



**HAL**  
open science

## Constrained random diffeomorphisms for data assimilation

Valentin Resseguier, Yicun Zhen, Bertrand Chapron

► **To cite this version:**

Valentin Resseguier, Yicun Zhen, Bertrand Chapron. Constrained random diffeomorphisms for data assimilation. *Stochastic Transport in Upper Ocean Dynamics 2022*, 2023. hal-03994686

**HAL Id: hal-03994686**

**<https://hal.science/hal-03994686v1>**

Submitted on 17 Feb 2023

**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.

# Constrained random diffeomorphisms for data assimilation

Valentin Resseguier<sup>1,2</sup>, Yicun Zhen<sup>3</sup>, Bertrand Chapron<sup>4</sup>

<sup>1</sup> INRAE, OPAALE, Rennes, France,  
valentin.resseguier@scalian.com,

<sup>2</sup> LAB, SCALIAN DS, Rennes, France

<sup>3</sup> Department of Oceanography, Hohai University, Nanjing, Jiangsu, China

<sup>4</sup> Laboratoire d’Océanographie Physique et Spatiale (LOPS), Ifremer, France

## 1 Introduction

For ensemble-based data assimilation purposes, there is definite need for relevant ensemble sampling tools. Indeed, the quality and spreading of these ensembles have deep implications in the quality of the data assimilation (Dufée et al, 2022), and – until recently – those so-called covariance inflation tools have mostly relied on unsuitable linear Gaussian frameworks (Tandeo et al, 2020; Resseguier et al, 2020a). A promising alternative is the generation of ensembles through stochastic remapping of the physical space.

Consider a random mapping  $T$ , acting at every infinitesimal time step, such that  $T_t(x) - x$  is interpreted as a "location perturbation" expressed by

$$T_t(x) = x + a(t, x)\Delta t + e_i(t, x)\Delta\eta_i(t), \quad (1.1)$$

where  $a(t, x), e_i(t, x) \in \mathbb{R}^n$ ,  $\Delta\eta_i(t) \sim \mathcal{N}(0, \Delta t)$  is a random number. In Eq.(1.1),  $a(t, x)\Delta t$  represent a deterministic location shift, and  $e_i\Delta\eta_i$  random ones. At every time step, the random mapping  $T$  shall then induce a perturbation to any tensor field  $\theta(t)$  (Zhen et al, 2022). For instance, one can perturb a differential form  $\theta(t)$  applying  $\theta(t) \rightarrow T_t^*\theta(t)$  with  $T_t^*$  the associated pull-back operator.

A rigorous mathematical definition and calculation of  $T_t$  and  $T_t^*$  can be obtained in terms of stochastic flows of diffeomorphisms and its Lie derivatives (e.g., Bethencourt De Leon, 2021). Yet, to rapidly assess  $T_t^*\theta$ , a Taylor expansion and usage of Ito’s lemma can be used. Given coordinates  $(x^1, \dots, x^n)$ , when  $\theta$  is a differential  $k$ -form, it can be written as

$$\theta = \sum_{i_1 < \dots < i_k} f^{i_1, \dots, i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k} \quad (1.2)$$

with  $f$  a smooth function. Then

$$T_t^*\theta = \sum_{i_1 < \dots < i_k} f^{i_1, \dots, i_k}(T_t(x))T_t^*(dx^{i_1} \wedge \dots \wedge dx^{i_k}) \quad (1.3)$$

leading to a compact expression

$$T_t^* \theta = \theta + \mathcal{M}(t, \theta) \Delta t + \mathcal{N}_i(t, \theta) \Delta \eta_i(t), \quad (1.4)$$

with some differential  $k$ -forms  $\mathcal{M}(t, \theta)$  and  $\mathcal{N}_i(t, \theta)$ , (see Zhen et al, 2022, Appendix B for full proof and definitions of  $\mathcal{M}$  and  $\mathcal{N}$ ).

In this paper, we present and discuss the potential of this random mapping scheme to possibly prescribe  $\theta$ , and the parameters  $a$  and  $e_i$  to ensure that certain quantities, i.e. mass, vorticity, helicity, energy, are conserved.

## 2 Induced Stochastic PDE

From the expression of  $T_t^* \theta$ , a SPDE is derived from an original PDE, when  $\theta$  is a differential form. Suppose  $S$  is the full state variable of the dynamical system:

$$\frac{\partial S}{\partial t} = g(S). \quad (2.1)$$

Let  $f$  be a component or a collection of components of  $S$ . We then associate  $f$  to a differential form  $\theta$ , i.e. there is an invertible map  $\mathcal{F}$  that maps the space of  $f$  to the space of  $\theta$ , such that  $\mathcal{F}(f) = \theta$ . Typically, if  $f$  is a tracer, it is often associated to the 0-form  $\theta = f$ . If  $f$  is the density  $\rho$ , we might associate the  $n$ -form  $\theta = \rho dx^{i_1} \wedge \dots \wedge dx^{i_n}$ . Consider the propagation equation for  $f$

$$df = g^f(S) dt. \quad (2.2)$$

It implies a propagation equation for  $\theta$ :

$$d\theta = g^\theta(S) dt. \quad (2.3)$$

The proposed discrete-time perturbation at each time step consists of the following two steps:

$$\begin{cases} \tilde{\theta}(t + \Delta t) = \theta(t) + g^\theta(S(t)) \Delta t & (2.4) \\ \theta(t + \Delta t) = T_t^* \tilde{\theta}(t + \Delta t) & (2.5) \end{cases}$$

with  $T_t^* \tilde{\theta}(t + \Delta t) = \tilde{\theta}(t + \Delta t) + \mathcal{M}(t, \tilde{\theta}(t + \Delta t)) \Delta t + \mathcal{N}_i(t, \tilde{\theta}(t + \Delta t)) \Delta \eta_i(t) + o(\Delta t)$  for the associated differential forms  $\mathcal{M}(t, \tilde{\theta})$  and  $\mathcal{N}_i(t, \tilde{\theta})$ .

The deterministic PDE (2.4) and  $\|\tilde{\theta}(t + \Delta t) - \theta(t)\|$  scales in  $O(\Delta t)$ . There is no noise term to induce a scaling in  $O(\sqrt{\Delta t})$ . Therefore, it can be assumed that there exists  $C > 0$  so that  $\|\mathcal{M}(t, \tilde{\theta}(t + \Delta t)) - \mathcal{M}(t, \theta(t))\| < C \Delta t$  and  $\|\mathcal{N}_i(t, \tilde{\theta}(t + \Delta t)) - \mathcal{N}_i(t, \theta(t))\| < C \Delta t$ , for  $\Delta t$  small enough. Accordingly,

$$\begin{aligned} T_t^* \tilde{\theta}(t + \Delta t) &= \tilde{\theta}(t + \Delta t) + \left( \mathcal{M}(t, \theta(t)) + \mathcal{O}(\Delta t) \right) \Delta t \\ &\quad + \left( \mathcal{N}_i(t, \theta(t)) + \mathcal{O}(\Delta t) \right) \Delta \eta_i(t) + o(\Delta t) \\ &= \tilde{\theta}(t + \Delta t) + \mathcal{M}(t, \theta(t)) \Delta t + \mathcal{N}_i(t, \theta(t)) \Delta \eta_i(t) + o(\Delta t) \end{aligned} \quad (2.6)$$

Therefore,

$$\theta(t + \Delta t) = \theta(t) + g^\theta(S(t))\Delta t + \mathcal{M}(t, \theta(t))\Delta t + \mathcal{N}_i(t, \theta(t))\Delta\eta_i + o(\Delta t). \quad (2.7)$$

It suggests the following stochastic propagation equation for  $\theta$ :

$$d\theta = g^\theta(S)dt + \mathcal{M}(t, \theta)dt + \mathcal{N}_i(t, \theta)d\eta_i. \quad (2.8)$$

Since there is a 1-1 correspondence between  $\theta$  and  $f$ , Eq.(2.3) also suggests a stochastic propagation equation for  $f$ , which can be written as

$$df = g^f(S)dt + \mathcal{M}^f(f)dt + \mathcal{N}_i^f(f)d\eta_i. \quad (2.9)$$

We denote the additional terms in Eq.(2.9) by

$$d_s f := \mathcal{M}^f(f)dt + \mathcal{N}_i^f(f)d\eta_i. \quad (2.10)$$

Then Eq.(2.9) can be written as:

$$df = g^f(S)dt + d_s f. \quad (2.11)$$

### 3 Comparison with other perturbation schemes

Obtained above,  $d_s f$  is completely determined by  $T_t^* \theta$ , but is not directly related to the original dynamics Eq.(2.2). Once the expression of  $T$  in Eq.(1.1) and the choice of  $\theta$  are determined, the perturbation term  $d_s f$  is prescribed. However, the choice of  $\theta$  is up to the user, and may then be related to the original dynamics.

In the following, we thus demonstrate that both the stochastic advection by Lie transport (SALT) equation (Holm, 2015) and the location uncertainty (LU) equation (Mémín, 2014; Resseguier et al, 2017, 2020b) can be properly recovered using the proposed perturbation scheme.

#### 3.1 Comparison with the LU equations

The Reynolds transport theorem is central to the LU setting. The Reynolds transport theorem expresses an integral conservation equation for the transport of any conserved quantity within a fluid, connected to its corresponding differential equation. A link between the proposed perturbation approach and the LU formulation can be anticipated to be related to differential  $n$ -forms. But first, we consider a key ingredient of LU: the stochastic material derivative of function (differential 0-forms).

**0-forms in the LU framework** Dropping the forcing terms, LU equation for compressible and incompressible flow writes (Resseguier et al, 2017).

$$\partial_t f + \mathbf{w}^* \cdot \nabla f = \nabla \cdot (\frac{1}{2} \mathbf{a} \nabla f) - \boldsymbol{\sigma} \dot{\mathbf{B}} \cdot \nabla f \quad (3.1)$$

$$\mathbf{w}^* = \mathbf{w} - \frac{1}{2} (\nabla \cdot \mathbf{a})^\top + \boldsymbol{\sigma} (\nabla \cdot \boldsymbol{\sigma})^\top \quad (3.2)$$

where  $\mathbf{a} = \boldsymbol{\sigma}_{\bullet k} \boldsymbol{\sigma}_{\bullet k}^T$  and  $f$  can be any quantity that is assumed to be transported by the flow, i.e.  $Df/Dt = 0$  where  $D/Dt$  is the Itô material derivative. For instance,  $f$  could be the velocity (dropping forces in the SPDE), the temperature, or the buoyancy.

Separating the terms of the SPDE related to the deterministic dynamics from the term associated to the stochastic scheme, it comes

$$\mathbf{d}^{\text{LU}} f = g^f(S)dt + \mathbf{d}_s^{\text{LU}} f, \quad (3.3)$$

where

$$g^f(S) = -\mathbf{w} \cdot \nabla f \quad (3.4)$$

$$\mathbf{d}_s^{\text{LU}} f = -(\mathbf{w}^* - \mathbf{w}) \cdot \nabla f dt - \boldsymbol{\sigma} d\mathbf{B} \cdot \nabla f + \nabla \cdot (\frac{1}{2} \mathbf{a} \nabla f) dt \quad (3.5)$$

Besides, from our proposed scheme applied to a 0-form  $\theta = f$ , we obtain:

$$\mathbf{d}_s f = \left( a^p \partial_{x^p} f + \frac{1}{2} e_i^p e_i^q \partial_{x^p} \partial_{x^q} f \right) dt + e_i^p \partial_{x^p} f d\eta_i \quad (3.6)$$

To physically interpret this equation, we rewrite:

$$\frac{\mathbf{d}_s f}{dt} = -V^p \partial_{x^p} f + \partial_{x^p} \left( \left( \frac{1}{2} e_i^p e_i^q \right) \partial_{x^q} f \right) \quad (3.7)$$

where

$$V^p = -a^p + \frac{1}{2} \partial_{x^q} (e_i^p e_i^q) - e_i^p \frac{d\eta_i}{dt} \quad (3.8)$$

Terms of advection and diffusion are recognized. The matrix  $\frac{1}{2} e_i e_i^T$  is symmetric non-negative and represents a diffusion matrix. The  $p$ -th component of the advecting velocity  $V^p$  is composed of the drift  $-a^p$ , a correction  $\frac{1}{2} \partial_{x^q} (e_i^p e_i^q)$ , and a stochastic advecting velocity  $-e_i^p \frac{d\eta_i}{dt}$ .

Direct calculation yields that Eq.(3.5) coincides with Eq.(3.7) when  $\mathbf{a} = \boldsymbol{\sigma}_{\bullet k} \boldsymbol{\sigma}_{\bullet k}^T = e_i e_i^T$  and  $\boldsymbol{\sigma} \mathbf{B} = -e_i d\eta_i$  and

$$T_t(x) = x + e_i^q \partial_{x_q} e_i \Delta t + e_i \Delta \eta_i = x - \mathbf{w}_S^c \Delta t + (-\mathbf{w}_S^c \Delta t - \boldsymbol{\sigma} \Delta \mathbf{B}), \quad (3.9)$$

where

$$\mathbf{w}_S^c = -\frac{1}{2} (\boldsymbol{\sigma}_{\bullet k} \cdot \nabla) \boldsymbol{\sigma}_{\bullet k} = -\frac{1}{2} (\nabla \cdot \mathbf{a})^\top + \frac{1}{2} \boldsymbol{\sigma} (\nabla \cdot \boldsymbol{\sigma})^\top. \quad (3.10)$$

The LU equation can thus be derived by choosing  $\theta = f$  and  $T_t$  by Eq.(3.9). Note, the term  $(-\mathbf{w}_S^c \Delta t - \boldsymbol{\sigma} \Delta \mathbf{B}) = (\frac{1}{2} e_i^q \partial_{x_q} e_i \Delta t + e_i \Delta \eta_i)$  is the Itô noise plus its Itô-to-Stratonovich correction. Hence, it corresponds to the Stratonovich noise  $e_i \circ d\eta_i$  of the flow associated to  $T_t$ . The additional drift  $-\mathbf{w}_S^c \Delta t$  is different in nature. It is related to the advection correction  $\mathbf{w}_S^c \cdot \nabla f$  in the LU setting. Indeed, in the LU framework, the Itô drift,  $\mathbf{w}$ , is seen as the resolved large-scale velocity. That is why, in this framework, the deterministic dynamics (3.4) involves the Itô drift,  $\mathbf{w}$ . This is also the reason why, under the LU derivation,

the advected velocity is assumed to be given by the Itô drift,  $\mathbf{w}$ . It differs from the Stratonovich drift  $\mathbf{w}_S = \mathbf{w} + \mathbf{w}_S^c$ , used as advected velocity in SALT approach or in Mikulevicius and Rozovskii (2004) (where the Stratonovich drift is denoted  $u$ ). Interested readers are referred to (Resseguier et al, 2020b, Appendix A) for a discussion on these assumptions and for the complete table of SALT-LU notations correspondences. Note however that in all these approaches, the advecting velocity is always the Stratonovich drift. This can be seen e.g., in the Stratonovich form of LU equations, derived in (Resseguier, 2017, Appendix 10.1) and (Resseguier et al, 2020a, 6.1.3):

$$\partial_t f + \mathbf{w}_S \cdot \nabla f = -(\boldsymbol{\sigma} \circ \dot{\mathbf{B}}) \cdot \nabla f, \quad (3.11)$$

where  $\boldsymbol{\sigma} \circ \dot{\mathbf{B}}$  is the Stratonovich noise of the SPDE. Since the advecting velocity  $\mathbf{w}_S$  and the resolved velocity  $\mathbf{w}$  differ by a drift  $\mathbf{w}_S^c$ , the term  $\mathbf{w}_S^c \cdot \nabla f$  is interpreted as an advection correction, being part of the stochastic scheme (3.5). Accordingly, the remapping  $T_t$  involves an additional drift  $-\mathbf{w}_S^c \Delta t$ .

To also understand (3.9), the inverse flow can be considered:

$$T_t^{-1}(x) = x - e_i \Delta \eta_i = x + \boldsymbol{\sigma} \Delta \mathbf{B}. \quad (3.12)$$

If  $T_t$  represents a necessary perturbation to match, at each time step, a true solution,  $T_t^{-1}$  measures the difference, at each time step, between this true solution and a model forecast. Therefore, the LU equation can be derived using the proposed perturbation scheme, choosing  $\theta = f$  and assuming that a true solution differs from a model forecast by a displacement prescribed by Eq.(3.12).

**n-forms in the LU framework** The LU physical justification relies on a stochastic interpretation of fundamental conservation laws, typically conservation of extensive properties (i.e. integrals of functions over a spatial volume) like momentum, mass, matter and energy (Resseguier et al, 2017). These extensive properties can be expressed by integrals of differential  $n$ -forms. For instance, the mass and the momentum are integrals of the differential  $n$ -forms  $\rho dx^1 \wedge \dots \wedge dx^n$  and  $\rho \mathbf{w} dx^1 \wedge \dots \wedge dx^n$ , respectively. In the LU framework, a stochastic version of the Reynolds transport theorem (Resseguier et al, 2017, Eq. (28)) is used to deal with these differential  $n$ -forms  $\theta = f dx^1 \wedge \dots \wedge dx^n$ . Assuming an integral conservation  $\frac{d}{dt} \int_{V(t)} f = 0$  on a spatial domain  $V(t)$  transported by the flow, it leads to the following SPDE:

$$\frac{Df}{Dt} + \nabla \cdot (\mathbf{w}^* + \boldsymbol{\sigma} \dot{\mathbf{B}}) f = \frac{d}{dt} \left\langle \int_0^t D_t f, \int_0^t \nabla \cdot \boldsymbol{\sigma} \dot{\mathbf{B}} \right\rangle = (\nabla \cdot \boldsymbol{\sigma}_{\bullet i})(\nabla \cdot \boldsymbol{\sigma}_{\bullet i})^T f \quad (3.13)$$

where  $D/Dt$  denotes the Itô material derivative. Forcing terms are dropped for the sake of readability. This SPDE can be rewritten using the expression of that material derivative (Eq. (9) and (10) of Resseguier et al (2017)):

$$\partial_t f + \nabla \cdot (\mathbf{w}_S f) = \frac{1}{2} \nabla \cdot (\mathbf{a} \nabla f) + \frac{1}{2} \nabla \cdot (\boldsymbol{\sigma}_{\bullet i} (\nabla \cdot \boldsymbol{\sigma}_{\bullet i})^T f) - \nabla \cdot (\boldsymbol{\sigma} \dot{\mathbf{B}} f) \quad (3.14)$$

$$(3.15)$$

The original deterministic equation and stochastic perturbation correspond to

$$g^f(S) = -\nabla \cdot (\mathbf{w}f) \quad (3.16)$$

$$\mathbf{d}_s^{\text{LU}} f = (-\nabla \cdot (\mathbf{w}_S^c f) + \frac{1}{2} \nabla \cdot (\mathbf{a} \nabla f) + \frac{1}{2} \nabla \cdot (\boldsymbol{\sigma}_{\bullet i} (\nabla \cdot \boldsymbol{\sigma}_{\bullet i})^T f)) \mathbf{d}t - \nabla \cdot (\boldsymbol{\sigma} \mathbf{d}\mathbf{B} f) \quad (3.17)$$

$$= -\nabla \cdot ((-\frac{1}{2} \nabla \cdot \mathbf{a})^T \mathbf{d}t + \boldsymbol{\sigma} \mathbf{d}\mathbf{B}) f + \nabla \cdot (\frac{1}{2} \mathbf{a} \nabla f) \mathbf{d}t \quad (3.18)$$

We can now compare these LU equations to our new stochastic scheme applied to  $n$ -form  $\theta = f dx^1 \wedge \cdots \wedge dx^n$ : This implies that

$$\begin{aligned} \mathbf{d}_s f = & \left( (\partial_{x^p} a^p + \frac{1}{2} J_i) f + (a^p + e_i^p \partial_{x^q} e_i^q) \partial_{x^p} f + \frac{1}{2} e_i^p e_i^q \partial_{x^p} \partial_{x^q} f \right) \mathbf{d}t \\ & + (\partial_{x^p} e_i^p f + e_i^p \partial_{x^p} f) \mathbf{d}\eta_i, \end{aligned} \quad (3.19)$$

where  $J_i = \partial_{x^p} e_i^p \partial_{x^q} e_i^q - \partial_{x^p} e_i^q \partial_{x^q} e_i^p$ . Rewritten, it leads to:

$$\frac{\mathbf{d}_s f}{\mathbf{d}t} = -\partial_{x^p} (\tilde{V}^p f) + \partial_{x^p} ((\frac{1}{2} e_i^p e_i^q) \partial_{x^q} f) \quad (3.20)$$

where

$$\tilde{V}^p = V^p - (e_i^p \partial_{x^q} e_i^q) = -a^p + \frac{1}{2} (\partial_{x^q} e_i^p e_i^q - e_i^p \partial_{x^q} e_i^q) - e_i^p \frac{\mathbf{d}\eta_i}{\mathbf{d}t} \quad (3.21)$$

Again a advection-diffusion equation is recognized, but of different nature. Indeed, as expected for a  $n$ -form, the PDE is similar to a density conservation equation. Moreover, the advecting drift is slightly different to take into account the cross-correlations between  $f(T_t(x))$  and  $T_t^*(dx^1 \wedge \cdots \wedge dx^n)$ .

Identifying  $\mathbf{a} = \boldsymbol{\sigma}_{\bullet k} \boldsymbol{\sigma}_{\bullet k}^T = e_i e_i^T$  and  $\boldsymbol{\sigma} \dot{\mathbf{B}} = -e_i \mathbf{d}\eta_i$ ,

$$\tilde{V} = -a^p + \frac{1}{2} (\partial_{x^q} e_i^p e_i^q - e_i^p \partial_{x^q} e_i^q) - e_i^p \frac{\mathbf{d}\eta_i}{\mathbf{d}t} = -(\frac{1}{2} \nabla \cdot \mathbf{a})^T + \boldsymbol{\sigma} \dot{\mathbf{B}} \quad (3.22)$$

i.e.

$$a^p = \frac{1}{2} (\partial_{x^q} e_i^p e_i^q - e_i^p \partial_{x^q} e_i^q) + \frac{1}{2} \partial_{x^q} (e_i^p e_i^q) = e_i^q \partial_{x^q} e_i^p \quad (3.23)$$

A remapping is thus obtained to write

$$T_t(x) = x + e_i^q \partial_{x^q} e_i \Delta t + e_i \Delta \eta_i = x - \mathbf{w}_S^c \Delta t + (-\mathbf{w}_S^c \Delta t - \boldsymbol{\sigma} \Delta \mathbf{B}), \quad (3.24)$$

already derived for differential 0-form in LU framework (Eq. (3.9)). Therefore, the proposed perturbation mapping can also encompass the LU framework for  $n$ -forms, and its capacity – given by the Reynolds transport theorem – to deal with extensive properties.

Moreover, for incompressible flows, LU equation further imposes that

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma} = 0 \\ \nabla \cdot \nabla \cdot \mathbf{a} = 0 \end{cases} \quad (3.25)$$

Translating it into our present notation, it reads as

$$\begin{cases} \partial_{x_p} e_i^p = 0 \text{ for each } i \\ \partial_{x_p} \partial_{x_q} (e_i^p e_i^q) = 0 \end{cases}$$

Following straightforward calculation, Eq.(3.25) is found equivalent to that  $T_t^* \theta = \theta$  for  $\theta = dx^1 \wedge \dots \wedge dx^n$ . Such a result is expected since constraints Eq. (3.25) are obtained from the LU density conservation.

### 3.2 The SALT perturbation scheme

Holm (2015) derived the original SALT equation following a stochastically constrained variational principle  $\delta S = 0$ , for which

$$\begin{cases} S(u, q) = \int \ell(u, q) dt \\ dq + \mathcal{L}_{\mathbf{d}x_t} q = 0. \end{cases} \quad (3.26)$$

where  $\ell(u, q)$  is the Lagrangian of the system,  $\mathcal{L}$  is the Lie derivative, and  $x_t(x)$  is defined by (using our notation)

$$x_t(x) = x_0(x) + \int_0^t u(x, s) ds - \int_0^t e_i(x) \circ \mathbf{d}\eta_i(s), \quad (3.27)$$

in which  $u$  is the velocity vector field. The  $\circ$  means that the integral is defined in the Stratonovich sense, instead of in the Ito sense. Hence,  $\mathbf{d}x_t = u(x, t)dt - e_i \circ \mathbf{d}\eta_i$  refers to an infinitesimal stochastic tangent field on the domain. We can express  $\mathbf{d}x_t = T_t(x) - x + udt$ . Note the difference between Ito's notation and Stratonovich's notation, i.e.  $e_i \circ \mathbf{d}\eta_i \neq e_i d\eta_i$ . The initial expression of  $T_t$  essentially follows Ito's notation. In this subsection, it comes that  $T_t(x) \neq x - e_i \Delta \eta_i$ . Instead, it becomes  $T_t(x) = x + \frac{1}{2} e_i^p \partial_{x_p} e_i \Delta t - e_i \Delta \eta_i$ .

In the second equation of Eq.(3.26),  $q$  is assumed to be a quantity advected by the flow.  $q$  can correspond to any differential form that is not uniquely determined by the velocity (since the SALT equation for the velocity is usually determined by the first equation of Eq.(3.26)). Holm (2015) evaluates the Lie derivative  $\mathcal{L}_{\mathbf{d}x_t} q$  using Cartan's formula:

$$\mathcal{L}_{\mathbf{d}x_t} q = d(i_{\mathbf{d}x_t} q) + i_{\mathbf{d}x_t} dq. \quad (3.28)$$

This Lie derivative  $\mathcal{L}_{\mathbf{d}x_t} q$  corresponds to  $T_t^* q - q + \int^q(S)dt$ , if we assume that the deterministic forecast of  $q$  is simply the advection of  $q$  by  $u$ . More generally,  $\mathcal{L}_{\mathbf{d}x_t - udt} q = T_t^* q - q$ . Therefore, the SALT equation for  $q$  is the same as our perturbation for  $q$ . Note, the Cartan's formula can not be directly applied to calculate the Lie derivative if the expression of  $\mathbf{d}x_t$  is in Ito's notation.

Within the SALT setting, the velocity  $u$  comes from the first equation of Eq.(3.26). For most cases, the velocity  $u$  is associated with the momentum, a differential 1-form  $m = u^j dx^j = u^1 dx^1 + \dots + u^n dx^n$ . When the Lagrangian



includes the kinetic energy, Holm (2015) observed that the stochastic noises contribute a term  $\mathcal{L}_{\mathbf{d}x_t}\theta$ , where  $\theta$  is a differential 1-form related to the momentum 1-form. In particular,  $\theta = \mathbf{m}$  for the ‘‘Stratonovich stochastic Euler-Poincaré flow’’ example, and  $\theta = \mathbf{m} + R^j dx^j$  for the ‘‘Stochastic Euler-Boussinesq equations of a rotating stratified incompressible fluid’’.

Already pointed out, the operator  $\mathcal{L}_{\mathbf{d}x_t}$  is closely related to  $T_t^*$ , and the SALT momentum equation can thus also be derived using our proposed perturbation scheme by properly choosing  $\theta$ , without relying on Lagrangian mechanics.

Another way to appreciate the correspondence to SALT is by looking at the final SPDE. If we choose  $\theta$  to be a differential 1-form to represent the momentum  $f$ , i.e.  $\theta = f^j dx^j$  we obtain (Zhen et al, 2022):

$$\begin{aligned} \mathbf{d}_s f^j = & (a^p \partial_{x^p} f^j + \frac{1}{2} e_i^p e_i^q \partial_{x^p} \partial_{x^q} f^j + \partial_{x^j} a^p f^p + \partial_{x^j} e_i^p e_i^q \partial_{x^q} f^p) \mathbf{d}t \\ & + (e_i^p \partial_{x^p} f^j + \partial_{x^j} e_i^p f^p) \mathbf{d}\eta_i \end{aligned} \quad (3.29)$$

Regrouping the terms for physical interpretation, it writes:

$$\frac{\mathbf{d}_s f^j}{\mathbf{d}t} = -V^p \partial_{x^p} f^j + \partial_{x^p} \left( \left( \frac{1}{2} e_i^p e_i^q \right) \partial_{x^q} f^j \right) + \partial_{x^j} \left( a^p + e_i^p \frac{\mathbf{d}\eta_i}{\mathbf{d}t} \right) f^p + \partial_{x^j} e_i^p e_i^q \partial_{x^q} f^p \quad (3.30)$$

Two last terms of the right-hand side complete the advection-diffusion terms, already appearing in (3.7). The first one,  $\partial_{x^j} \left( -a^p - e_i^p \frac{\mathbf{d}\eta_i}{\mathbf{d}t} \right) f^p$ , is reminiscent to the additional terms appearing in SALT momentum equations (Holm, 2015; Resseguier et al, 2020b). The second term,  $-\partial_{x^j} e_i^p e_i^q \partial_{x^q} f^p$ , comes from cross-correlation in Itô notation.

## 4 Conclusion

As demonstrated, both SALT and LU equations can be recovered using a prescribed definition of a random diffeomorphism  $T_t$  used to perturb the physical space. However, compared with SALT and LU settings, the proposed perturbation scheme does not directly rely on a particular physics. Hence, the random mapping is more flexible and can be applied to any PDE. Interestingly, similarities and differences can then be identified and studied between the proposed use of the random diffeomorphism and the existing stochastic physical SALT and LU settings. For instance, the proposed derivation provides interesting interpretation the operator  $\mathcal{L}_{\mathbf{d}x_t - u \mathbf{d}t}$ , appearing in the SALT equation. This term can indeed represent an infinitesimal forecast error at every forecast time step.

To apply the proposed perturbation scheme to any specific model, the diffeomorphism parameters  $a$  and  $e_i$  must be determined specifically. Hence it is necessary to learn these parameters from existing data, experimental runs, or additional physical considerations. This framework naturally provides new perspectives to generate ensembles through constrained stochastic mappings applied in the physical space.

## **Acknowledgment**

This work is supported by the ERC project 856408-STUOD and SCALIAN DS.

## Bibliography

- Bethencourt De Leon A (2021) On the effect of stochastic lie transport noise on fluid dynamic equations. PhD thesis, Imperial College London
- Dufée B, Mémin E, Crisan D (2022) Stochastic parametrization: an alternative to inflation in ensemble kalman filters. *Quarterly Journal of the Royal Meteorological Society* 148(744):1075–1091
- Holm D (2015) Variational principles for stochastic fluid dynamics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 471(2176)
- Mémin E (2014) Fluid flow dynamics under location uncertainty. *Geophysical & Astrophysical Fluid Dynamics* 108(2):119–146, DOI 10.1080/03091929.2013.836190
- Mikulevicius R, Rozovskii B (2004) Stochastic Navier–Stokes equations for turbulent flows. *SIAM Journal on Mathematical Analysis* 35(5):1250–1310
- Resseguier V (2017) Mixing and fluid dynamics under location uncertainty. PhD thesis, Rennes 1 university
- Resseguier V, Mémin E, Chapron B (2017) Geophysical flows under location uncertainty, part I random transport and general models. *Geophysical & Astrophysical Fluid Dynamics* 111(3):149–176
- 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
- Tandeo P, Ailliot P, Bocquet M, Carrassi A, Carrassi A, Miyoshi T, Pulido M, Pulido M, Zhen Y (2020) A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Monthly Weather Review*
- Zhen Y, Resseguier V, Chapron B (2022) Physically constrained covariance inflation from location uncertainty and optimal transportation. arXiv preprint arXiv:221104207