 63(5):958–977
- Orszag S (1970) Analytical theories of turbulence. *Journal of Fluid Mechanics* 41(02):363–386



- Pannekoucke O, Massart S (2008) Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Quarterly Journal of the Royal Meteorological Society* 134(635):1425–1438
- Pannekoucke O, Emili E, Thual O (2014) Modelling of local length-scale dynamics and isotropizing deformations. *Quarterly Journal of the Royal Meteorological Society* 140(681):1387–1398
- Papadakis N, Mémin E, Cuzol A, Gengembre N (2010) Data assimilation with the weighted ensemble Kalman filter. *Tellus A* 62(5):673–697
- Papanicolaou G, Kohler W (1974) Asymptotic theory of mixing stochastic ordinary differential equations. *Communications on Pure and Applied Mathematics* 27(5):641–668
- Parzen E (1962) On estimation of a probability density function and mode. *The annals of mathematical statistics* 33(3):1065–1076
- Peavoy D, Franzke C, Roberts G (2015) Systematic physics constrained parameter estimation of stochastic differential equations. *Computational Statistics & Data Analysis* 83:182–199
- Pegion P, Bates GM G, Hamill T, Kolczynski W, Whitaker J, Zhu Y (2016) Stochastic parametrization development in the noaa/ncep global forecast system. In: *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading
- Penland C (2003a) Noise out of chaos and why it won't go away. *Bulletin of the American Meteorological Society* 84(7):921
- Penland C (2003b) A stochastic approach to nonlinear dynamics: A review (extended version of the article-” noise out of chaos and why it won't go away”). *Bulletin of the American Meteorological Society* 84(7):925–925
- Penland C, Matrosova L (1994) A balance condition for stochastic numerical models with application to the El Nino-southern oscillation. *Journal of climate* 7(9):1352–1372
- Penland C, Sardeshmukh P (1995) The optimal growth of tropical sea surface temperature anomalies. *Journal of climate* 8(8):1999–2024
- Pinier B, Mémin E, Laizet S, Lewandowski R (2019) Stochastic flow approach to model the mean velocity profile of wall-bounded flows, DOI 10.1103/PhysRevE.99.063101, URL <https://link.aps.org/doi/10.1103/PhysRevE.99.063101>
- Plant R, Craig GC (2008) A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences* 65(1):87–105
- Pope S (1983) A lagrangian two-time probability density function equation for inhomogeneous turbulent flows. *The Physics of fluids* 26(12):3448–3450
- Pope S (1994) Lagrangian PDF methods for turbulent flows. *Annu Rev Fluid Mech* 26:23–63
- Pope S, Chen Y (1990) The velocity-dissipation probability density function model for turbulent flows. *Physics of Fluids A: Fluid Dynamics* 2(8):1437–1449
- Prévôt C, Röckner M (2007) *A concise course on stochastic partial differential equations*, vol 1905. Springer
- Raynaud L, Boisserie M, Bouttier DL F, Ladabie C, Michel Y (2016) Model error representation in meteo france ensemble nwp system. In: *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading
- Resseguier V (2017) *Mixing and fluid dynamics under location uncertainty*. PhD thesis, Université Rennes 1
- Resseguier V, Mémin E, Chapron B (2015) Reduced flow models from a stochastic Navier-Stokes representation. *Annales de l'ISUP*
- Resseguier V, Mémin E, Chapron B (2017a) Geophysical flows under location uncertainty, part I: Random transport and general models. *Geophysical & Astrophysical Fluid Dynamics* 111(3):149–176
- Resseguier V, Mémin E, Chapron B (2017b) Geophysical flows under location uncertainty, part II: Quasi-geostrophic models and efficient ensemble spreading. *Geophysical & Astrophysical Fluid Dynamics* 111(3):177–208
- Resseguier V, Mémin E, Heitz D, Chapron B (2017c) Stochastic modelling and diffusion modes for POD models and small-scale flow analysis. *J of Fluid Mech* In press
- Resseguier V, Pan W, Fox-Kemper B (2019) Data-driven versus self-similar parameterizations for stochastic lie transport and location uncertainty Submitted, to be published in arXiv soon

- Reynolds C, Leutbecher M, Batt L, Chen S, Christensen H, Klasa C, Pegion P, Plant B, Raynaud L, Roberts N, Sandu I, Singleton A, Sommer M, Swinbank R, Tennant W, Theis S (2016) Reports from working group 3 : What are the pros/cons of existing model uncertainty schemes and how should these be measured? In: ECMWF/WWRP Workshop: Model Uncertainty, ECMWF, Reading
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27(3):832–837
- Sakov P, Haussaire JM, Bocquet M (2017) An iterative ensemble kalman filter in the presence of additive model error. *Quarterly Journal of the Royal Meteorological Society*
- Sapsis T (2013) Attractor local dimensionality, nonlinear energy transfers and finite-time instabilities in unstable dynamical systems with applications to two-dimensional fluid flows. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 469(2153), DOI 10.1098/rspa.2012.0550
- Sapsis T, Lermusiaux P (2012) Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty. *Physica D: Nonlinear Phenomena* 241(1):60–76
- Sapsis T, Majda A (2013a) Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems. *Journal of Nonlinear Science* 23(6):1039–1071
- Sapsis T, Majda A (2013b) Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences* 110(34):13705–13710
- Sapsis T, Majda A (2013c) A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems. *Physica D: Nonlinear Phenomena* 252:34–45
- Sardeshmukh PD, Sura P (2009) Reconciling non-gaussian climate statistics with linear dynamics. *Journal of Climate* 22(5):1193–1207
- Sardeshmukh PD, Compo GP, Penland C (2015) Need for caution in interpreting extreme weather statistics. *Journal of Climate* 28(23):9166–9187
- Sawford B (1991) Reynolds number effects in lagrangian stochastic models of turbulent dispersion. *Physics of Fluids A: Fluid Dynamics* 3(6):1577–1586
- Scheuerer M, Hamill TM (2015) Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review* 143(4):1321–1334
- Schumann U (1995) Stochastic backscatter of turbulence energy and scalar variance by random subgrid-scale fluxes. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, vol 451, pp 293–318
- Separovic L, Charron M, Erfani A, Gagnon N, Zadra A, Vaillancourt P (2016) Model error representation in the canadian ensemble prediction systems. In: *ECMWF/WWRP Workshop: Model Uncertainty*, ECMWF, Reading
- Shutts G (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* 612:3079–3012
- Shutts G (2015) A stochastic convective backscatter scheme for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* 141(692):2602–2616
- Smagorinsky J (1963) General circulation experiments with the primitive equation: I. the basic experiment. *Monthly Weather Review* 91:99–165
- Snyder C, Bengtsson T, Morzfeld M (2015) Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review* 143:4750–4761
- Talagrand O (1999) Evaluation of probabilistic prediction systems. In: *Workshop proceedings "Workshop on predictability"*, 20-22 October 1997, ECMWF, Reading, UK
- Tandeo P, Ailliot P, Bocquet M, Carrassi A, Miyoshi T, Pulido M, Zhen Y (2018) Joint estimation of model and observation error covariance matrices in data assimilation: a review. *arXiv preprint arXiv:180711221*
- Thorarinsdottir TL (2017) Verification: assessment of calibration and
- Trevisan A, Palatella L (2011) On the Kalman filter error covariance collapse into the unstable subspace. *Nonlinear Processes in Geophysics* 18(2):243–250
- Trevisan A, Uboldi F (2004) Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle system. *Journal of the atmospheric sciences* 61(1):103–113

- 
- Vallis G (2006) Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation. Cambridge University Press
- Veneziani M, Griffa A, Reynolds A, Mariano A (2004) Oceanic turbulence and stochastic models from subsurface Lagrangian data for the Northwest Atlantic Ocean. *Journal of physical oceanography* 34(8):1884–1906
- Williams CK, Rasmussen CE (2006) Gaussian processes for machine learning. the MIT Press 2(3):4
- Wouters J, Gottwald GA (2018) Stochastic model reduction for slow-fast systems with moderate time-scale separation. arXiv preprint arXiv:180409537
- Wouters J, Lucarini V (2012) Disentangling multi-level systems: averaging, correlations and memory. *Journal of Statistical Mechanics: Theory and Experiment* 2012(03):P03003
- Yang Y (2014) Study of variational ensemble methods for image assimilation. PhD thesis, Rennes 1
- Yang Y, Mémin E (2017) High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4denvar. *Tellus A: Dynamic Meteorology and Oceanography* 69(1):1308772
- Yang Y, Mémin E (2018) Estimation of physical parameters under location uncertainty using an ensemble 2-expectation-maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*