



HAL
open science

Maximum likelihood estimation of subgrid flows from tracer image sequences

Valentin Resseguier

► **To cite this version:**

Valentin Resseguier. Maximum likelihood estimation of subgrid flows from tracer image sequences. STUOD 2023, 2024. hal-04549000

HAL Id: hal-04549000

<https://hal.science/hal-04549000v1>

Submitted on 16 Apr 2024

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.

Maximum likelihood estimation of subgrid flows from tracer image sequences

Valentin Resseguier¹

Acta team, OPAALE, INRAE,
valentin.resseguier@inrae.fr,
home page: <https://www.valentinresseguier.com>

Abstract. From a sequence of tracer satellite images, several methods (e.g. optical flow) exist to successfully estimate the main advecting current. Yet, this estimate is limited in resolution. To go beyond, we propose a new parametric estimation method to estimate second-order statistics of the residual small-scale velocity. We first express stochastic transport in a discrete setting to apply standard MLE techniques. Then we propose an efficient method to solve the MLE optimization problem through a fast log-likelihood gradient evaluation algorithm.

1 Model and assumptions

1.1 Objectives

Let us consider several observed oceanic tracers q_i :

$$\frac{Dq_i}{Dt} = \dot{Q}_i, \quad (1.1)$$

where \dot{Q}_i is a source term smooth in time (e.g. $\dot{Q}_i = \nu_i \Delta q_i$). For instance, q_1 would be the SST and q_2 the SSS on a bounded spatial domain $\Omega \subset \mathbb{R}^2$. We can probably use Ocean colour or SSH as well. We assume that we can observe a set of snapshots of these tracers $(q_i(\mathbf{x}, t_k))_{1 \leq k \leq N}$. We aim at estimate the two-dimensional velocity field \mathbf{v} – which transports the tracer – from those tracer snapshots.

First, we use a classical method (e.g. optical flow Corpetti et al, 2006; Dérian et al, 2013; Ghahenoei et al, 2015; Sun et al, 2016) to estimate a (two-dimensional) velocity field from the tracer snapshots q_i . Note that the velocity estimate is probably different for each tracer q_i , that is why we denote it \mathbf{w}_i . Say that we have S tracers, we can compute the mean drift estimate $\tilde{\mathbf{w}} = \frac{1}{S} \sum_i \mathbf{w}_i$. This estimate is hopefully accurate but limited in resolution, typically by the resolution of tracer images and by the optical flow algorithm efficiency. Therefore, we refer to this term as large-scale velocity component. This optical flow procedure is obviously of main importance but we do not address it here. A large literature already deals with it. We assume that an optical flow algorithm – says the most efficient optical flow algorithm of the literature – is applied before our method comes into play.

Then, we note that the tracer is also transported by a small-scale velocity $\mathbf{v}' = \mathbf{v} - \tilde{\mathbf{w}}$. The paper focuses on this residual velocity field. We aim at estimating it or at least estimating its statistics. For this purpose, we make use of modern machine learning and statistics informed by physics. Specifically, we consider maximum

likelihood estimation, stochastic calculus, and processes statistics (Rao, 1999; Sørensen, 2004; van Waaij and van Zanten, 2016) guided by stochastic fluid dynamics (Kraichnan, 1994; Piterbarg and Ostrovskii, 1997; Mémin, 2014; Holm, 2015; Resseguier et al, 2020a; Zhen et al, 2023). At long term we expect that our work will benefit for the recent advances in processes statistics for linear and linear SPDE with additive and multiplicative noises (Pasemann and Stannat, 2020; Altmeyer and Reiß, 2021; Altmeyer et al, 2022, 2023; Gaudlitz and Reiß, 2023; Janák and Reiß, 2023).

1.2 Simplifications of the problem

Then, we neglect its time correlations of the residual velocity field. This assumption is supported by the fact that this velocity is small-scale. The time step Δt between two tracer snapshots is finite, and possibly larger than the small-scale velocity correlation time. Therefore, it is probably hopeless to estimate a time-correlated small-scale velocity field. The best we can do is probably estimating the statistics of a time-subsampling version of the small-scale velocity (subsamped at the time step Δt). And this time-subsampling version is time-uncorrelated if Δt is larger than the correlation time of the true small-scale velocity.

We will also assume that \mathbf{v}' is Gaussian, homogeneous and isotropic in space. Therefore, we can parameterize \mathbf{v}' as the spatial convolution of space-time white noise:

$$\mathbf{v}' = \check{\boldsymbol{\sigma}} * \dot{B}, \quad (1.2)$$

where $*$ denotes the 2-dimensional spatial convolution, $\check{\boldsymbol{\sigma}}$ is a 2-dimensional vector of spatial filters and \dot{B} is space-time white noise. The spatial filter imposes a spatial correlation. The covariance of \mathbf{v}' is:

$$\mathbb{E}\{\mathbf{v}'(\mathbf{x}, t_1)\mathbf{v}'(\mathbf{y}, t_2)\} = \frac{1}{dt_1}\delta(t_1 - t_2)(\check{\boldsymbol{\sigma}} * \check{\tilde{\boldsymbol{\sigma}}})(\mathbf{x} - \mathbf{y}), \quad (1.3)$$

where $\check{\tilde{\boldsymbol{\sigma}}}(\mathbf{x}) \triangleq \check{\boldsymbol{\sigma}}(-\mathbf{x})$. To simplify the notations, we will denote:

$$\mathbf{a}(\mathbf{x} - \mathbf{y}) \triangleq \frac{1}{dt} \mathbb{E}\{(\check{\boldsymbol{\sigma}} * dB_t)(\mathbf{y} + \mathbf{x})(\check{\boldsymbol{\sigma}} * dB_t)^T(\mathbf{y})\} = (\check{\boldsymbol{\sigma}} * \check{\tilde{\boldsymbol{\sigma}}})(\mathbf{x} - \mathbf{y}) \quad (1.4)$$

Under some assumptions on the filter $\check{\boldsymbol{\sigma}}$, we can show that : $\mathbf{a}(0) = a_0\mathbb{I}_d$ where $a_0 = \frac{1}{2}\text{tr}(\mathbf{a}(0)) = \frac{1}{2}\|\check{\boldsymbol{\sigma}}\|_{\mathcal{L}^2(\Omega)}^2$ is a positive constant, sometimes called absolute diffusivity, Kubo-type formula or variance tensor Mémin (2014). It is equal to the variance of the small-scale velocity multiplied by its correlation time.

1.3 Stochastic transport

Since \mathbf{v}' is assumed time-uncorrelated, the transport equation (1.1) can be interpreted as a dynamic under Location Uncertainty (LU) (Mémin, 2014; Resseguier et al, 2017a) or Stochastic Advection by Lie Transport (SALT) (Holm, 2015). With Itô notations, it reads:

$$\partial_t q_i + (\tilde{\mathbf{w}} + \mathbf{v}') \cdot \nabla q_i \approx \partial_t q_i + (\mathbf{w}_i + \mathbf{v}') \cdot \nabla q_i = \frac{a_0}{2} \Delta q_i + \dot{Q}_i. \quad (1.5)$$

The term $\mathbf{v}' \cdot \nabla q_i$ acts as a time-uncorrelated random forcing. Eulerian stochastic transport equations always involve the Lagrangian displacement Stratonovich drift as advecting velocity. This is true for both

Itô and Stratonovich notations of the Eulerian SPDEs and for both SALT and LU (Resseguier et al, 2020b). That Stratonovich drift corresponds to the Itô drift plus a possible correction. Since the optical flow will estimate the advecting velocity, we identify \mathbf{w}_i as the Lagrangian displacement Stratonovich drift. Note that all stochastic differential equations of this paper are expressed with Itô notations. We refer to Resseguier et al (2020b) for a comparison between SALT and LU, and to Resseguier et al (2020a) for a review of SALT/LU models and calibration methods. Note that Kraichnan (1994); Piterbarg and Ostrovskii (1997) and references therein have also studied in details stochastic transport of passive tracers by delta-correlated velocities.

2 Quadratic co-variation for turbulence amplitude estimation

We aim at estimating the statistics of \mathbf{v}' . In this section, we treat the estimation of the variance tensor a_0 . The random forcing $\mathbf{v}' \cdot \nabla q_i$ being delta-correlated, we can estimate the variance tensor from the following algorithm.

We compute for every point \mathbf{x} of the grid:

1. $dq_i(\mathbf{x}, t) = q_i(\mathbf{x}, t + \Delta t) - q_i(\mathbf{x}, t)$, the tracer time increments,
2. $\mathbf{w}_i(\mathbf{x}, t)$ from an optical flow algorithm,
3. $d\tilde{q}_i(\mathbf{x}, t) = dq_i(\mathbf{x}, t) - \mathbf{w}_i \cdot \nabla q_i \Delta t$,
At this step, we can also subtract some known source terms \dot{Q}_i if any.
4. $\overline{d\tilde{q}_i}(\mathbf{x}, t)$ a local time average of $d\tilde{q}_i(\mathbf{x}, t)$,
5. $dq'_i(\mathbf{x}, t) = d\tilde{q}_i(\mathbf{x}, t) - \overline{d\tilde{q}_i}(\mathbf{x}, t)$.

This step should subtract the effect of the unknown smooth forcing \dot{Q}_i . It acts as a high-pass filter to keep only the highly oscillating components of $d\tilde{q}_i$. According to the stochastic transport equation (1.5), we should have:

$$dq'_i(\mathbf{x}, t) \approx -(\boldsymbol{\sigma} * \Delta B) \cdot \nabla q_i + (\text{something small}) \times \Delta t \quad (2.1)$$

with $\Delta B = B_{t+\Delta t} - B_t \propto \sqrt{\Delta t}$ a Brownian increment.

6. At this point, we may check – by usual statistical tests – that the increments $dq'_i(\mathbf{x}, t)$ is approximately time uncorrelated.

In the LU-SALT theoretical framework, the delta correlation of these increments comes from the delta correlation of the subgrid velocity \mathbf{v}' . This model assumption is consistent with the fact that the subgrid velocity has short correlation time. However, in practice, the subgrid velocity correlation time is finite. It is a recurrent issue for the data driven modeling of systems combining fast and slowly evolving components (Pavliotis and Stuart, 2007; Papavasiliou et al, 2009; Cotter and Pavliotis, 2009; Azencott et al, 2010, 2013; Reich, 2023). If spurious correlations remain among the increments, we can down-sample the data to force the noise terms to be as decorrelated as possible. The literature proposes several time subsampling rate, generally related to the correlation time of those increments. Resseguier et al (2021)

applied this method in a LU context. They estimate a minimal time subsampling rate from the empirical time correlation function of the subgrid velocity. Here, the empirical time correlation function of the increments dq'_i could be used instead.

7. Now, we can compute what is called the quadratic co-variation of q_i and q_j , denoted $\langle q_i, q_j \rangle$, in stochastic calculus:

$$\langle q_i, q_j \rangle(\mathbf{x}) = \sum_k dq'_i(\mathbf{x}, t_k) dq'_j(\mathbf{x}, t_k). \quad (2.2)$$

8. We also compute the tracer gradients cross-correlations:

$$c_{ij} = \int_0^{N\Delta t} \nabla q_i \cdot \nabla q_j dt \quad (2.3)$$

9. Finally, we obtain the variance tensor by a simple (overdetermined) linear system:

$$a_0 \approx \frac{\sum_{ijp} c_{ij}(\mathbf{x}_p) \langle q_i, q_j \rangle(\mathbf{x}_p)}{\sum_{ijq} c_{ij}^2(\mathbf{x}_p)}. \quad (2.4)$$

Indeed, the decorrelation between two time increments yields:

$$\langle q_i, q_j \rangle \approx \int_0^{N\Delta t} \nabla q_i^T (\boldsymbol{\sigma} * \Delta B) (\boldsymbol{\sigma} * \Delta B)^T \nabla q_j \approx \int_0^{N\Delta t} \nabla q_i^T \mathbf{a}(0) \nabla q_j dt = a_0 c_{ij}. \quad (2.5)$$

Theoretically, only one tracer is needed here, even though we expect a higher accuracy with several observed fields.

3 Parametric model for the small-scale velocity statistics

The variance tensor, a_0 , gives the "amplitude" of the small-scale velocity. But, we may want more information (e.g. correlation length, covariance or spectrum). For this purpose, we propose in this section a parametric model for the spatial "covariance", $a(\mathbf{x} - \mathbf{y})$, of that velocity introduced in equation (1.4)). This model will depend on the variance tensor a_0 (the "variance") and on some other parameters θ_r . Section 4 will propose a method to estimate these parameters by maximum likelihood. Several choices of parametric covariance are possible (e.g., Gaussian or Matérn covariance). Here, we propose a self-similar model for turbulence statistics inline with previous work related to stochastic transport (Kraichnan, 1994; Mémin, 2014; Resseguier et al, 2017b, 2020b).

A representation in Fourier space will be convenient for our ultimate estimation procedure. Hence we limit the present study to fields with periodic boundary conditions and we introduce the unitary Fourier transform $\widehat{\boldsymbol{\zeta}}(\mathbf{k}) = \int_{\Omega} d\mathbf{x} \boldsymbol{\zeta}(\mathbf{x}) e^{-2i\pi \mathbf{k} \cdot \mathbf{x}}$ for any function $\boldsymbol{\zeta}$.

3.1 Spectrum matrix for divergence-free velocity

In order to simplify the homogeneous model (1.2) while enforcing the divergence-free and spatial stationarity constraints, we can define the small-scale velocity $\boldsymbol{\sigma}dB_t$ with its streamfunction:

$$\boldsymbol{\sigma}dB_t = \nabla^\perp \psi_\sigma dB_t = \nabla^\perp \check{\psi}_\sigma * dB_t, \quad (3.1)$$

where $*$ denote a spatial convolution and ∇^\perp the two-dimensional curl. Furthermore, we consider a Matérn for the streamfunction covariance:

$$\gamma_{\psi_\sigma}(\mathbf{x}) = \frac{1}{dt} \mathbb{E} \{ (\psi_\sigma(\mathbf{y} + \mathbf{x}) dB_t) (\psi_\sigma(\mathbf{y}) dB_t) \} = \left(\check{\psi}_\sigma * \check{\psi}_\sigma \right) (\mathbf{x}) = D g_{\frac{\beta+1}{2}}(2\pi\kappa_m \|\mathbf{x}\|), \quad (3.2)$$

where $g_\nu(r) = r^\nu \mathcal{K}_\nu(r)$, \mathcal{K}_ν is the modified Bessel function of second kind, $1/2\pi\kappa_m$ is the correlation length, and D is a constant defined in Appendix. We will show further below that $-\beta$ is the velocity spectrum slope. This covariance choice is physically relevant since it highlights an important symmetry of turbulence: the self-similar distribution of energy. Indeed, the corresponding streamfunction spectrum is Williams and Rasmussen (2006); Lim and Teo (2009):

$$S_{\psi_\sigma}(\mathbf{k}) = \left| \widehat{\check{\psi}_\sigma}(\mathbf{k}) \right|^2 = \widehat{\gamma}_{\psi_\sigma}(\mathbf{k}) = S_{\psi_\sigma}(0) \left(1 + \left(\frac{k}{\kappa_m} \right)^2 \right)^{-\frac{\beta+3}{2}} \quad (3.3)$$

with a constant $S_{\psi_\sigma}(0)$ defined in Appendix and the wavenumber $k = \|\mathbf{k}\|$. The small-scale velocity spectrum matrix is

$$\widehat{\mathbf{a}}(\mathbf{k}) = \frac{1}{dt} \mathbb{E} \left\{ \widehat{(\boldsymbol{\sigma}dB_t)}(\mathbf{k}) \widehat{(\boldsymbol{\sigma}dB_t)}^H(\mathbf{k}) \right\} = (2\pi i \mathbf{k}^\perp) (2\pi i \mathbf{k}^\perp)^H S_{\psi_\sigma} = \mathbf{J} (2\pi \mathbf{k}) (2\pi \mathbf{k})^T \mathbf{J}^T S_{\psi_\sigma}, \quad (3.4)$$

where \mathbf{J} is the matrix which performs a $\frac{\pi}{2}$ -rotation. Equation (3.3) confirms that the velocity spectrum slope is $-\beta$. Note that $\widehat{\mathbf{a}}(\mathbf{k})$ is of rank 1. Therefore, it is not invertible and even its pseudo inverse, $\widehat{\mathbf{a}}(\mathbf{k})^\dagger$, is not defined. This singularity will induce major difficulties in the following estimation procedure (see section 4 below). In other words, our methodology cannot be applied with solely a solenoidal small-scale velocity. A workaround is the consideration of a divergent component.

3.2 Spectrum matrix from Helmholtz decomposition of the small-scale velocity

Now, we do not impose divergence free anymore. From Helmholtz-Hodge theorem, we can write the hidden small-scale velocity as a sum of a solenoidal and a potential components:

$$\boldsymbol{\sigma} \dot{B} = \nabla^\perp \check{\psi}_\sigma * \dot{B}^{\nabla^\perp} + \nabla \check{\phi}_\sigma * \dot{B}^\nabla, \quad (3.5)$$

where \dot{B}^{∇^\perp} and \dot{B}^∇ are two independent white noises. Here, we have implicitly assumed that the two components are independent from one another. We stick to Matérn covariances, using it from both compo-

nents:

$$S_{\psi_\sigma}(\mathbf{k}) = S_{\psi_\sigma}(0) \left(1 + \left(\frac{k}{\kappa_m^{\nabla^\perp}} \right)^2 \right)^{-\frac{\beta^{\nabla^\perp} + 3}{2}}, \quad (3.6)$$

$$S_{\phi_\sigma}(\mathbf{k}) = S_{\phi_\sigma}(0) \left(1 + \left(\frac{k}{\kappa_m^\nabla} \right)^2 \right)^{-\frac{\beta^\nabla + 3}{2}}. \quad (3.7)$$

To simplify, we may choose $\kappa_m^{\nabla^\perp} = \kappa_m^\nabla = \kappa_m$ and set $2\pi\kappa_m$ to be the smallest well resolved scale contained in the large-scale velocities $\tilde{\mathbf{w}}$. Then, the diagonalization of the spectrum matrix is straightforward :

$$\hat{\mathbf{a}}(\mathbf{k}) = (2\pi\mathbf{k}^\perp)(2\pi\mathbf{k}^\perp)^T S_{\psi_\sigma} + (2\pi\mathbf{k})(2\pi\mathbf{k})^T S_{\phi_\sigma} = \mathbf{K}\mathbf{S}\mathbf{K}^T, \quad (3.8)$$

denoting $\mathbf{K} = \begin{bmatrix} \tilde{\mathbf{k}} & \tilde{\mathbf{k}}^\perp \end{bmatrix}$, $\mathbf{S} = \begin{bmatrix} S^\nabla & 0 \\ 0 & S^{\nabla^\perp} \end{bmatrix} = \begin{bmatrix} (2\pi k)^2 S_{\phi_\sigma} & 0 \\ 0 & (2\pi k)^2 S_{\psi_\sigma} \end{bmatrix}$, with $\tilde{\mathbf{k}} = \mathbf{k}/k$, $S^\nabla = (2\pi k)^2 S_{\phi_\sigma}$, and $S^{\nabla^\perp} = (2\pi k)^2 S_{\psi_\sigma}$ the normalized wave-vector, the divergent and the solenoidal subgrid velocity spectra respectively. Note that the spectrum matrix is now full rank.

3.3 Velocity covariance

Once the parameters optimized, we can use of the know subgrid spectra (3.8) and covariance:

$$\mathbf{a}(\mathbf{x}) = \mathbf{J}\mathbf{H}_{\gamma_{\psi_\sigma}}\mathbf{J}^T + \mathbf{H}_{\gamma_{\phi_\sigma}}, \quad (3.9)$$

$$\begin{aligned} &= (2\pi\kappa_m)^2 D^{\nabla^\perp} \left(g_{\frac{\beta^{\nabla^\perp}-1}{2}}(2\pi\kappa_m\|\mathbf{x}\|) \mathbb{I}_d - g_{\frac{\beta^{\nabla^\perp}-3}{2}}(2\pi\kappa_m\|\mathbf{x}\|) ((\mathbf{x}^\perp)(\mathbf{x}^\perp)^T) \right) \\ &+ (2\pi\kappa_m)^2 D^\nabla \left(g_{\frac{\beta^\nabla-1}{2}}(2\pi\kappa_m\|\mathbf{x}\|) \mathbb{I}_d - g_{\frac{\beta^\nabla-3}{2}}(2\pi\kappa_m\|\mathbf{x}\|) (\mathbf{x}\mathbf{x}^T) \right), \end{aligned} \quad (3.10)$$

with constants D^{∇^\perp} , D^∇ defined in Appendix as functions of the respective variance tensors $a_0^{\nabla^\perp}$ and a_0^∇ , turbulence kinetic energy spectrum slope β^{∇^\perp} and β^∇ , and the correlation length $1/\kappa_m$. For synthetic notations, we also introduce $e^\alpha = a_0^{\nabla^\perp}/a_0^\nabla$ the variance ratio between the solenoidal and potential components.

3.4 Parametric model summary

We have introduced a parametric model for the subgrid velocity depending on the variance tensor a_0 and 3 other parameters $\boldsymbol{\theta} = (\beta^{\nabla^\perp}, \beta^\nabla, \alpha)$. Finally, our parametric model can be summarized as follow:

$$\hat{\mathbf{a}}(\mathbf{k}) = \mathbf{K}\mathbf{S}\mathbf{K}^T = \begin{bmatrix} \tilde{\mathbf{k}} & \tilde{\mathbf{k}}^\perp \end{bmatrix} \begin{bmatrix} S^\nabla(\mathbf{k}) & 0 \\ 0 & S^{\nabla^\perp}(\mathbf{k}) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{k}} & \tilde{\mathbf{k}}^\perp \end{bmatrix}^T, \quad (3.11)$$

$$S^\nabla(\mathbf{k}) = a_0 \left(\frac{1}{1+e^\alpha} \right) \left(\frac{2\pi((\beta^\nabla)^2 - 1)}{(2\pi\kappa_m)^2} \right) \left(\frac{k}{\kappa_m} \right)^2 \left(1 + \left(\frac{k}{\kappa_m} \right)^2 \right)^{-\frac{\beta^\nabla + 3}{2}}, \quad (3.12)$$

$$S^{\nabla^\perp}(\mathbf{k}) = a_0 \left(\frac{e^\alpha}{1+e^\alpha} \right) \left(\frac{2\pi((\beta^{\nabla^\perp})^2 - 1)}{(2\pi\kappa_m)^2} \right) \left(\frac{k}{\kappa_m} \right)^2 \left(1 + \left(\frac{k}{\kappa_m} \right)^2 \right)^{-\frac{\beta^{\nabla^\perp} + 3}{2}}. \quad (3.13)$$

Such parametrization naturally ensures that the total variance is always a_0 and that the solenoidal and potential variances are $a_0^{\nabla^\perp} = a_0 \left(\frac{e^\alpha}{1+e^\alpha} \right)$ and $a_0^\nabla = a_0 \left(\frac{1}{1+e^\alpha} \right)$ respectively, both restricted to $[0, a_0]$.

4 Maximum likelihood estimation of turbulence correlations

The variance tensor a_0 (the "variance") can be estimated by quadratic variation as explained in section 2. The other parameters $\boldsymbol{\theta}$ will be estimated by maximum likelihood. However, due to the multiplicative structure of the random forcing in (1.5), the tracers q_i are not Gaussian. So, the tracers likelihood is not trivial. Fortunately, the Girsanov theorem gives the expression of the mutual likelihood of the whole set $(q_i(\mathbf{x}_q, t_k))_{\substack{1 \leq q \leq M \\ 1 \leq k \leq N}}$. Note that most of the results of this section are relatively general. They can be applied to most homogeneous subgrid spatial covariance parametrisations in stochastic transport contexts.

Piterbarg and Rozovskii (1996); Piterbarg and Ostrovskii (1997) have already proposed surface current estimations from satellite image by maximum likelihood estimation. However, their algorithm estimates a constant velocity field only. Then, they repeat the operation on a multitude of patches to eventually obtain a gridded velocity field. We believe that such coarse-grid velocity field is probably better estimated by state-of-the-art optical flow methods, and we do not address this issue here. We rather try to extract additional statistical information from the residue of coarse-scale current estimations.

4.1 Girsanov theorem

In the literature of processes statistics, a lot of parametric estimation methods rely on likelihood. Indeed, denoting $\boldsymbol{\theta}$ the parameters, even for a non-Gaussian process $\mathbf{Q}(t)$, such as

$$\frac{d}{dt}\mathbf{Q}(t) = \mathbf{F}(\mathbf{Q}(t)|\boldsymbol{\theta}) + \mathbf{G}(\mathbf{Q}(t)|\boldsymbol{\theta})\dot{\mathbf{W}}, \quad (4.1)$$

where $\dot{\mathbf{W}}$ is a vector of independent white noise, there is a simple expression of the joint likelihood, $p(\{\mathbf{Q}_t | 0 \leq t \leq N\Delta t\} | \boldsymbol{\theta})$. Here above $\dot{\mathbf{W}}$ is to be interpreted in a Itô sense. We assume that $\mathbf{G}\mathbf{G}^T$ is invertible where $\mathbf{G} = \mathbf{G}(\mathbf{Q}(t)|\boldsymbol{\theta})$. We will discuss the validity of this assumption later in the paper. The Girsanov theorem (Oksendal, 1998) leads as explained in Rao (1999) (Eq. (3.3.2) page 147) to the following log-likelihood:

$$\begin{aligned} l(\{\mathbf{Q}(t) | 0 \leq t \leq N\Delta t\} | \boldsymbol{\theta}) &= \int_0^{N\Delta t} \log |\boldsymbol{\Sigma}(\mathbf{Q}(t)|\boldsymbol{\theta})|^{-\frac{1}{2}} dt + \int_0^{N\Delta t} (\mathbf{F}^T \boldsymbol{\Sigma}^{-1})(\mathbf{Q}(t)|\boldsymbol{\theta}) d\mathbf{Q}(t) \\ &\quad - \frac{1}{2} \int_0^{N\Delta t} (\mathbf{F}^T \boldsymbol{\Sigma}^{-1} \mathbf{F})(\mathbf{Q}(t)|\boldsymbol{\theta}) dt. \end{aligned} \quad (4.2)$$

where

$$\boldsymbol{\Sigma}^{-1} = \mathbf{G}[[\mathbf{G}^T \mathbf{G}]^\dagger]^2 \mathbf{G}^T = \mathbf{G}[[\mathbf{G}^T \mathbf{G}]^\dagger]^2 \mathbf{G}^T [\mathbf{G}\mathbf{G}^T][\mathbf{G}\mathbf{G}^T][\mathbf{G}\mathbf{G}^T]^{-2} = \mathbf{G}\mathbf{G}^T [\mathbf{G}\mathbf{G}^T]^{-2} = [\mathbf{G}\mathbf{G}^T]^{-1}, \quad (4.3)$$

i.e. $\boldsymbol{\Sigma} = \mathbf{G}\mathbf{G}^T = [\mathbf{G}\mathbf{G}^T](\mathbf{Q}(t)|\boldsymbol{\theta})$ is the noise conditional covariance given the current state $\mathbf{Q}(t)$ and the parameter $\boldsymbol{\theta}$. Here, we have added the normalizing constant logarithm, $\int_0^{N\Delta t} \log |\boldsymbol{\Sigma}|^{-\frac{1}{2}} dt$, since $\boldsymbol{\Sigma}$ depends on the parameters $\boldsymbol{\theta}$ to be estimated.

Note that the Girsanov theorem does not give us the conditional probability density function of $\mathbf{Q}(t)$ at time t but only the joint probability density function of $\{\mathbf{Q}(t)|0 \leq t \leq t\}$. The above formula is widely used to perform maximum likelihood estimations and Bayesian estimations in finance (Rao, 1999; Sørensen, 2004; van Waaij and van Zanten, 2016) and more recently to linear and nonlinear SPDE in biology and reaction-diffusion systems with additive noise (Pasemann and Stannat, 2020; Altmeyer and Reiß, 2021; Altmeyer et al, 2022, 2023; Gaudlitz and Reiß, 2023) and multiplicative noise (Janák and Reiß, 2023). Note that the Girsanov theorem is also valid in infinite dimension (Da Prato and Zabczyk, 1992; Liu and Röckner, 2015). (Janák and Reiß, 2023) treat the variance estimator aside, before performing MLE. This prevents theoretical estimation issues. That is why we operate similarly here: first estimating the variance tensor and then estimating the other covariance parameters.

As already discussed in section 2, another important issue to deal with is the finite time correlation time of the observed increments. It is a common problem that biases MLE (Pavliotis and Stuart, 2007; Papavasiliou et al, 2009; Cotter and Pavliotis, 2009; Azencott et al, 2010, 2013; Reich, 2023). As explained section 2, we address this issue by subsampling the data.

An alternative estimation method can also be derived from the discretized-in-time version of the stochastic differential equation (4.1). \mathbf{Q} is not Gaussian but $d\mathbf{Q}(t)$ and thus $\mathbf{Q}(t+dt)$ are conditionally Gaussian given $\mathbf{Q}(t)$ and $\boldsymbol{\theta}$. By factorizing the conditional Gaussian distribution from $t = 0$ to $t = N\Delta t$, we obtain a similar likelihood expression.

4.2 Application of the Girsanov theorem

After spatial discretization, the stochastic equation (1.5) reduces to the form (4.1). A sequence of satellite images of a tracer could hence be used to estimate a parametrization of the stochastic model. In this case,

$$\mathbf{Q}_{pi}(t) = q_i(\mathbf{x}_p, t), \quad (4.4)$$

$$\mathbf{F}_{pi}(t) = \left(-\mathbf{w}_i \cdot \nabla q_i + \frac{\alpha_0}{2} \Delta q_i + \dot{Q}_i \right) (\mathbf{x}_p, t), \quad (4.5)$$

$$(\mathbf{G}\dot{\mathbf{W}})_{pi}(t) = (-\mathbf{v}' \cdot \nabla q_i) (\mathbf{x}_p, t), \quad (4.6)$$

$$\Sigma_{pi,rj}(t) = (\mathbf{G}\mathbf{G}^T)_{pi,rj}(t) = \nabla q_i^T(\mathbf{x}_p, t) \mathbf{a}(\mathbf{x}_p - \mathbf{x}_r) \nabla q_j(\mathbf{x}_r, t), \quad (4.7)$$

$$\forall \mathbf{Q}', (\boldsymbol{\Sigma}\mathbf{Q}')_{pi}(t) = \nabla q_i^T(\mathbf{x}_p, t) \left(\mathbf{a} * \left(\sum_j \nabla q_j q'_j \right) \right) (\mathbf{x}_p, t) \text{ where } q'_j(\mathbf{x}_r) \triangleq \mathbf{Q}'_{jr} \quad (4.8)$$

In practice, we shall use finite-dimensional approximations for every calculations steps. For any d -dimensional vector fields $\boldsymbol{\zeta}$, we must define spatially-discretized version. We represent $\boldsymbol{\zeta}$ by $M \times d$ -dimensional vectors, \mathbf{Z} . More precisely, we denote $Z_{pi} = \zeta_i(\mathbf{x}_p)$, We may also give a matrix notation $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$ to the convolution.

$$\forall \mathbf{Z}, \mathbf{A}(\boldsymbol{\theta})\mathbf{Z} = \mathbf{a}(\boldsymbol{\theta}) * \boldsymbol{\zeta} \quad (4.9)$$

To simplify, we will consider periodic boundary conditions and a discretisation over a uniform spatial grid. It prevents technical problems with convolutions and Fourier transform. Then, we introduce the local matrix $\mathbf{Y}_{is}(\mathbf{x}, t) = \partial_{x_s} q_i(\mathbf{x}, t)$ and the associated block-diagonal matrix \mathbb{Y} for the point-wise application of the small matrix \mathbf{Y} :

$$\mathbb{Y}_{rj,pi} = \mathbb{Y}_{rj,pi}(\mathbf{Q}(t)) = Y_{ji}(\mathbf{x}_r, t)\delta_{rp} = \partial_i q_j(\mathbf{x}_r, t)\delta_{rp}. \quad (4.10)$$

We can eventually rewrite the operator $\Sigma = \Sigma(\mathbf{Q}(t)|\boldsymbol{\theta})$ as :

$$\forall \mathbf{Q}', \Sigma \mathbf{Q}' = \mathbb{Y} \mathbf{A}(\boldsymbol{\theta}) \mathbb{Y}^T \mathbf{Q}', \quad (4.11)$$

i.e. $\Sigma = \mathbb{Y} \mathbf{A}(\boldsymbol{\theta}) \mathbb{Y}^T$ where $\mathbb{Y} = \mathbb{Y}(\mathbf{Q}(t))$.

4.3 Inversion of the operator Σ

If the matrix \mathbf{Y} is everywhere sufficiently well conditioned (in particular if we observe enough tracers and if the fronts of different tracers are not aligned), we can locally consider the pseudo-inverse of \mathbf{Y} , that we will denote \mathbf{Y}^\dagger . Accordingly, we can obtain an approximate inverse of the operator Σ :

$$\Sigma^{-1} = (\mathbb{Y}^\dagger)^T \mathbf{A}^{-1}(\boldsymbol{\theta}) \mathbb{Y}^\dagger. \quad (4.12)$$

where $\mathbf{A}^{-1}(\boldsymbol{\theta})$ is a deconvolution operation. It can be computed in Fourier space or using other usual deconvolution methods. Needless to say that $\mathbb{Y}_{rj,pi}^\dagger = Y_{ji}^\dagger(\mathbf{x}_r, t)\delta_{rp}$ is also block-diagonal, enabling such large matrix computation in reasonable time. Using a given parametric form for \mathbf{a} (see section 3), we can compute \mathbf{A}^{-1} in Fourier space. For continuous Fourier transform, it would read:

$$\forall \hat{\zeta}, \widehat{[\mathbf{a}^*]^{-1} \zeta} = \widehat{[\hat{\mathbf{a}}^*]^{-1} \hat{\zeta}} = \hat{\mathbf{a}}^{-1} \hat{\zeta}, \quad (4.13)$$

assuming that the small-scale spectrum matrix $\hat{\mathbf{a}}$ has full rank. From (3.8), we have an explicit expression of the inverse:

$$\hat{\mathbf{a}}^{-1} = \mathbf{K} \mathbf{S}^{-1} \mathbf{K}^T. \quad (4.14)$$

In practice, we shall use the Fast Fourier Transform (FFT). For an uniform spatial grid of M points, we represent $\hat{\zeta}$ by $\hat{\mathbf{Z}} = \mathbf{P}^H \mathbf{Z}$. More precisely, we denote $\hat{Z}_{pi} = \hat{\zeta}_i(\mathbf{k}_p)$, $P_{pi,rj} = \delta_{ij} e^{2i\pi \mathbf{x}_p \cdot \mathbf{k}_r}$ and $\mathbf{P}^H = \overline{\mathbf{P}}^T$ its conjugate transpose, i.e. $P_{rj,pi}^H = \delta_{ij} e^{-2i\pi \mathbf{x}_p \cdot \mathbf{k}_r}$. The inverse discrete Fourier transform is defined by matrix $\frac{1}{M} \mathbf{P}$ since $(\mathbf{P}^H)^{-1} = \frac{1}{M} \mathbf{P}$. We can now express the deconvolution (4.13) with the block-diagonal matrix $\hat{\mathbf{A}}^{-1}(\boldsymbol{\theta}) = \mathbf{P}^{-1} \mathbf{A}^{-1}(\boldsymbol{\theta}) \mathbf{P}$

$$\forall \mathbf{Z}, \mathbf{P}^H (\mathbf{A}^{-1}(\boldsymbol{\theta}) \mathbf{Z}) = (\mathbf{P}^{-1} \mathbf{A}^{-1}(\boldsymbol{\theta}) \mathbf{P}) \hat{\mathbf{Z}} \quad (4.15)$$

where $\left(\widehat{\mathbf{A}}^{-1}(\boldsymbol{\theta})\right)_{rj,pi} = (\mathbf{P}^{-1}\mathbf{A}^{-1}(\boldsymbol{\theta})\mathbf{P})_{rj,pi} = \frac{1}{M}(\widehat{\mathbf{a}}^{-1})_{ji}(\mathbf{k}_r, \boldsymbol{\theta})\delta_{rp}$. Finally, from (4.12) we obtain a simple matrix form for the inverse covariance:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{M}(\mathbb{Y}^\dagger)^T \mathbf{P} \widehat{\mathbf{A}}^{-1}(\boldsymbol{\theta}) \mathbf{P}^H \mathbb{Y}^\dagger. \quad (4.16)$$

The efficiency of the FFT algorithm together with the block-diagonal structures of the other matrices ensure a low computational cost. With the expression of \mathbf{F} and $\boldsymbol{\Sigma}^{-1}$, we can now compute the expression of the log-likelihood (4.2).

4.4 Gradient of the likelihood

In order to estimate the parameters θ_r , we will need to maximize the log-likelihood, by e.g. gradient descent. Such algorithm necessitates the log-likelihood derivative along each parameters. Since \mathbf{F} does not depend on $\boldsymbol{\theta}$, we have $\partial_{\theta_r} l = \int_0^{N\Delta t} d\partial_{\theta_r} l$ with :

$$d\partial_{\theta_r} l = -\frac{1}{2}\partial_{\theta_r} \log |\boldsymbol{\Sigma}| dt + \mathbf{F}^T \partial_{\theta_r} (\boldsymbol{\Sigma}^{-1}) d\mathbf{Q}(t) - \frac{1}{2}\mathbf{F}^T \partial_{\theta_r} (\boldsymbol{\Sigma}^{-1}) \mathbf{F} dt, \quad (4.17)$$

where

$$\partial_{\theta_r} (\boldsymbol{\Sigma}^{-1}) = \frac{1}{M}(\mathbb{Y}^\dagger)^T \mathbf{P} \partial_{\theta_r} (\widehat{\mathbf{A}}^{-1}) \mathbf{P}^H \mathbb{Y}^\dagger, \quad (4.18)$$

$$\partial_{\theta_r} \left(\widehat{\mathbf{A}}^{-1}\right)_{rj,pi} = \frac{1}{M} \partial_{\theta_r} (\widehat{\mathbf{a}}^{-1})_{ji}(\mathbf{k}_r) \delta_{rp}, \quad (4.19)$$

$$\partial_{\theta_r} \left(\widehat{\mathbf{a}}^{-1}\right) = \mathbf{K} \partial_{\theta_r} (\mathbf{S}^{-1}) \mathbf{K}^T = \mathbf{K} \begin{bmatrix} \partial_{\theta_r} (1/S^\nabla) & 0 \\ 0 & \partial_{\theta_r} (1/S^{\nabla\perp}) \end{bmatrix} \mathbf{K}^T \quad (4.20)$$

The normalizing constant can be differentiate with Jacobi formula:

$$\partial_{\theta_r} \log |\boldsymbol{\Sigma}| = \text{tr} (\boldsymbol{\Sigma}^{-1} \partial_{\theta_r} \boldsymbol{\Sigma}), \quad (4.21)$$

$$= \text{tr} \left(\left(\frac{1}{M} (\mathbb{Y}^\dagger)^T \mathbf{P} \widehat{\mathbf{A}}^{-1} \mathbf{P}^H \mathbb{Y}^\dagger \right) \left(\frac{1}{M} \mathbb{Y} \mathbf{P} \partial_{\theta_r} (\widehat{\mathbf{A}}) \mathbf{P}^H \mathbb{Y}^T \right) \right), \quad (4.22)$$

$$= \text{tr} \left(\widehat{\mathbf{A}}^{-1} \partial_{\theta_r} (\widehat{\mathbf{A}}) \right), \quad (4.23)$$

$$= \frac{1}{M} \sum_{ir} (\widehat{\mathbf{a}}^{-1} \partial_{\theta_r} \widehat{\mathbf{a}})_{ii}(\mathbf{k}_r), \quad (4.24)$$

$$= \frac{1}{M} \sum_r \text{tr} \left((\mathbf{K} \mathbf{S}^{-1} \mathbf{K}^T) (\mathbf{K} \partial_{\theta_r} (\mathbf{S}) \mathbf{K}^T) \right), \quad (4.25)$$

$$= \frac{1}{M} \sum_r \left(\partial_{\theta_r} \log S^\nabla(k_r) + \partial_{\theta_r} \log S^{\nabla\perp}(k_r) \right), \quad (4.26)$$

which can evaluate analytically from (3.12)-(3.13). We skip these straightforward calculations for concision and readability. We now gather the different terms to obtain the full the log likelihood gradient to be used

in the gradient descent algorithm, to eventually find the optimal covariance parameters $\boldsymbol{\theta}$:

$$\begin{aligned} d\partial_{\theta_r} l = & -\frac{1}{2M} \sum_r \left(\partial_{\theta_r} \log S^\nabla(k_r) + \partial_{\theta_r} \log S^{\nabla^\perp}(k_r) \right) dt \\ & + \left(\mathbf{F}(\mathbf{Q}(t)|\boldsymbol{\theta})^T \frac{1}{M} (\mathbb{Y}^\dagger(t))^T \mathbf{P} \mathbb{K} \partial_{\theta_r} (\mathbb{S}^{-1}) \mathbb{K}^T \mathbf{P}^H \mathbb{Y}^\dagger(t) \right) \left(d\mathbf{Q}(t) - \frac{1}{2} \mathbf{F}(\mathbf{Q}(t)|\boldsymbol{\theta}) dt \right), \end{aligned} \quad (4.27)$$

where \mathbb{K} and \mathbb{S}^{-1} are the block-diagonal versions of \mathbf{K} and \mathbf{S}^{-1} :

$$\mathbb{S}_{rj,pi}^{-1} = (1/S_{jj}(\mathbf{k}_r)) \delta_{ji} \delta_{rp} \quad \text{and} \quad \mathbb{K}_{rj,pi} = K_{ji}(\mathbf{k}_r) \delta_{rp}. \quad (4.28)$$

Since all matrices except \mathbf{P} are block diagonal, their evaluations have a complexity $O(M)$. Only \mathbf{P} necessities a complexity $O(M \log(M))$ (Fast Fourier Transform algorithm). Evaluating one gradient step requires the time integration of the above formula over N time steps. The computational cost of one gradient is hence $O(NM \log(M))$ only.

5 Conclusion

We have proposed a new approach to estimate statistics of a hidden subgrid velocity field from a sequence of tracer images. The first contribution of the paper is to convert LU-SALT into (4.1) to then apply standard MLE techniques. The second important contribution is an efficient method to solve the MLE optimization problem through a fast log-likelihood gradient evaluation algorithm (4.27). We rely on a parametric model and Fourier-based representations of that velocity to tackle the curse of dimensionality of the problem. As statistics tools, we choose quadratic co-variation and maximum likelihood estimation for their reliable theoretical grounds. Notably, if the subgrid velocity component is solenoidal, its spectrum matrix has rank 1 locally and cannot be inverted. As such, we cannot apply our method. We have to consider both divergent and rotational terms to obtain full rank spectrum matrix, evaluate the joint tracers series distribution, and perform MLE.

Measurement errors are neglected here. True measurement errors are generally weak for satellite images of the oceanic tracers (e.g., SST) even though the image resolution is always limited. Fake measurement errors are sometimes considered to mimic the effect of observed aliased geophysical signals (e.g., nugget in altimetry optimal interpolation (i.e. kriging)). This aliasing is often filtered out by regularization. We do not consider it explicitly here, but this preprocessing could be pursued before applying our method. If we work on L4 satellite products, this would be the case. Nevertheless, by forgetting resolution issues we may increase model errors. These errors may be large and may lead to statistics robustness issues (e.g., when estimating quadratic variations, correlation times, and other parameters through MLE). Indeed, even though erroneous advection – and thus structure mislocation – is well modeled by SALT-LU, we may debate about the dynamics assumed for the partially-resolved geophysical observables. Errors in the dynamics itself may not be negligible. To alleviate this issue together with regularizing the statistics, we could probably add an additive noise in the modeled dynamics on top of the multiplicative noise. We could try to jointly learn this additional stochastic forcing terms following Piterbarg and Rozovskii (1996). Alternatively, the additive white noise forcing variance could be inferred from the tracer quadratic variation (2.2). This statistic

would contain the information of both the forcing variance and the subgrid velocity variance. The number of statistics being large (one by grid point), the least square problem can be solved easily. Considering additive noise complexify the evaluation of the inverse covariance matrix Σ but the Woodbury identity can probably yields a tractable solution.

In future work, we could combine our approach with stochastic optical flow algorithms (Corpetti and Mémin, 2012; Cai et al, 2018). Furthermore, we will adapt our algorithm to deal with non-periodic boundary conditions. Then, we shall apply this methodology to synthetic and real satellite images of tracers like SSS or SST. Our method may also be generalized to treat not only 0-form like SSS or SST, but also more complex differential forms through stochastic geometric fluid dynamics (Holm, 2015; Zhen et al, 2023). As such, we will be able to directly treat surface currents observations from high-frequency radars or wind observations from Doppler radars. More generally, we hope that our work will enable new calibration methods for SALT and LU, but also new observation capabilities from current and future satellite observations, and better physical understanding of tracer budgets from them (e.g. Michel et al, 2007).

Acknowledgment

We thank Bertrand Chapron and the reviewers for helpful discussions, comments and inspiring references.

A Variance parameters for streamfunction, potential and velocity covariances

From properties of the modified Bessel function of second kind, we get the following formula:

$$\partial_r g_\nu(r) = -r g_{\nu-1}(r), \quad (\text{A.1})$$

$$\partial_r^2 g_\nu(r) = -g_{\nu-1}(r) + r^2 g_{\nu-2}(r), \quad (\text{A.2})$$

$$\nabla (g_\nu(\|\mathbf{x}\|)) = -\mathbf{x} g_{\nu-1}(\|\mathbf{x}\|), \quad (\text{A.3})$$

$$H(g_\nu(\|\mathbf{x}\|)) = -g_{\nu-1}(\|\mathbf{x}\|)\mathbb{I}_d + g_{\nu-2}(\|\mathbf{x}\|)\mathbf{x}\mathbf{x}^T, \quad (\text{A.4})$$

$$g_\nu(r) \xrightarrow{r \rightarrow 0} \frac{\Gamma(\nu)}{2^{1-\nu}}, \quad (\text{A.5})$$

with Γ is the Gamma function. From the last equation, we obtain the normalization factor of the streamfunction and potential covariances:

$$D = \frac{2^{\frac{1-\beta}{2}}}{\Gamma(\frac{\beta+1}{2})} \gamma_{\psi_\sigma}(0) \quad (\text{A.6})$$

From the other formula above, we can evaluate the divergence-free velocity covariance:

$$\mathbf{a}^{\nabla^\perp}(\mathbf{x}) = \frac{1}{dt} \mathbb{E} \left\{ \left(\boldsymbol{\sigma}^{\nabla^\perp} dB_t \right) (\mathbf{y} + \mathbf{x}) \left(\boldsymbol{\sigma}^{\nabla^\perp} dB_t \right)^T (\mathbf{y}) \right\}, \quad (\text{A.7})$$

$$= \mathbf{J} \mathbf{H}_{\gamma_{\psi_\sigma}} \mathbf{J}^T, \quad (\text{A.8})$$

$$= D^{\nabla^\perp} (2\pi\kappa_m)^2 \left(g_{\frac{\beta^{\nabla^\perp}-1}{2}} \mathbb{I}_d - g_{\frac{\beta^{\nabla^\perp}-3}{2}} (\mathbf{J}\mathbf{x})(\mathbf{J}\mathbf{x})^T \right). \quad (\text{A.9})$$

Finally, the variance tensor is $\mathbf{a}^{\nabla^\perp}(0) = a_0^{\nabla^\perp} \mathbb{I}_d$ with

$$a_0^{\nabla^\perp} = \frac{(2\pi\kappa_m)^2}{\beta^{\nabla^\perp} - 1} \gamma_{\psi_\sigma}(0) \text{ and } D^{\nabla^\perp} = \frac{2^{\frac{1-\beta^{\nabla^\perp}}{2}}}{\Gamma\left(\frac{\beta^{\nabla^\perp}+1}{2}\right)} \frac{\beta^{\nabla^\perp} - 1}{(2\pi\kappa_m)^2} a_0^{\nabla^\perp}, \quad (\text{A.10})$$

About the amplitude spectrum $S_{\psi_\sigma}(0)$, we know from Williams and Rasmussen (2006); Lim and Teo (2009) that:

$$S_{\psi_\sigma}(0) = \frac{4\pi\Gamma\left(\frac{\beta^{\nabla^\perp}+3}{2}\right)}{2\pi\kappa_m\Gamma\left(\frac{\beta^{\nabla^\perp}+1}{2}\right)} \gamma_{\psi_\sigma}(0) = \frac{2\pi(\beta^{\nabla^\perp}+1)}{(2\pi\kappa_m)^2} \gamma_{\psi_\sigma}(0) = \frac{2\pi((\beta^{\nabla^\perp})^2-1)}{(2\pi\kappa_m)^4} a_0^{\nabla^\perp}. \quad (\text{A.11})$$

The formula for the potential component are similar, replacing ∇^\perp by ∇ , \mathbf{J} by \mathbb{I}_d , and ψ_σ by ϕ_σ .

Bibliography

- Altmeyer R, Reiß M (2021) Nonparametric estimation for linear spdes from local measurements. *The Annals of Applied Probability* 31(1):1–38
- Altmeyer R, Bretschneider T, Janák J, Reiß M (2022) Parameter estimation in an spde model for cell repolarization. *SIAM/ASA Journal on Uncertainty Quantification* 10(1):179–199
- Altmeyer R, Cialenco I, Pasemann G (2023) Parameter estimation for semilinear spdes from local measurements. *Bernoulli* 29(3):2035–2061
- Azencott R, Beri A, Timofeyev I (2010) Adaptive sub-sampling for parametric estimation of gaussian diffusions. *Journal of Statistical Physics* 139(6):1066–1089
- Azencott R, Beri A, Jain A, Timofeyev I (2013) Sub-sampling and parametric estimation for multiscale dynamics. *Communications in Mathematical Sciences* 11(4):939–970
- Cai S, Mémin E, Dérian P, Xu C (2018) Motion estimation under location uncertainty for turbulent fluid flows. *Experiments in Fluids* 59(1):8
- Corpetti T, Mémin E (2012) Stochastic models for local optical flow estimation. In: *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVN 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*, Springer, pp 701–712
- Corpetti T, Heitz D, Arroyo G, Mémin E, Santa-Cruz A (2006) Fluid experimental flow estimation based on an optical-flow scheme. *Experiments in fluids* 40(1):80–97
- Cotter C, Pavliotis G (2009) Estimating eddy diffusivities from noisy lagrangian observations. *Communications in Mathematical Sciences* 7(4):805–838
- Da Prato G, Zabczyk J (1992) *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications, Cambridge University Press
- Dérian P, Héas P, Herzet C, Mémin E (2013) Wavelets and optical flow motion estimation. *Numerical Mathematics: Theory, Methods and Applications* 6(1):116–137
- Gaudlitz S, Reiß M (2023) Estimation for the reaction term in semi-linear spdes under small diffusivity. *Bernoulli* 29(4):3033–3058
- Ghalenoei E, Sharifi MA, Hasanlou M (2015) Estimating and fusing optical flow, geostrophic currents and sea surface wind in the waters around kish island. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40:221–226
- Holm D (2015) Variational principles for stochastic fluid dynamics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 471(2176)
- Janák J, Reiß M (2023) Parameter estimation for the stochastic heat equation with multiplicative noise from local measurements. arXiv preprint arXiv:230300074
- Kraichnan R (1994) Anomalous scaling of a randomly advected passive scalar. *Physical review letters* 72(7):1016
- 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):105,202
- Liu W, Röckner M (2015) *Stochastic partial differential equations: an introduction*. Springer
- Mémin E (2014) Fluid flow dynamics under location uncertainty. *Geophysical & Astrophysical Fluid Dynamics* 108(2):119–146, DOI 10.1080/03091929.2013.836190

- Michel S, Chapron B, Tournadre J, Reul N (2007) Sea surface salinity variability from a simplified mixed layer model of the global ocean. *Ocean Science Discussions* 4(1):41–106
- Oksendal B (1998) *Stochastic differential equations*. Springer-Verlag
- Papavasiliou A, Pavliotis G, Stuart A (2009) Maximum likelihood drift estimation for multiscale diffusions. *Stochastic Processes and their Applications* 119(10):3173–3210
- Pasemann G, Stannat W (2020) Drift estimation for stochastic reaction-diffusion systems. *Electronic Journal of Statistics* 14:547–579
- Pavliotis GA, Stuart A (2007) Parameter estimation for multiscale diffusions. *Journal of Statistical Physics* 127:741–781
- Piterbarg L, Ostrovskii A (1997) *Advection and diffusion in random media: implications for sea surface temperature anomalies*. Kluwer Academic
- Piterbarg L, Rozovskii B (1996) Maximum likelihood estimators in the equations of physical oceanography. In: *Stochastic modelling in physical oceanography*, Springer, pp 397–421
- Rao P (1999) *Statistical inference for diffusion type processes*. Arnold
- Reich S (2023) Frequentist perspective on robust parameter estimation using the ensemble kalman filter. *Stochastic Transport in Upper Ocean Dynamics* p 237
- 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-geostrophy and efficient ensemble spreading. *Geophysical & Astrophysical Fluid Dynamics* 111(3):177–208
- Resseguier V, Li L, Jouan G, Dérian P, Mémin E, Bertrand C (2020a) New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics. *Archives of Computational Methods in Engineering* pp 1–82
- Resseguier V, Pan W, Fox-Kemper B (2020b) Data-driven versus self-similar parameterizations for stochastic advection by lie transport and location uncertainty. *Nonlinear Processes in Geophysics* 27(2):209–234
- Resseguier V, Picard AM, Memin E, Chapron B (2021) Quantifying truncation-related uncertainties in unsteady fluid dynamics reduced order models. *SIAM/ASA Journal on Uncertainty Quantification* 9(3):1152–1183
- Sørensen H (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review* 72(3):337–354
- Sun H, Song Q, Shao R, Schlicke T (2016) Estimation of sea surface currents based on ocean colour remote-sensing image analysis. *International journal of remote sensing* 37(21):5105–5121
- van Waaij J, van Zanten H (2016) Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. *Electronic Journal of Statistics* 10(1):628–645
- Williams CK, Rasmussen CE (2006) *Gaussian processes for machine learning*. the MIT Press 2(3):4
- Zhen Y, Resseguier V, Chapron B (2023) Physically constrained covariance inflation from location uncertainty. *EGUsphere* 2023:1