Information Barriers for Noisy Lagrangian Tracers in Filtering Random Incompressible Flows

Nan Chen, Andrew J Majda and Xin T Tong

Department of Mathematics, and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

E-mail: `chennan@cims.nyu.edu`, `jonjon@cims.nyu.edu`, `tong@cims.nyu.edu`

Abstract. An important practical problem is the recovery of a turbulent velocity field from Lagrangian tracers that move with the fluid flow. Here the filtering skill of L moving Lagrangian tracers in recovering random incompressible flow fields defined through a finite number of random Fourier modes is studied with full mathematical rigor. Despite the inherent nonlinearity in measuring noisy Lagrangian tracers, it is shown below that there are exact closed analytic formulas for the optimal filter for the velocity field involving Riccati equations with random coefficients for the covariance matrix. This mathematical structure allows a detailed asymptotic analysis of filter performance both as time goes to infinity and as the number of noisy Lagrangian tracers, L, increases. In particular, the asymptotic gain of information from L-tracers grows only like $\ln L$ in a precise fashion, i.e., an exponential increase in the number of tracers is needed to reduce the uncertainty by a fixed amount; in other words, there is a practical information barrier. The proofs proceed through a rigorous mean field approximation of the random Ricatti equation. Also, as an intermediate step, geometric ergodicity with respect to the uniform measure on the period domain is proved for any fixed number L of noisy Lagrangian tracers. All of the above claims are confirmed by detailed numerical experiments presented here.

Submitted to: *Nonlinearity*

1. Introduction

Lagrangian tracers are drifters or floaters that follow a parcel of fluid's movement. Serving as moving observations, Lagrangian tracers gather real-time information as they travel, which is of significance especially for data collection in the center of the ocean since Eulerian measurements are available only near the shores [1, 2]. One key application of these Lagrangian data, which is also the main concern of this paper, is the inference of the current state of the velocity field.

Data assimilation with Lagrangian tracers has attracted considerable attention across the years. Various filtering approaches have been applied to Lagrangian data in different modeling contexts. This includes assimilation methods based on Taylor expansion [3], particle filter methods [4] and the ensemble Kalman filter (EnKF) [5, 6, 7, 8, 9]. These approximate filters produce satisfactory numerical results in the recovery of the velocity field. Related numerical experiments also shed light on the dependence of filter performance over different factors, such as tracers' trajectories, number of tracers and initial deployment of the tracers [6, 7].

In spite of the achievements mentioned above, little is known about filters for Lagrangian tracers in theory, as all the analyses were carried out by numerical experiments. The nonlinear dynamics of the Lagrangian tracers makes it difficult to establish a theoretical framework which in turn impedes systematic understanding. One goal of the present work is to introduce random incompressible flow models which, despite the nonlinearity in measuring noisy Lagrangian tracers, have mathematical structure with exact closed analytic formulas for an optimal filter for the velocity field.

The second motivation of this paper lies in understanding the uncertainty reduction with Lagrangian tracers. Without any observations, the equilibrium distribution of the underlying dynamics provides the least biased estimation of the current flow. Unfortunately, the strong turbulent nature brings about a large uncertainty in the equilibrium distribution among many practical issues. Intuitively, the observations of the tracers' trajectories helps to reduce the uncertainty. This uncertainty reduction can be quantified using information theory, as is well documented in the literature [10, 11, 12, 13].

In principle, adding more tracers is expected to gain more information about the underlying flow. However, adding more tracers may not be an effective way to reduce the uncertainty. On one hand, this will inevitably increase the maintenance and computational cost of the filter. On the other hand, as recently revealed in various scenarios for prediction with model error [10, 11, 12], there might exist an information barrier for our filter. An information barrier indicates that a certain level of information is extremely difficult to acquire in practice, therefore adding tracers produces little effective information return. Evidently, identifying and understanding a potential information barrier is of importance for the optimal arrangement of the filter.

In the following, we consider a simplified set-up with a d dimensional incompressible random flow modeled by a finite number of Fourier modes with random amplitudes:

$$\vec{v}(\vec{x}, s) = \sum_{\mathbf{k} \in \mathbf{K}} \hat{v}_{\mathbf{k}}(s) \exp(\mathrm{i}\vec{k} \cdot \vec{x})\vec{r}_{\mathbf{k}}. \tag{1.1}$$

Since the whole velocity field $\vec{v}$ is real-valued, we require that the modes $\hat{v}_{\mathbf{k}}(s)$ with nonzero wavenumbers form conjugate pairs. The $\vec{r}_{\mathbf{k}}$ are vectors orthogonal to the wavenumber $\vec{k}$, due to the incompressibility of the whole velocity, i.e. $\nabla \cdot \vec{v} = 0$. Each $\hat{v}_{\mathbf{k}}(s)$ follows an Ornstein-Uhlenbeck (O.U.) process,

$$\mathrm{d}\hat{v}_{\mathbf{k}}(s) = -d_{\mathbf{k}}\hat{v}_{\mathbf{k}}(s)\mathrm{d}s + f_{\mathbf{k}}(s)\mathrm{d}s + \sigma_{\mathbf{k}}\mathrm{d}W_{\mathbf{k}}^{v}(s), \tag{1.2}$$

where $f_{\mathbf{k}}(s)$ is a prescribed external forcing. The joint dynamics of all these Fourier coefficients serve as the signal process in our filtering problem. This model can be regarded as a stochastic version of many canonical deterministic fluid models, cf. [14] and [15] Chapter 5.

On the other hand, the observation process is given by the trajectories of L Lagrangian tracers involving some noise, i.e. noisy Lagrangian tracers. This observation process is modeled by the following stochastic differential equations (SDE):

$$\mathrm{d}\vec{X}_l(s) = \vec{v}(\vec{X}_l(s), s)\mathrm{d}s + \sigma_x \mathrm{d}W_l^x(s)$$
$$= \sum_{\mathbf{k} \in \mathbf{K}} \hat{v}_{\mathbf{k}}(s) \exp(\mathrm{i}\vec{k} \cdot \vec{X}_l(s))\vec{r}_{\mathbf{k}}\mathrm{d}s + \sigma_x \mathrm{d}W_l^x(s), \quad l = 1, \ldots, L. \qquad (1.3)$$

Here is the key mathematical observation which forms the basis for the analysis presented in the remainder of the paper. Although the evolution of $\vec{X}_l(s)$ is nonlinear, it depends linearly on the signal processes $\hat{v}_{\mathbf{k}}(s)$. Therefore, our signal-observation system (1.1) and (1.3) fits into the category of "conditional Gaussian process", introduced by [16]. Such signal-observation processes are linear once the observations are given. With this conditional property, the conditional distribution of the signal with given observations, i.e. the posterior distribution, is Gaussian, as long as the signal process is initially distributed as Gaussian. Moreover, the posterior mean and covariance of this distribution follow explicit nonlinear differential equations with random coefficients, which can be seen as a generalization of those of a Kalman-Bucy filter.

Two significant advantages over other filters for Lagrangian data are possessed by our filter produced by this conditional Gaussian theory. First, it is an optimal or exact filter, as there are no approximation errors appearing as in EnKF for non-Gaussian models. Second, its evolution involves rather elementary formulas, which enables detailed theoretical analysis and simple numerical algorithms. These advantages will be exploited to study the filtering skill and explore the information barriers in the remainder of the paper.

In order to give an asymptotic characterization of the filter for $L \to \infty$, the assumption that the tracers are uniformly deployed initially is assumed. This is the most interesting setting in practice, because intuitively it generates a more complete observation, which is also numerically verified in [7] under a different setting. Moreover, Theorem 3.1 proved below establishes ergodicity of the uniform distribution for a finite number of tracer in the periodic setting, in other words, the uniform distribution is the attracting equilibrium distribution for the tracers' projection on the periodic domain. From this, we can establish a formal mean field dynamics of the random filter and use it to do asymptotic analysis.

Here, we mention qualitatively the main results of this paper. Theorems are proved to support these points.

- The posterior covariance matrix approaches a deterministic matrix R_L, with R_L being a diagonal matrix which scales as $L^{-1/2}$, cf. Theorem 3.3, part (i);
- The posterior mean, i.e. the maximum likelihood estimator produced by the filter, converges to the true value of the signal, cf. Theorem 3.3, part (ii) and (iii);

- The total uncertainty reduction gained by the observations, being measured either in relative entropy or mutual information, asymptotically increases as $\frac{1}{4}|\mathbf{K}| \ln L$, where $|\mathbf{K}|$ is the number of modes included; in other words, we gain only $\frac{1}{4}$ nat of information for each mode with each additional order of magnitude of L, cf. Corollary 3.4.

Judging from the last item, the marginal information gained from each additional tracer increases very slowly with the number of tracer, L. This will form an information barrier in practice, when only reasonable amount of tracers are available.

The rest of this paper is organized as follows. In Section 2, we set-up the random incompressible model, establish the filter using conditional Gaussian theory and set-up the uncertainty quantification with explicit formulas. Section 3 states our theoretical results. The major lines of proof are in Section 4 with additional verification done in the appendix. The corresponding numerical experiments, showing the uncertainty reduction and information barrier, are included in Section 5. We conclude in Section 6.

2. Set-up and terminology

2.1. Random incompressible flow model

To set-up the intuition behind (1.1), consider a two dimensional version of this incompressible random flow model with $\vec{k} = (k_1, k_2)^T$,

$$\vec{v}(\vec{x}, s) = \hat{v}_{\vec{0},1}(s) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \hat{v}_{\vec{0},2}(s) \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \sum_{|k_1| \vee |k_2| \leq N, \vec{k} \neq \vec{0}} \hat{v}_{\vec{k}}(s) \exp(i\vec{k} \cdot \vec{x}) \frac{i\vec{k}^{\perp}}{|\vec{k}|}. \tag{2.1}$$

where $\vec{k}^{\perp} = (-k_2, k_1)^T$ and $|k_1| \vee |k_2|$ denotes the maximum value of $|k_1|$ and $|k_2|$. The first two real modes represent a random background mean sweep; the remaining complex-valued modes form conjugate pairs and represent superposition of random plane waves. (2.1) can be easily generalized to any higher dimension, which has a general simplified form as (1.1): $\sum_{\mathbf{k} \in \mathbf{K}} \hat{v}_{\mathbf{k}}(s) \exp(i\vec{k} \cdot \vec{x}) \vec{r}_{\mathbf{k}}$, as long as the following properties of (2.1) are preserved:

(i) $\vec{r}_{\mathbf{k}}$ is orthogonal to the wavenumber $\vec{k}$; when there are multiple $\mathbf{k}$s correspond to one same wavenumber $\vec{k}$, we assume $\vec{r}_{\mathbf{k}}$ are orthogonal with each other without loss of generality;

(ii) The modes with wavenumber $\vec{k} = \vec{0}$ are real-valued;

(iii) The complex-valued modes come into conjugate pairs, as $\hat{v}_{\mathbf{k}} = \hat{v}_{-\mathbf{k}}^*, \hat{r}_{\mathbf{k}} = \hat{r}_{-\mathbf{k}}^*$, where $-\mathbf{k}$ stands for the conjugate index of $\mathbf{k}$.

For the sake of conciseness, we leave the lengthy but simple model description in Appendix A.1.

In the following, we group all the signal processes, $\{\hat{v}_{\mathbf{k}}(s)\}_{\mathbf{k} \in \mathbf{K}}$, into a $|\mathbf{K}|$-dim vector process

$$\mathbf{U}_s = \begin{bmatrix} \vdots \\ \hat{v}_{\mathbf{k}}(s) \\ \vdots \end{bmatrix}.$$

There exists an obvious bijection between $\vec{v}(\,\cdot\,, s)$ and $\mathbf{U}_s$: the latter can be obtained by taking Fourier transform of the former while the former can be written as

$$\vec{v}(\vec{x}, s) = \sum_{\mathbf{k} \in \mathbf{K}} \hat{v}_{\mathbf{k}}(s) \exp(i\vec{k} \cdot \vec{x})\vec{r}_{\mathbf{k}} = P_X(\vec{x}) \cdot \mathbf{U}_s, \tag{2.2}$$

with $P_X(\vec{x}) = [\cdots, \exp(i\vec{k} \cdot \vec{x})\vec{r}_{\mathbf{k}}, \cdots]$ being a $d \times |\mathbf{K}|$-dim matrix.

Since each of its components follows (1.2), $\mathbf{U}_s$ is the solution to the following linear SDE:

$$d\mathbf{U}_s = -\Gamma\mathbf{U}_s ds + \mathbf{F}_s ds + \Sigma_u dB_s, \tag{2.3}$$

where Γ is a strictly positive diagonal matrix, $\mathbf{F}_s$ represents the deterministic forcing, and B_s is a canonical $|\mathbf{K}|$-dim real-valued Wiener process. In order to preserve the conjugate pairs, Σ_u will have $\sigma_{\mathbf{k}}$ as $(\mathbf{k}, \mathbf{k})$th term for the real-valued modes, and $\{\mathbf{k}, -\mathbf{k}\}^2$ 2×2 sub-block being $\frac{1}{\sqrt{2}}\sigma_{\mathbf{k}}\left[\begin{smallmatrix} 1 & i \\ 1 & -i \end{smallmatrix}\right]$ for the complex conjugate pairs; the other components are 0. Notice that $\mathbf{\Sigma} = \Sigma_u\Sigma_u^*$ is a real diagonal matrix with $(\mathbf{k}, \mathbf{k})$th entry being $\sigma_{\mathbf{k}}^2$. More explanation of Σ_u's form can be found in Appendix A.1. Here, we assume F_s is periodic with period T.

Without any observations, the least biased status of the velocity field, or equivalently $\mathbf{U}_s$, is given by its time-dependent equilibrium distribution, also known as the climatology in climate science. This statistical attractor of $\mathbf{U}_s$ can be computed by applying the Fokker-Planck equation to (2.3). Since (2.3) is linear, the equilibrium distribution is Gaussian $\pi_t^{att} = \mathcal{N}(\vec{m}_s^{att}, \vec{R}_{att})$, where $\vec{m}_s^{att}$ has the same period with $\mathbf{F}_s$ and satisfies

$$d\vec{m}_s^{att} = -\Gamma\vec{m}_s^{att} ds + \mathbf{F}_s ds. \tag{2.4}$$

The covariance has no dependence of the deterministic forcing and is the solution of the system

$$0 = -\Gamma R_{att} - R_{att}\Gamma^* + \Sigma_u\Sigma_u^*. \tag{2.5}$$

Making use of the fact that Γ is diagonal, one can easily solve

$$\vec{m}_t^{att} = \int_{-\infty}^{t} \exp(-\Gamma(t - s))\mathbf{F}_s ds, \quad R_{att} = \frac{1}{2}\Gamma^{-1}\Sigma_u\Sigma_u^*.$$

Usually the information obtained from the equilibrium distribution is very limited. This distribution usually serves as the prior knowledge of $\mathbf{U}_0$ before observations are taken. We assume this below for simplicity in explanation.

2.2. Filtering with noisy Lagrangian tracers

Given one realization of the velocity field, $\vec{v}_{s \geq 0}$, the trajectory of each one of the L noisy tracers can be modeled as the solution to the following SDE:

$$\begin{aligned} d\vec{X}_l(s) &= \vec{v}(\vec{X}_l(s), s)ds + \sigma_x dW_l^x(s) \\ &= P_X(\vec{X}_l(s))\mathbf{U}_s ds + \sigma_x dW_l^x(s), \quad l = 1, \ldots, L. \end{aligned} \tag{2.6}$$

The $\{W_l^x(s)\}_{l \leq L}$ are 2-dim real-valued independent Wiener processes. They describe the random instrumental error over the tracers. In the language of filtering, they are called the observation errors.

We will also group all $\vec{X}_l(s)$ into an observation vector $\mathbf{X}_s$ of dimension $Ld \times 1$:

$$\mathbf{X}_s := \begin{bmatrix} \vec{X}_1(s) \\ \vdots \\ \vec{X}_L(s) \end{bmatrix}.$$

Based on (2.6) and (2.2), $\mathbf{X}_s$ follows

$$\mathrm{d}\mathbf{X}_s = \mathbf{P}_X(\mathbf{X}_s)\mathbf{U}_s\mathrm{d}s + \sigma_x\mathrm{d}W_X(s). \tag{2.7}$$

where $\mathbf{P}_X(\mathbf{X}_s)$ is a $Ld \times |\mathbf{K}|$ matrix-valued function of $\mathbf{X}_s$ defined as follow with W_X:

$$\mathbf{P}_X(\mathbf{X}_s) = \begin{bmatrix} P_X(\vec{X}_1(s)) \\ \vdots \\ P_X(\vec{X}_L(s)) \end{bmatrix} = \begin{bmatrix} \cdots & \exp(i\vec{k} \cdot \vec{X}_1(s))\vec{r}_\mathbf{k} & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \exp(i\vec{k} \cdot \vec{X}_L(s))\vec{r}_\mathbf{k} & \cdots \end{bmatrix}, \quad W_X(s) = \begin{bmatrix} W_1^x(s) \\ \vdots \\ W_L^x(s) \end{bmatrix}. \tag{2.8}$$

From previous discussion, filtering $\vec{v}(\vec{x}, t)$ with $(\vec{X}_1(s), \ldots, \vec{X}_L(s))_{s \leq t}$ is equivalent to filtering $\mathbf{U}_t$ using $\mathbf{X}_{s \leq t}$, which have their joint evolution described by (2.3) and (2.7).

The nonlinearity of this problem comes from the tracers, as the evolution of $\mathbf{X}_s$ depends nonlinearly on itself in (2.7). However, once its realization is fixed, both (2.3) and (2.7) will be linear in $\mathbf{U}_s$. Moreover the initial distribution of $\mathbf{U}_0$ is taken as π_0^{att}, which is Gaussian.

Therefore our problem fits tightly into the condition of Theorem 12.7 of [16]. Since this theorem is of pivotal importance, we attach one simplified version of it as Theorem Appendix A.1. Based on its statement, conditioned on the observation of $\mathbf{X}_{s \leq t}$, $\mathbf{U}_t$ is distributed as $\pi_{t|t} = \mathcal{N}(\vec{m}_t, R_t)$. The posterior mean, $\vec{m}_t$, and covariance, R_t, are solutions to the following:

$$\mathrm{d}R_t = [-\Gamma R_t - R_t\Gamma^* + \mathbf{\Sigma} - \sigma_x^{-2}R_t\mathbf{P}(t)R_t]\mathrm{d}t, \tag{2.9}$$

$$\mathrm{d}\vec{m}_t = -\Gamma\vec{m}_t\mathrm{d}t + \mathbf{F}_t\mathrm{d}t + \sigma_x^{-2}R_t\mathbf{P}_X(\mathbf{X}_t)(\mathrm{d}\mathbf{X}_s - \mathbf{P}_X(\mathbf{X}_t)\mathrm{d}t), \tag{2.10}$$

where we denote $\mathbf{\Sigma} := \Sigma_v\Sigma_v^*$ and $\mathbf{P}(t) = \mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t)$, which has explicit random matrix entries as

$$[\mathbf{P}(t)]_{\mathbf{j},\mathbf{k}} = \sum_{l=1}^{L} \exp((\vec{k} - \vec{j}) \cdot \vec{X}_l(t))(r_\mathbf{j}^* r_\mathbf{k}). \tag{2.11}$$

It is worth noticing that since P_X is a periodic function, it is equivalent to observe either $\vec{X}_l(t)$ or its projection on torus $\mathbb{T}^d = (0, 2\pi]^d$.

Without lost of generality, we assume the initial locations of the tracers $\vec{X}_1(0), \ldots, \vec{X}_L(0)$ are distributed independently. Then conditioned on the realization of $\vec{v}_{s \geq 0}$, $\vec{X}_1(s), \ldots, \vec{X}_L(s)$ are conditionally independent processes. Their joint law conditioned on the realization of $\vec{v}_{s \geq 0}$ will be denoted as $\mathbb{P}_{\vec{v}_{s \geq 0}}$ and the expectation taken with it denoted as $\mathbb{E}_{\vec{v}_{s \geq 0}}$. Regarding

regularity, $\mathbb{P}_{\vec{v}_{s\geq 0}}$ and other conditional laws in this paper are all regular versions of conditional probabilities; their existence relies on the fact that the underlying space is Polish, cf. Theorem 6.3 [17]. In order to differentiate this distribution from the joint law of $\vec{v}_{s\geq 0}$ and $(\vec{X}_l(s))_{0\leq s,l\leq L}$, we will call the latter the super ensemble probability as it includes all realization of $\vec{v}_{s\geq 0}$ and denote it as $\mathbb{P}$.

2.3. Uncertainty reduction

Denote by $\pi_{t|t}$ the posterior distribution of the velocity field with observation from tracers, which is a Gaussian distribution with mean $\vec{m}_t$ and covariance R_t. One practical question is how much uncertainty is reduced in $\pi_{t|t}$ compared with the unfiltered velocity field, meaning the absence of observations. Here the prior distribution is given by the equilibrium distribution π_t^{att}, with mean $\vec{m}_t^{att}$ and covariance R_{att}. Empirical information theory, as in [11, 12, 18, 19], is applied to measure this uncertainty reduction.

In the information-theoretic framework, one natural way to measure the lack of information in one probability density q, when the real distribution is p, is through the relative entropy

$$\mathcal{P}(p,q) = \int p \ln \frac{p}{q}.$$

Despite the lack of symmetry in its arguments, the relative entropy, $\mathcal{P}(p,q)$ provides an attractive framework for assessing the uncertainty reduction as well as measuring model error due to its two "distance-like" properties: (i) $\mathcal{P}(p,q)$ is always positive unless $p = q$, and (ii) it is invariant under general nonlinear change of variables. These properties make relative entropy an ideal measurement for uncertainty quantifications purposes.

When both $p \sim \mathcal{N}(\vec{m}_p, R_p)$ and $q \sim \mathcal{N}(\vec{m}_q, R_q)$ are Gaussian, the relative entropy has an explicit formula [20, 21]:

$$\mathcal{P}(p,q) = \left[\tfrac{1}{2}(\vec{m}_p - \vec{m}_q)^T R_q^{-1}(\vec{m}_p - \vec{m}_q)\right] + \tfrac{1}{2}\left[\text{tr}(R_p R_q^{-1}) - N - \ln\det(R_p R_q^{-1})\right], \quad (2.12)$$

where N is the dimension of both the distribution. The first term in brackets in (2.12) is called the "signal", it measures the lack of information in the mean weighted by model covariance; whereas the second term in brackets is called the "dispersion" and involves only the covariance ratio.

For our purposes, of key importance is the so-called Bayesian-update interpretation of relative entropy. It states that if $p = \pi_{t|t}$ is a posterior distribution conditioned on the observation $\mathbf{X}_{s\leq t}$ and q is the corresponding prior distribution, which is the case for $q = \pi_t^{att}$, then $\mathcal{P}(\pi_{t|t}, \pi_t^{att})$ measures the additional information beyond π_t^{att} gained by having observed $\mathbf{X}_{s\leq t}$, namely, the uncertainty reduction.

The above discussions and formulae also apply to complex-valued random vectors, as long as their conjugates are integrated properly. The relative entropy of a random variable $X + Yi$ can be naturally defined as the relative entropy for the real-valued pair (X, Y).

More discussion and verifications of formula (2.12) in this setting can be found in Section Appendix A.2.

Moreover, as the $\mathbf{U}_t$ has its complex entries come in as conjugate pairs, we can apply (2.12) with its mean and covariance under $\pi_{t|t}$ and π_t^{att}:

$$\mathcal{P}(\pi_{t|t}, \pi_t^{att}) = \tfrac{1}{2}\big[(\vec{m}_t - \vec{m}_t^{att})^* R_{att}^{-1}(\vec{m}_t - \vec{m}_t^{att})\big] \qquad \text{Signal}$$
$$+ \tfrac{1}{2}\big[\text{tr}(R_t R_{att}^{-1}) - |\mathbf{K}| - \ln\det(R_t R_{att}^{-1})\big] \qquad \text{Dispersion.} \qquad (2.13)$$

Another important quantity that also characterizes the amount of information of $\mathbf{U}_t$ that $\mathbf{X}_{s\leq t}$ can provide is their mutual information. This quantity is also widely used for other scenarios of uncertainty quantifications, [22, 23, 24, 25]. For two random variable X, Y with joint density p and marginal density p_X, p_Y, the mutual information is defined as:

$$\mathcal{I}(X;Y) := \int p(\mathrm{d}x, \mathrm{d}y) \ln \frac{p(\mathrm{d}x, \mathrm{d}y)}{p_X(\mathrm{d}x)p_Y(\mathrm{d}y)}.$$

In our scenario, we can show that $\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s\leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_t^{att}))$, i.e. the "super-ensemble" expectation value of the relative entropy over all the observations [24]. With direct computation in Lemma Appendix A.3, we also have a concise formula for this quantity:

$$\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s\leq t}) = \frac{1}{2}\ln\det R_{att} - \frac{1}{2}\mathbb{E}\ln\det R_t. \qquad (2.14)$$

Notice the bijection between $\mathbf{U}_s$ and $\vec{v}_s$, so (2.14) is also $\mathcal{I}(\vec{v}_t, (\vec{X}_l(s))_{s\leq t, l\leq L})$.

Note that the relative entropy $\mathcal{P}(\pi_{t|t}, \pi_t^{att})$ is stochastic since it is based on a single realization of $\mathbf{X}_{s\leq t}$ while the mutual information, which averages with respect to all the realizations, is deterministic.

The final remark is that both the relative entropy and mutual information can be applied to any marginal parts of $\mathbf{U}_t$. As these marginal parts are also Gaussian distribution, it suffices to replace the vectors, matrices and dimension in (2.13) and (2.14) with corresponding sub-vector, sub-matrix and dimension. One can read Corollary 3.5 and its proof for a detailed explanation of this.

3. Statements of results

Our results consists of three parts.

The first issue we address before analyzing the filter is the locations of the tracers. This impacts dramatically how the filter performs. Indeed, if all the tracers are clustered in one location, then intuitively we should have little inference of the flow at a distance away. Then in this situation, the filter would perform very differently comparing to an ideal situation when tracers are uniformly distributed in the whole area. Reference such as [5, 7] also address similar issues in a different setting.

Consider the process for the noisy Lagrangian tracer trajectory given in (1.3), our first result indicates that the ideal situation is probably the most interesting case:

Theorem 3.1. *For a.s. realization of $\vec{v}(\vec{x},s)_{s\geq 0}$ formulated by (1.1) , the uniform distribution, λ, on the d-dim torus $\mathbb{T}^d := (0,2\pi]^d$, is the unique invariant measure for process $\tilde{X}_l(t) := \vec{X}_l(t) \mod \mathbb{T}^d$, for any fixed l. Moreover, the tracer's projection on $\mathbb{T}^d$ is geometricly ergodic under the conditional distribution for given $\vec{v}_{s\geq 0}$, i.e. there exits a suitable constant $q < 0$ such that:*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \|\mathbb{P}_{\vec{v}_{s\geq 0}}(\tilde{X}_l(n) \in \cdot) - \lambda\|_{TV} \leq q.$$

Here, $\mathbb{P}_{\vec{v}_{s\geq 0}}(\tilde{X}_l(n) \in \cdot)$ is the conditional distribution of $\tilde{X}_l(n)$ for given $\vec{v}_{s\geq 0}$. The choice of q is independent of $\vec{v}_{s\geq 0}$ and the distribution of $\vec{X}_l(0)$.

The notation $\|\cdot\|_{TV}$ stands for total variation distance. Recall the total variation distance between two measure $\mathbb{P}$ and $\mathbb{Q}$ is defined as:

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} = \sup_{|f|\leq 1} \mathbb{E}_{\mathbb{P}}f - \mathbb{E}_{\mathbb{Q}}f,$$

with the supremum runs over all measurable functions f, $|f| \leq 1$.

Since the tracers are mutually independent, by Theorem 3.1, the distribution of the configuration of their projection on $\mathbb{T}^d$ will approach $\lambda^{\otimes L} := \lambda \otimes \cdots \otimes \lambda$ in a finite amount of time. It is therefore more interesting and theoretically convenient to assume the tracers are distributed at their steady state in the beginning.

On the other hand, in the view of (2.9) and (2.10), our filter process $(\vec{m}_t, R_t)$ is a Markov process. Also by Lemma Appendix A.4, $(\vec{m}_t, R_t)$ are bounded in probability. Therefore even if the tracers are not distributed as $\lambda^{\otimes L}$, we can shift the time till the point that $\lambda^{\otimes L}$ is approximately reached. In other word, assuming the following loses little generality:

Assumption 3.2. *The projection of the initial location of each tracer on $\mathbb{T}^d$ is independently and uniformly distributed on $\mathbb{T}^d$. The prior distribution is chosen as $\mathbf{U}_0 \sim \mathcal{N}(\vec{m}_0, R_0)$ with $|\vec{m}_0| \leq M, m\sqrt{L}^{-1}I \leq R_0 \leq MI$ for some constants $0 < m < M$.*

Under Assumption 3.2, our second result discusses the limiting behavior of the posterior mean and covariance $\vec{m}_t$ and R_t for large L. Recall that a sequence of random variables X_L converges to a in probability, written as $X_L \to a$ in $\mathbb{P}$, if for any $\epsilon > 0$, $\mathbb{P}(|X_n - a| \geq \epsilon) \to 0$ as $L \to \infty$. We will say X_L approaches Y_L in $\mathbb{P}$ if $X_L - Y_L \to 0$ in $\mathbb{P}$. We will say a deterministic sequence x_L grows asymptotically as $f(L)$ if their ratio goes to 1 as $L \to \infty$. We will use $\|\cdot\|$ to denote the L^2 norm of a matrix.

Theorem 3.3. *When filtering a random incompressible flow $\vec{v}(\vec{x},s)$ described by (1.1) using L Lagrangian tracers under Assumption 3.2, the following hold as $L \to \infty$:*

(i) *For any fixed $t > 0$ and a.s. realization of $\vec{v}_{s\geq 0}$, the posterior covariance R_t approaches R_L in the conditional distribution with given $\vec{v}_{s\geq 0}$:*

$$\|R_L^{-1}\|\|R_t - R_L\| \to 0 \quad in \ \mathbb{P}_{\vec{v}_{s\geq 0}}.$$

The sequence R_L are deterministic diagonal matrices with

$$[R_L]_{\mathbf{k},\mathbf{k}} = \frac{\sigma_{\mathbf{k}}^2}{d_{\mathbf{k}} + \sqrt{d_{\mathbf{k}}^2 + L\sigma_x^{-2}\sigma_{\mathbf{k}}^2|\vec{r}_{\mathbf{k}}|^2}}. \tag{3.1}$$

(ii) *For any fixed $t > 0$, the difference between the posterior mean and the signal, $\vec{m}_t - \mathbf{U}_t$, converges to 0 in $\mathbb{P}$, the super ensemble probability that integrates all realization of $\vec{v}_{s\geq 0}$.*

(iii) *There exists a constant time s_0, such that for a.s. fixed realization of $\vec{v}_{s\geq 0}$, $t > s_0$, $\vec{m}_t - \mathbf{U}_t$ converges to 0 in $\mathbb{P}_{\vec{v}_{s\geq 0}}$.*

In other words, asymptotically $\pi_{t|t}$ can be approximated by $\mathcal{N}(\mathbf{U}_t, R_L)$. We only need to plug in the formulas for relative entropy, (2.13), and mutual information, (2.14) to have formally the following corollary on the asymptotic behavior of uncertainty reduction:

Corollary 3.4. *Under the condition of Theorem 3.3, the following hold as $L \to \infty$:*

(i) *The dispersion part of the relative entropy in (2.13) approaches the following in $\mathbb{P}_{\vec{v}_{s\geq 0}}$:*

$$\frac{1}{2}\sum_{\mathbf{k}\in\mathbf{K}}\left(\ln\left(\tfrac{1}{2}(1 + \sqrt{1 + L\sigma_x^{-2}d_{\mathbf{k}}^{-2}\sigma_{\mathbf{k}}^2|\vec{r}_{\mathbf{k}}|^2}\right) - 1\right). \tag{3.2}$$

(ii) *The signal part in (2.13) will converge to $\frac{1}{2}(\mathbf{U}_t - \vec{m}_t^{att})^*R_{att}^{-1}(\mathbf{U}_t - \vec{m}_t^{att})$ in $\mathbb{P}$ for $t > 0$ and $\mathbb{P}_{\vec{v}_{s\geq 0}}$ for a.s. $\vec{v}_{s\geq 0}$ when $t > s_0$, with s_0 defined by Theorem 3.3 part (iii).*

(iii) *The mutual information between the velocity field $\vec{v}_t$ and the observations $(\vec{X}_l(s))_{s\leq t, l\leq L}$ grows asymptocally as $\frac{1}{4}|\mathbf{K}|\ln L$.*

Notice that under $\mathcal{N}(\mathbf{U}_t, R_L)$, each component is independent as R_L is diagonal, so there is the following version of Corollary 3.4 for marginal processes:

Corollary 3.5. *Under the condition of Theorem 3.3, let $\widetilde{\mathbf{K}} \subset \mathbf{K}$ be a subset that does not separate any conjugate pairs. Denote $\widetilde{\mathbf{U}}_t = [\cdots, \hat{v}_{\mathbf{k}}(t), \cdots]_{\mathbf{k}\in\widetilde{\mathbf{K}}}^T$, then the following holds as $L \to \infty$:*

(i) *The dispersion part of the relative entropy of $\widetilde{\mathbf{U}}_t$ approaches the following in $\mathbb{P}_{\vec{v}_{s\geq 0}}$:*

$$\frac{1}{2}\sum_{\mathbf{k}\in\widetilde{\mathbf{K}}}\left(\ln\left(\tfrac{1}{2}(1 + \sqrt{1 + L\sigma_x^{-2}d_{\mathbf{k}}^{-2}\sigma_{\mathbf{k}}^2|\vec{r}_{\mathbf{k}}|^2}\right) - 1\right). \tag{3.3}$$

(ii) *The signal part of the relative entropy of $\widetilde{\mathbf{U}}_t$ will converge to $\frac{1}{2}(\widetilde{\mathbf{U}}_t - \tilde{m}_t^{att})^*\tilde{R}_{att}^{-1}(\widetilde{\mathbf{U}}_t - \tilde{m}_t^{att})$ in $\mathbb{P}$ for $t > 0$ and $\mathbb{P}_{\vec{v}_{s\geq 0}}$ for a.s. $\vec{v}_{s\geq 0}$ when $t > s_0$, with s_0 defined by Theorem 3.3 part (iii). $\tilde{m}_t^{att}$ is a sub-vector of $\vec{m}_t^{att}$ and $\tilde{R}_{att}$ is a diagonal subblock of R_{att}, with indexes in $\widetilde{\mathbf{K}}$.*

(iii) *The mutual information between the velocity field $\vec{v}_t$ and the observations $(\vec{X}_l(s))_{s\leq t, l\leq L}$ grows asymptocally as $\frac{1}{4}|\widetilde{\mathbf{K}}|\ln L$.*

4. Proof of theorems

4.1. Conditional geometric ergodicity of tracers

In this section we will prove Theorem 3.1. We will use the following notation:

$$\mathbb{P}^x_{\vec{v},n}(\,\cdot\,) := \mathbb{P}_{\vec{v}_{s\geq 0}}(\tilde{X}_l(n) \in \,\cdot\,|\tilde{X}_l(0) = x), \quad \forall x \in \mathbb{T}^d.$$

Recall that $\tilde{X}_l(n)$ is the process associated with the noisy Lagrangian tracer equation's projection on $\mathbb{T}^d$:

$$\tilde{X}_l(t) = \tilde{X}_l(0) + \int_0^t \vec{v}(\tilde{X}_s, s)ds + \int_0^t \sigma_x dW^x_l(s) \qquad \text{mod } \mathbb{T}^d.$$

We also denote the supremum velocity in a velocity field as:

$$|\vec{v}_s|_\infty := \sup_{\vec{x}} |\vec{v}(\vec{x}, s)|.$$

We write $p : \mathbb{R}^d \to \mathbb{T}^d$ as the canonical covering map and we denote $p^{-1}x$ as the unique inverse image of x in $(0, 2\pi]^d$.

Also recall that a coupling of two measures is a joint measure with its marginal distributions being the two measures respectively. The essence of the following proof is constructing a coupling between the trajectories of two tracers $\tilde{X}_l(s)$ starting from two different locations but transported by the same realization of $\vec{v}_{s\geq 0}$. By the Goldstein's theorem, Section 14 [26], we have

$$\|\mathbb{P}^x_{\vec{v},n} - \mathbb{P}^y_{\vec{v},n}\|_{TV} \leq 2\mathbb{P}_{\vec{v}_{s\geq 0}}(\text{two tracers are coupled into one trajectory from time } n),$$

so if there exist a coupling such that two tracers are eventually coupled, we will have our claim. In a rough sense, we will construct this coupling one time step by one time step. The key argument in (4.2) below implies two tracers can be coupled at the end of one time step with a probability bounded away from 0 no matter where they start. Hence the probability that the tracers are not coupled will decay geometrically in time. And in order to show (4.2), we will show it is measure theoretically equivalent to the case when $\vec{v}_s$ is trivial using the Girsanov theorem as long as $\vec{v}_s$ is bounded in L^2.

Proof. To show the invariance for the measure λ, notice that $\vec{v}(\vec{x}, s)$ is spatially periodic, it suffices for us to verify that locally the density $f \equiv 1$ is a solution to the Fokker-Plank equation for (2.6). Using the Fokker-Plank equation on $\mathbb{T}^d$, see [27] Section 8.1, the evolution of a density under $\mathbb{P}_{\vec{v}_{s\geq 0}}$ is

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^2 \frac{\partial}{\partial x_i}[\vec{v}_i(\vec{x}, t)f(\vec{x}, t)] + \frac{\sigma_x^2}{2}\sum_{i=1}^2\sum_{j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j}[\delta_{ij}f(\vec{x}, t)].$$

When $f \equiv 1$, due to $\nabla \cdot \vec{v}_t \equiv 0$, the right hand side above is 0, hence $f \equiv 1$ is a time independent invariant solution of the equation above. The uniqueness of invariant measure is proved by the geometric ergodicity argument below.

To get the geometric ergodicity of tracer, notice that $\vec{v}(\vec{x}, s)$ is formulated as in (2.2), and $\mathbf{U}_s$ is an O.U. process with strictly positive damping, so $\vec{v}_s$ is ergodic. By the Birkhoff's ergodic theorem, if we denote the periodicity of π_s^{att} as T, then

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t |\vec{v}_s|_\infty^2 \mathrm{d}s \to \frac{1}{T} \int_0^T \mathbb{E}_{\pi_s^{att}} |P_X(\vec{x}) \cdot \mathbf{U}_s|_\infty \mathrm{d}s \quad \mathbb{P}\text{-}a.s.$$

Let N be the value twice of the right hand side, and denote

$$k(\vec{v}, n) := \#\left\{ k : 0 \leq k \leq n - 1, \int_k^{k+1} |\vec{v}(s)|_\infty^2 \mathrm{d}s < N \right\}.$$

Then using $\int_0^n |\vec{v}_s|_\infty^2 \geq 2N(n - k(\vec{v}, n))$, we can conclude

$$\liminf_{n \to \infty} \frac{k(\vec{v}, n)}{n} \geq \frac{1}{2} \quad \mathbb{P}\text{-}a.s. \tag{4.1}$$

We fix a $\vec{v}_{s \geq 0}$ such the event of (4.1) takes place. We will first prove the following claim: there exists a uniform $c > 0$ such that for any x and y in $\mathbb{T}^d$,

$$\|\mathbb{P}_{v,1}^x - \mathbb{P}_{v,1}^y\|_{TV} \leq 2 - c \mathbf{1}_{\int_0^1 |\vec{v}(s)|_\infty^2 \mathrm{d}s < N}. \tag{4.2}$$

This claim is trivial for $\{\int_0^1 |\vec{v}(s)|_\infty^2 \mathrm{d}s \geq N\}$, so we assume the opposite holds. Consider the following joint processes on $\mathbb{R}^d$,

$$\vec{X}_s = p^{-1}x + \sigma_x W_s, \quad \vec{Y}_s = p^{-1}y + \sigma_x W_s'$$

where W_s and W_s' are two d-dim canonical Wiener processes. Denote their laws respectively as $\mathbb{Q}_x$ and $\mathbb{Q}_y$. Since $p^{-1}x$ and $p^{-1}y$ lie in $(0, 2\pi]^d$, so $|p^{-1}x - p^{-1}y| \leq 2\sqrt{d}\pi$. Then since X_1 and Y_1 are distributed as $\mathcal{N}(p^{-1}x, \sigma_x^2 I_d), \mathcal{N}(p^{-1}y, \sigma_x^2 I_d)$, with $|p^{-1}x - p^{-1}y| \leq 2\sqrt{d}\pi$ we can find an universal $a > 0$ such that

$$\|\mathcal{N}(p^{-1}x, \sigma_x^2 I_d) - \mathcal{N}(p^{-1}y, \sigma_x^2 I_d)\|_{TV} \leq 2 - a, \quad \forall x, y \in \mathbb{T}^2.$$

Due to the relation between coupling and total variation, cf. Theorem (I.5.2) in [26], we can find a maximal coupling $Q^{x,y}$ of the random variables $\vec{X}_1, \vec{Y}_1$: $Q^{x,y}(\vec{X}_1 = \vec{Y}_1) \geq a$. Then conditioned on the value of $\vec{X}_1, \vec{Y}_1$, we can extend it to a coupling, $\mathbb{Q}_1^{x,y}$, of processes $(\vec{X}_s)_{s \leq 1}, (\vec{Y}_s)_{s \leq 1}$, explicitly defined as:

$$\mathbb{Q}_1^{x,y}(\vec{X}_{s \leq 1} \in \mathrm{d}u, \vec{Y}_{s \leq 1} \in \mathrm{d}v)$$
$$:= \int_{w,z} Q^{x,y}(\vec{X}_1 \in \mathrm{d}w, \vec{Y}_1 \in \mathrm{d}z) \mathbb{Q}_x(\vec{X}_{s \leq 1} \in \mathrm{d}u | \vec{X}_1 = w) \mathbb{Q}_y(\vec{Y}_{s \leq 1} \in \mathrm{d}v | \vec{Y}_1 = z).$$

Then as $Q^{x,y}$ is the marginal distribution of $(\vec{X}_1, \vec{Y}_1)$ for $\mathbb{Q}_1^{x,y}$:

$$\mathbb{Q}_1^{x,y}(\vec{X}_1 = \vec{Y}_1) = Q^{x,y}(\vec{X}_1 = \vec{Y}_1) \geq a, \quad \forall x, y \in \mathbb{T}^d. \tag{4.3}$$

Now denote

$$z_{\vec{v}}(t) := \exp(\mathcal{M}_t - \tfrac{1}{2}\langle \mathcal{M} \rangle_t), \quad \mathcal{M}_t = \sigma_x^{-1}\int_0^t \vec{v}(\vec{X}_s, s)\mathrm{d}W_s + \vec{v}(\vec{Y}_s, s)\mathrm{d}W_s'. \tag{4.4}$$

Since $\int_0^1 (|\vec{v}(\vec{X}_s, s)|^2 + |\vec{v}(\vec{Y}_s, s)|^2)\mathrm{d}s \leq 2\int_0^1 |\vec{v}_s|_\infty^2 \mathrm{d}s < 2N$, the Novikov condition, Corollary 3.5.13 [28], holds for $z_v(t)$, meaning $z_v(t)$ is a martingale. By the Girsanov theorem, under measure μ defined by $d\mu := z_{\vec{v}}(1)d\mathbb{Q}_1^{x,y}$, the processes

$$\tilde{W}_t = W_t - \sigma_x^{-1}\int_0^t \vec{v}(\vec{X}_s, s)\mathrm{d}s, \quad \tilde{W}_t' = W_t' - \sigma_x^{-1}\int_0^t \vec{v}(\vec{Y}_s, s)\mathrm{d}s$$

are canonical Wiener processes. The $\vec{X}_s, \vec{Y}_s$ on the other hands are solutions to the equations:

$$\vec{X}_t = p^{-1}x + \int_0^t \vec{v}(\vec{X}_s, s)\mathrm{d}s + \sigma_x\tilde{W}_t, \quad \vec{Y}_t = p^{-1}y + \int_0^t \vec{v}(\vec{Y}_s, s)\mathrm{d}s + \sigma_x\tilde{W}_t'.$$

So $\vec{X}_s$ and $\vec{Y}_s$ can be seen as lifts of the processes described by (2.6) starting from point x and y. In other words, the joint law of $(p\vec{X}_1, p\vec{Y}_1)$ is a coupling of $\mathbb{P}_{\vec{v},1}^x$ and $\mathbb{P}_{\vec{v},1}^y$. Therefore

$$\|\mathbb{P}_{\vec{v},1}^x - \mathbb{P}_{\vec{v},1}^y\|_{TV} = \sup_{|f|\leq 1}\mathbb{P}_{\vec{v},1}^x f - \mathbb{P}_{\vec{v},1}^y f = \sup_{|f|\leq 1}\mathbb{E}_\mu(f(\vec{X}_1) - f(\vec{Y}_1)) \leq 2\mu(\vec{X}_1 \neq \vec{Y}_1). \tag{4.5}$$

On the other hand, let $M = 2\sigma_x^{-1}\sqrt{N/a}$, since under $\mathbb{Q}_1^{x,y}$ for $t \leq 1$,

$$\langle \mathcal{M} \rangle_t = \mathbb{E}_\mathbb{Q}\mathcal{M}_t^2 = \sigma_x^{-2}\mathbb{E}_\mathbb{Q}\int_0^t (|\vec{v}(\vec{X}_s, s)|^2 + |\vec{v}(\vec{Y}_s, s)|^2)\mathrm{d}s \leq 2\sigma_x^{-2}\int_0^t |\vec{v}_s|_\infty^2 \mathrm{d}s \leq 2\sigma_x^{-2}N.$$

Using the definition of $z_v(1)$, (4.4), and the Chebyshev's inequality,

$$\mathbb{Q}_1^{x,y}\big(z_v(1) \leq \exp(-M - \sigma_x^{-2}N)\big) \leq \mathbb{Q}_1^{x,y}(\mathcal{M}_1 \leq -M) \leq \frac{\langle \mathcal{M} \rangle_1}{M^2} = \frac{1}{2}a$$

From this and (4.3) we derive

$$\begin{aligned}
\mu(\vec{X}_1 = \vec{Y}_1) &= \int 1_{\vec{X}_1 = \vec{Y}_1}z_v(1) \cdot d\mathbb{Q}_1^{x,y}\\
&\geq \mathbb{Q}_1^{x,y}(\vec{X}_1 = \vec{Y}_1, z_v(1) \geq \exp(-M - \sigma_x^{-2}N))\exp(-M - \sigma_x^{-2}N)\\
&\geq \big(\mathbb{Q}_1^{x,y}(\vec{X}_1 = \vec{Y}_1) - \mathbb{Q}_1^{x,y}(z_v(1) \leq \exp(-M - \sigma_x^{-2}N))\big)\exp(-M - \sigma_x^{-2}N)\\
&\geq \frac{1}{2}a\exp(-M - \sigma_x^{-2}N)
\end{aligned}$$

Based on (4.5), $c = a\exp(-M - \sigma_x^{-2}N) > 0$ satisfies our claim (4.2).

The next step is applying (4.2) at each time step and use the Markov property to form an induction. Denote

$$\mathbb{Q}^x_{\vec{v},n-1} = \mathbb{P}^x_{\vec{v},n-1} \wedge \lambda, \quad \tilde{\mathbb{P}}^x_{\vec{v},n-1} = \mathbb{P}^x_{\vec{v},n-1} - \mathbb{Q}^x_{\vec{v},n-1}, \quad \tilde{\lambda}_{n-1} = \lambda - \mathbb{Q}^x_{\vec{v},n-1}.$$

In other words, the density of $\mathbb{Q}^x_{\vec{v},n-1}$ is the minimum of the ones of $\mathbb{P}^x_{\vec{v},n-1}$ and λ; it represents the maximum volume of measure that $\mathbb{P}^x_{\vec{v},n-1}$ and λ can be coupled with. Notice that,

$$\tilde{\mathbb{P}}^x_{\vec{v},n-1}(\mathbb{T}^d) = \tilde{\lambda}_{n-1}(\mathbb{T}^d) = \frac{1}{2}\|\mathbb{P}^x_{\vec{v},n-1} - \lambda\|_{TV}. \tag{4.6}$$

Then by letting

$$\tilde{\mathbb{R}}^x_{\vec{v},n-1}(dx, dy) := \begin{cases} \frac{1}{\tilde{\lambda}_{n-1}(\mathbb{T}^d)}\tilde{\mathbb{P}}^x_{\vec{v},n-1}(dx)\tilde{\lambda}_{n-1}(dy) & \tilde{\lambda}_{n-1}(\Omega) > 0; \\ 0 & \tilde{\lambda}_{n-1}(\Omega) = 0. \end{cases}$$

one can verify that $\tilde{\mathbb{R}}^x_{\vec{v},n-1}$ is a coupling of $\tilde{\mathbb{P}}^x_{\vec{v},n-1}$ and $\tilde{\lambda}_{n-1}$, that is, its marginal distribution is $\tilde{\mathbb{P}}^x_{\vec{v},n-1}$ and $\tilde{\lambda}_{n-1}$ in x and y direction.

Notice that by the Markov property, using the shift operator $\theta : (\theta^n \vec{v})(s) = \vec{v}(s + n)$, the following equation holds by conditioning on the value of $\vec{X}_l(n-1)$,

$$\mathbb{P}^x_{\vec{v},n}f = \int \mathbb{P}^x_{\vec{v},n-1}(dy)\mathbb{P}^y_{\theta^n\vec{v},1}f.$$

Therefore for any measurable f, $|f| \leq 1$, making use of that λ is invariant under $\mathbb{P}^x_{\vec{v},n-1}$

$$\begin{aligned}
\mathbb{P}^x_{\vec{v},n}f - \lambda f &= \int \mathbb{P}^x_{\vec{v},n-1}(dy)\mathbb{P}^y_{\theta^{n-1}\vec{v},1}f - \lambda(dy)\mathbb{P}^y_{\theta^{n-1}\vec{v},1}f \\
&= \int \tilde{\mathbb{P}}^x_{\vec{v},n-1}(dy)\mathbb{P}^y_{\theta^{n-1}\vec{v},1}f - \tilde{\lambda}_{n-1}(dz)\mathbb{P}^z_{\theta^{n-1}\vec{v},1}f \\
&= \int \tilde{\mathbb{R}}^x_{\vec{v},n-1}(dy, dz)(\mathbb{P}^y_{\theta^{n-1}\vec{v},1}f - \mathbb{P}^z_{\theta^{n-1}\vec{v},1}f) \\
&\leq \frac{1}{2}\|\mathbb{P}^x_{\vec{v},n-1} - \lambda\|_{TV}(2 - c1_{\int_{n-1}^n |\vec{v}(s)|^2_\infty ds < N}).
\end{aligned}$$

We apply (4.6) at the last step. By taking the supreme over all measurable f, $|f| \leq 1$, we have

$$\|\mathbb{P}^x_{\vec{v},n} - \lambda\|_{TV} \leq (1 - \tfrac{c}{2}1_{\int_{n-1}^n |\vec{v}(s)|^2_\infty ds < N})\|\mathbb{P}^x_{\vec{v},n-1} - \lambda\|_{TV}.$$

Using induction, $\|\mathbb{P}^x_{\vec{v},n} - \lambda\|_{TV} \leq 2(1 - \tfrac{c}{2})^{k(\vec{v},n)}$. Due to the Markov property on the initial condition,

$$\mathbb{P}_{\vec{v}_{s\leq n}}(\tilde{X}_l(n) \in A) = \int \mathbb{P}(\tilde{X}_l(0) \in dx)\mathbb{P}^x_{\vec{v},n}(A)dx.$$

So apply the Jason's inequality,

$$\|\mathbb{P}_{\vec{v}_{s\leq n}}(\tilde{X}_l(n) \in \cdot) - \lambda\|_{TV} \leq \int \mathbb{P}(\tilde{X}_l(0) \in dx)\|\mathbb{P}^x_{\vec{v},n} - \lambda\|_{TV} \leq 2(1 - \tfrac{c}{2})^{k(\vec{v},n)}.$$

Recalling (4.1), we have

$$\limsup_{n \to \infty} \frac{1}{n} \ln \|\mathbb{P}_{\vec{v}_{s \le n}}(\tilde{X}_l(n) \in \cdot) - \lambda\|_{TV} \le -\frac{1}{2} \ln(1 - \tfrac{c}{2}) < 0.$$

$\square$

4.2. The mean field dynamics of the filter

In this section we will prove Theorem 3.3. Let us first briefly explain the intuition. Since the posterior covariance R_t follows its own dynamics (2.9) but enters in evolution of $\vec{m}_t$ through (2.10), it is intuitive to study R_t first. Recall from (2.9) that R_t is the solution to a Riccati equation with random coefficients:

$$dR_t = [-\Gamma R_t - R_t \Gamma^* + \Sigma - \sigma_x^{-2} R_t \mathbf{P}(t) R_t] dt.$$

The randomness comes from the tracers $\mathbf{X}_t$, which enters the evolution by the second order coefficient

$$\mathbf{P}(t) = \mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t) = \sum_{l=1}^{L} P_X(\vec{X}_l(t))P_X(\vec{X}_l(t)). \tag{4.7}$$

The second equality above comes from the fact that $\mathbf{P}_X(\mathbf{X}_t)$ is a concatenation of $P_X(\vec{X}_l(t))$ as in (2.8). The randomness of $\mathbf{P}(t)$ makes R_t also a random process. However, under Assumption 3.2, by Theorem 3.1, $\tilde{X}_1(t), \ldots, \tilde{X}_L(t)$ are i.i.d. samples from the uniform distribution λ under $\mathbb{P}_{\vec{v}_{s \ge 0}}$. Therefore the value of $\mathbf{P}(t)$ as a sum of i.i.d. random variables will not deviate too far from its mean in $\mathbb{P}_{\vec{v}_{s \ge 0}}$ when L is large. So intuitively, R_t will stay close to the solution of the formal meal field average of (2.9) as

$$dR_t = [-\Gamma R_t - R_t \Gamma + \Sigma - \sigma_x^{-2} L R_t \mathbf{M} R_t] dt, \quad \mathbf{M} := \mathbb{E}_\lambda \mathbf{P}_X^*(X)\mathbf{P}_X(X). \tag{4.8}$$

On the other hand, matrix Riccati equations are known to be a contraction for Hermitian matrices. Therefore, R_t will also converge to R_L in time, which is the stable point of (4.8); it will also be shown in the proof of Theorem 3.3 part i) that this convergence is exponentially fast with a rate which scales as $\sqrt{L}$. Using this accelerating convergence, R_t approaches R_L very rapidly.

Since we will deal with variables that depend on L, we will say some sequence x_L is of order $f(L)$, written as $x_L \sim f(L)$, if $Mf(L) \ge x_L \ge mf(L)$ for some constants $M, m > 0$ as $L \to \infty$.

The next lemma computes explicitly the value of $\mathbf{M}$ and R_L, and also give a probabilistic control over the deviation of $\mathbf{P}(t)$ from $L\mathbf{M}$.

Lemma 4.1. *i)* $\mathbf{M}$ *in (4.8) is a diagonal matrix with* $\mathbf{k}$-$\mathbf{k}$ *th entry* $|r_{\mathbf{k}}|^2$.

ii) The stable point of (4.8), R_L, *is a diagonal matrix with* $\mathbf{k}$-$\mathbf{k}$ *th entry being*

$$[R_L]_{\mathbf{k},\mathbf{k}} = \frac{\sigma_{\mathbf{k}}^2}{d_{\mathbf{k}} + \sqrt{d_{\mathbf{k}}^2 + L\sigma_x^{-2}\sigma_{\mathbf{k}}^2|\vec{r}_{\mathbf{k}}|^2}}.$$

iii) Define $\tilde{P}_s = \mathbf{P}(s) - L\mathbf{M}$, $q_L := L^{-1} \sup_{s \leq t} \|\tilde{P}_s\|$, *then for any* $q > 1/2$, $L^{1-q} q_L \to 0$ *in* $\mathbb{P}_{\vec{v}_{s \geq 0}}$ *for a.s.* $\vec{v}_{s \geq 0}$.

Proof. It suffices for us to prove each part with a fixed realization of $\vec{v}_{s \geq 0}$ so that the claim of Theorem 3.1 holds.

Part i) Recall that by (2.11), $\mathbf{P}(t)$ has each of its matrix entries explicitly as:

$$(\mathbf{P}(t))_{\mathbf{j},\mathbf{k}} = \sum_{l=1}^{L} \exp(\mathrm{i}(\vec{k} - \vec{j}) \cdot \vec{X}_l(t))(\vec{r}_{\mathbf{j}}^* \vec{r}_{\mathbf{k}}).$$

When $\mathbf{j} = \mathbf{k}$, the quantity above is a constant, $L|\vec{r}_{\mathbf{k}}|^2$. Hence $\mathbf{M}$ has the described diagonal terms.

When $\vec{j} = \vec{k}$, $\mathbf{j} \neq \mathbf{k}$, the quantity above is 0 as $\vec{r}_{\mathbf{j}} \perp \vec{r}_{\mathbf{k}}$. (This is why we require $\vec{r}_{\mathbf{k}}$ to be orthogonal to each other if they share the same wavenumber as in Section 2.)

When $\vec{j} \neq \vec{k}$, notice that if we let $Y_l = \exp(\mathrm{i}(\vec{k} - \vec{j}) \cdot \vec{X}_l(t))$, then using $\tilde{X}_l(t) \sim \lambda$ under $\mathbb{P}_{\vec{v}_{s \geq 0}}$, we have

$$\mathbb{E}_{\vec{v}_{s \geq 0}} Y_l = \tfrac{1}{|\mathbb{T}^d|} \int_{\mathbb{T}^d} \exp(\mathrm{i}(\vec{k} - \vec{j}) \cdot x) dx = 0.$$

In summary, $\mathbf{M}$ is diagonal.

Part ii) Since (4.8) has all matrices being diagonal, we can solve the equation in each diagonal entry, which is solving $0 = -2d_{\mathbf{k}} x + \sigma_{\mathbf{k}}^2 - L\sigma_x^{-2}|\vec{r}_{\mathbf{k}}|^2 x^2$. Directly solving these equations gives ii).

Part iii) Notice that $\tilde{P}_s := \mathbf{P}(s) - L\mathbf{M}$ has its diagonal terms being 0, its off diagonal terms the same as $\mathbf{P}(s)$. In order to show our claim of q_L, it suffices to show $L^{-q}([\tilde{P}_s]_{\mathbf{j},\mathbf{k}})_{s \leq t}$ converge to the constant 0 process in $\mathbb{P}_{\vec{v}_{s \geq 0}}$, because by the following we have our claim proved:

$$L^{1-q} q_L = \sup_{s \leq t} L^{-q} \|\tilde{P}_s\| \leq |\mathbf{K}| \max_{\mathbf{j} \neq \mathbf{k}} \sup_{t \leq T} L^{-q} |[\tilde{P}_s]_{\mathbf{j},\mathbf{k}}| \overset{L \to \infty}{\longrightarrow} 0 \quad \text{in} \quad \mathbb{P}_{\vec{v}_{s \geq 0}}.$$

Since $z_s^L := L^{-q} [\tilde{P}_s]_{\mathbf{j},\mathbf{k}}$ are continuous in s, so it suffices to verify the conditions of Theorem 7.5 of [29]. This verification is done by Lemma Appendix A.5. $\qquad\square$

Proof of Theorem 3.3, part (i). Fix any $0 < c_0 \leq \frac{1}{2} \min_{\mathbf{k}} |\vec{r}_{\mathbf{k}}|^2$. In the following, if we say some constant is uniform, then that constant works for all $c_0 \leq \frac{1}{2} \min_{\mathbf{k}} |\vec{r}_{\mathbf{k}}|^2$ because only a bounded set of c_0 is considered. We introduce this technicality because we may decrease c_0 to a smaller value which depends on these constants in part (iii) of this theorem. For part (i) and (ii), any fixed $c_0 \leq \frac{1}{2} \min_{\mathbf{k}} |\vec{r}_{\mathbf{k}}|^2$ will be suffice.

Recall that

$$q_L := L^{-1} \sup_{s \leq t} \|\mathbf{P}(s) - L\mathbf{M}\|,$$

one has

$$L(\mathbf{M} - q_L I) \leq \mathbf{P}(s) \leq L(\mathbf{M} + q_L I), \quad s \leq t.$$

We will focus on the event $\{q_L \leq c_0\}$, which by Lemma 4.1 (iii) has $\mathbb{P}_{\vec{v}_{s \geq 0}}$ probability goes to 1.

By the comparison theorem of Riccati equations, Theorem 4.1.4 of [30], if we let $R_{\pm,t}$ be the solution of

$$dR_{\pm,t} = [-\Gamma R_{\pm,t} - R_{\pm,t}\Gamma + \Sigma - L\sigma_x^{-2}R_{\pm,t}(\mathbf{M} \mp q_L I)R_{\pm,t}]dt, \quad R_{\pm,0} = R_0; \quad (4.9)$$

then R_s is bounded by the two: $R_{-,s} \leq R_s \leq R_{+,s}, s \leq t$.

On the other hand, since $R_{\pm,t}$ follows (4.9), according to Theorem 8.5 [31], they will converge in time to the stable points of (4.9) respectly, denoted by $R_\pm$, which are the solutions to the algebraic Riccati equations:

$$0 = -\Gamma R_\pm - R_\pm\Gamma + \Sigma - L\sigma_x^{-2}R_\pm(\mathbf{M} \mp q_L I)R_\pm,$$

or explicitly:

$$R_\pm = \begin{bmatrix} \ddots & & \\ & \frac{\sigma_{\mathbf{k}}^2}{d_{\mathbf{k}}+\sqrt{d_{\mathbf{k}}^2+L\sigma_x^{-2}\sigma_{\mathbf{k}}^2(|\vec{r}_{\mathbf{k}}|^2 \mp q_L)}} & \\ & & \ddots \end{bmatrix}.$$

By Theorem 8.5 of [31], when $q_L \leq c_0$ the rate that $R_{\pm,t}$ converges to $R_\pm$ in the Thompson's metric is at least

$$\|L\sigma_x^{-2}(\mathbf{M} - c_0I)^{1/2}\Sigma(\mathbf{M} - c_0I)^{1/2}\| \geq \frac{1}{2}L\sigma_x^{-2}\min_{\mathbf{k}}|\vec{r}_{\mathbf{k}}|^2\sigma_{\mathbf{k}}^2 =: \mu_L^2.$$

In other words: $p(R_{\pm,t}, R_\pm) \leq \exp(-2\mu_L t)p(R_0, R_\pm)$. According to [31], for two positive semidefinite Hermitian matrices A, B, the Thompson's metric is defined as:

$$p(A, B) := \ln\inf\{t : A \leq tB, B \leq tA\}.$$

Since $mL^{-1/2}I \leq R_{\pm,0} \leq MI$, when $q_L \leq c_0$ there is a uniform constant D_0 such that $p(R_0, R_\pm) \leq |\ln(D_0\sqrt{L})|$. Therefore $p(R_{\pm,t}, R_\pm) \leq |\ln(D_0\sqrt{L})|\exp(-2\mu_L t)$. Note that if $p(A, B) \leq a$ for two Hermitian matrices, then by definition of the Thompson's metric, $(e^{-a} - 1)B \leq A - B \leq (e^a - 1)B$, which implies $\|A - B\| \leq (e^a - 1)\|B\|$. Hence, $\|R_{\pm,t} - R_\pm\| \leq \|R_\pm\|[\exp(\ln(D_0\sqrt{L})\exp(-2t\mu_L)) - 1]$.

Notice the following bound holds for all $0 \leq x \leq 1, a > 0$,

$$\exp(ax) - 1 = ax + \frac{1}{2}(ax)^2 + \frac{1}{3!}(ax)^3 + \dots$$
$$\leq x[a + \frac{1}{2}a^2 + \frac{1}{3}a^3 + \dots] \leq e^a x.$$

Assuming $q_L \leq c_0$, as $\|R_\pm\| \sim 1/\sqrt{L}$ we have for a uniform constant D_1,

$$\|R_{\pm,t} - R_\pm\| \leq D_0\|R_\pm\|\sqrt{L}\exp(-2\mu_L t) = D_1\exp(-2t\mu_L).$$

On the other hand, according to the formulation of $R_\pm$, which is C^1 in q_L, there is a uniform constant C_1 such that $\|R_\pm - R_L\| \leq C_1 \sqrt{L}^{-1} q_L$ when $q_L \leq c_0$. Then from $R_{-,t} \leq R_t \leq R_{+,t}$, we can verify that when $q_L \leq c_0$,

$$\|R_t - R_L\| \leq \|R_{+,t} - R_L\| \vee \|R_{-,t} - R_L\| \leq C_1 \sqrt{L}^{-1} q_L + D_1 \exp(-2\mu_L t). \tag{4.10}$$

On the other hand, as $\|R_L^{-1}\| \sim \sqrt{L}$, there are some constants $\tilde{C}_1, \tilde{D}_1$ so that for any $\xi > 0$,

$$\mathbb{P}_{\vec{v}_{s\geq 0}}(\|R_L^{-1}\|\|R_t - R_L\| \geq \xi) \leq \mathbb{P}_{\vec{v}_{s\geq 0}}(q_L \leq c_0) + \mathbb{P}_{\vec{v}_{s\geq 0}}(\tilde{C}_1 q_L + \tilde{D}_1 \sqrt{L} \exp(-2\mu_L t) \leq \xi),$$

both probabilities on the right go to 0 by Lemma 4.1 and $\mu_L \sim \sqrt{L}$ as $L \to \infty$, therefore we conclude our claim. □

Since $R_t := \mathbb{E}((\mathbf{U}_t - \vec{m}_t) \cdot (\mathbf{U}_t - \vec{m}_t)^* | \mathbf{X}_{s\leq t})$, so using its convergence in the super ensemble probability $\mathbb{P}$ easily leads to the following proof of Theorem 3.3, part (ii).

Proof of Theorem 3.3, part (ii). Because $\mathbb{E}(|\mathbf{U}_t - \vec{m}_t|^2 | \mathbf{X}_{s\leq t}) = \mathrm{tr}(R_t)$. By part (i) of Theorem 3.3 , the law of total probability, i.e. $\mathbb{E}Y = \mathbb{E}\mathbb{E}_{\vec{v}_{s\geq 0}}Y$, and the dominated convergence theorem, for any $\xi > 0$,

$$\mathbb{P}(\|R_t\| > \xi^4) = \mathbb{E}\mathbb{P}_{\vec{v}_{s\geq 0}}(\|R_t\| > \xi^4) \to 0, \quad \text{as} \quad L \to \infty.$$

So for any fixed $\xi > 0$ apply first the law of total probability then the Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}(|\mathbf{U}_t - \vec{m}_t| \geq \xi) &\leq \mathbb{P}(\mathrm{tr}(R_t) \geq \xi^4) + \mathbb{P}(\mathrm{tr}(R_t) \leq \xi^4, |\mathbf{U}_t - \vec{m}_t| \geq \xi) \\
&= \mathbb{P}(\mathrm{tr}(R_t) \geq \xi^4) + \mathbb{E}[\mathbb{E}(1_{\mathrm{tr}(R_t)\leq\xi^4}1_{|\mathbf{U}_t - \vec{m}_t|\geq\xi}|\mathbf{X}_{s\leq t})] \\
&\leq \mathbb{P}(\mathrm{tr}(R_t) \geq \xi^4) + \mathbb{E}[\mathbb{P}(|\mathbf{U}_t - \vec{m}_t| \geq \xi|\mathbf{X}_{s\leq t}, \mathrm{tr}(R_t) \leq \xi^4)] \\
&\leq \mathbb{P}(\mathrm{tr}(R_t) \geq \xi^4) + \xi^2.
\end{aligned} \tag{4.11}$$

By letting $L \to \infty$ then $\xi \to 0$, we have shown the claim. □

4.3. Convergence of posterior mean

In order to show the convergence of $\vec{m}_t$ towards $\mathbf{U}_t$ in $\mathbb{P}_{\vec{v}_{s\geq 0}}$, we need to do some computations. The intuition can be revealed by replacing $d\mathbf{X}_t$ in (2.10) by its formulation (2.7), so we rewrite (2.10) as

$$\mathrm{d}\vec{m}_t = [-\Gamma\vec{m}_t + \mathbf{F}_t + \sigma_x^{-2}R_t\mathbf{P}(t)(\mathbf{U}_t - \vec{m}_t)]\mathrm{d}t + \sigma_x^{-1}R_t P_X^*(\mathbf{X}_t)\mathrm{d}W_x(t). \tag{4.12}$$

This can be seen as a linear process with random perturbation. Its damping coefficient is $\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t)$, which by part (i) of Theorem 3.3 can be seen as $\Gamma + \sigma_x^{-2}LR_L\mathbf{M}$ plus some deviations. As L increases, $\Gamma + \sigma_x^{-2}LR_L\mathbf{M}$ is of order $\sqrt{L}$, which is also the order of force that will apply to $\vec{m}_t$ if it deviates from $\mathbf{U}_t$.

Proof of Theorem 3.3, part (iii). First fix a realization of $\vec{v}_{s\geq 0}$ such that part (i) holds. We simplify the notation of (4.12) by defining a new d-dim vector process w_t:

$$\mathrm{d}w_t := \frac{1}{\sqrt{L}} P_X^* \mathrm{d}W_X(t) = \frac{1}{\sqrt{L}} \begin{bmatrix} \vdots \\ \sum_{l=1}^L \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{X}_l(t)} (\vec{r}_{\mathbf{k}}^* \mathrm{d}W_l^x(t)) \\ \vdots \end{bmatrix}.$$

The quadratic covariation of w_t is

$$\langle \mathrm{d}w_t, \mathrm{d}w_t \rangle = \frac{1}{L} \left[\sum_{l=1}^L \mathrm{e}^{\mathrm{i}(\vec{j}-\vec{k})\cdot\vec{X}_l(t)} \vec{r}_{\mathbf{j}}^* \, \vec{r}_{\mathbf{k}} \right]_{\mathbf{j},\mathbf{k}} \mathrm{d}t = \frac{1}{L}\mathbf{P}(t)\mathrm{d}t. \tag{4.13}$$

Notice that $L^{-1}\mathbf{P}(t)$ has each matrix entry of norm at most 1,

$$\langle \mathrm{d}w_s^*, \mathrm{d}w_s^* \rangle \leq L^{-1}\mathrm{tr}(\mathbf{P}(t)) \leq |\mathbf{K}|.$$

Subtracting the equation for $\mathrm{d}\mathbf{U}_t$ in (2.3) from $\mathrm{d}\vec{m}_t$ in (4.12), we have the evolution of the difference $\vec{n}_t = \mathbf{U}_t - \vec{m}_t$:

$$\mathrm{d}\vec{n}_t = -[\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t)\vec{n}_t]\mathrm{d}t - \Sigma_v\mathrm{d}B_t + \sigma_x^{-1}\sqrt{L}R_t\mathrm{d}w_t.$$

Let $\Psi(s,t)$ be the fundamental matrix group generated by $D_t := -d_0 I - \sigma_x^{-2}R_t\mathbf{P}(t)$, that is if $\dot{x}_t = D_t x_t$ then $x_t = \Psi(s,t)x_s, \forall s \leq t$. Then by the Duhamel formula,

$$\vec{n}_t = \Psi(0,t)\vec{n}_0 - \int_0^t \Psi(s,t)\Sigma_v\mathrm{d}B_s + \sigma_x^{-1}\sqrt{L}\int_0^t \Psi(s,t)R_s\mathrm{d}w_s. \tag{4.14}$$

Recall that all of the derivation of the proof for part (i) holds as long as

$$q_L = L^{-1}\sup_{s\leq t}\|\mathbf{P}(t) - L\mathbf{M}\| \leq c_0,$$

where c_0 is any small fixed number. Define the stopping time $\tau := \inf\{s \geq 0, L^{-1}\|\tilde{P}_s\| \geq c_0\}$. Then the event $B := \{\tau \leq t\}$ indicates $q_L \leq c_0$. By (4.14) and the union bound,

$$\mathbb{P}_{\vec{v}_{s\geq 0}}(|\vec{n}_t| \geq 3\epsilon) \leq \mathbb{P}_{\vec{v}_{s\geq 0}}(B^c) + \mathbb{P}_{\vec{v}_{s\geq 0}}(|\Psi(0,t)\vec{n}_0| \geq \epsilon, B) + \mathbb{P}_{\vec{v}_{s\geq 0}}\left(\left|\int_0^t \Psi(s,t)\Sigma_v dB_s\right| \geq \epsilon, B\right)$$

$$+ \mathbb{P}_{\vec{v}_{s\geq 0}}\left(\left|\sigma_x^{-1}\sqrt{L}\int_0^t \Psi(s,t)R_s dw_s\right| \geq \epsilon, B\right). \tag{4.15}$$

In order to show $\vec{n}_t \to 0$ in $\mathbb{P}_{\vec{v}_{s\geq 0}}$, it suffices to show for any $\epsilon > 0$, each of the four probabilities on the right above goes to 0. Using Lemma 4.1, we know $\mathbb{P}_{\vec{v}_{s\geq 0}}(B^c) \to 0$ as $L \to \infty$.

In order to bound the remaining terms, let us bound $\|\Psi(s,t)\|$ when the event $\{\tau \geq t\}$ holds first. Let $x_t = \Psi(s,t)x_s$, then $\dot{x}_t = -(d_0 + \sigma_x^{-2}R_t\mathbf{P}(t))x_t$. Recall that based on the

derivation in the proof of part (i), (4.10) holds when $\{t \leq \tau\}$. Notice this implies the following hold for $s \leq t$ and some uniform constants D_1, C_1 and $\tilde{C}_1$:

$$\|\tilde{R}_s\| \leq D_1 \exp(-2\mu_L s) + c_0 C_1 \sqrt{L}^{-1}, \quad \|R_L\| \leq \tilde{C}_1 \sqrt{L}^{-1}, \quad \|\tilde{P}_s\| \leq c_0.$$

Hence there are uniform constants C_2, D_2 such that when $s \leq t \leq \tau$,

$$\frac{\mathrm{d}}{\mathrm{d}s}|x_s|^2 = \mathrm{Re}(2x_s^* \dot{x}_s) = -2x_s^*(\Gamma + \sigma_x^{-2} L R_L \mathbf{M} + \sigma_x^{-2} L R \mathrm{Re}(\tilde{R}_s \mathbf{M} + (R_L + \tilde{R}_s)\tilde{P}_s))x_s$$

$$\leq -2x_s^*(\Gamma + \sigma_x^{-2} L R_L \mathbf{M})x_s + 2\sigma_x^{-2} L \left[[D_1 \exp(-2\mu_L s) + c_0 C_1 \sqrt{L}^{-1}]\|\mathbf{M}\| \right.$$

$$\left. + c_0(\tilde{C}_1 \sqrt{L}^{-1} + c_0 C_1 \sqrt{L}^{-1} + D_1 \exp(-2\mu_L s)) \right] |x_s|^2$$

$$\leq -2x_s^*(\sigma_x^{-2} L R_L \mathbf{M})x_s + 2[c_0 C_2 \sqrt{L} + L D_2 \exp(-2\mu_L s)]|x_s|^2.$$

Since $R_L \mathbf{M}$ is a diagonal matrix with all real positive entries, the following holds for an $a > 0$:

$$\sigma_x^{-2} x_s^* L R_L \mathbf{M} x_s \geq |x_s|^2 \min_{\mathbf{k}} \frac{\sigma_x^{-2} L \sigma_{\mathbf{k}}^2 |\vec{r}_{\mathbf{k}}|^2}{d_{\mathbf{k}} + \sqrt{d_{\mathbf{k}}^2 + \sigma_x^{-2} L \sigma_{\mathbf{k}}^2 |\vec{r}_{\mathbf{k}}|^2}} \geq |x_s|^2 a \sqrt{L} \sigma_x |\vec{r}_{\mathbf{k}}|.$$

If we let $d_L = \frac{1}{2} a \sqrt{L} \sigma_x |\vec{r}_{\mathbf{k}}|$, choose a $c_0 \leq \frac{1}{4} C_2^{-1} a$, then we can simplify the inequality of $\frac{d}{ds}|x_s|^2$ above by another bound:

$$\frac{\mathrm{d}}{\mathrm{d}s}|x_s|^2 \leq -2(d_L - L D_2 \exp(-2\mu_L s))|x_s|^2. \tag{4.16}$$

This choice of c_0 is legitimate, since recall that in the first line of the proof of part (i), c_0 can be any fixed positive small number.

Using the Gronwall's inequality and $\mu_L \sim \sqrt{L}$, one has with some D_3

$$|x_t| \leq \exp\left(-d_L(t-s) + L D_2 \int_s^t \exp(-2\mu_L r) dr \right) |x_s|$$

$$\leq \exp(-d_L(t-s) + D_3 \sqrt{L} \exp(-2\mu_L s))|x_s|,$$

in other words up to $t \leq \tau$,

$$\|\Psi(s,t)\| \leq \exp(-d_L(t-s) + D_3 \sqrt{L} \exp(-2\mu_L s)) =: \psi(s,t).$$

Let s_0 be a constant larger than $\limsup 2D_3 \sqrt{L}/d_L < \infty$, then if $s > \frac{1}{2} s_0$,

$$d_L s \geq D_3 \sqrt{L}, \quad \sqrt{L} \exp(-2\mu_L s) \leq 1.$$

So when $t > s_0$, the second probability in (4.15) is bounded by an indicator

$$\mathbb{P}_{\vec{v}_{s\geq 0}}(\Psi(0,t)\vec{n}_0 \geq \epsilon, \tau \leq t) \leq 1_{\{\exp(-D_3\sqrt{L})\geq\epsilon\}},$$

which goes to 0 as $L \to \infty$. For the third probability in (4.15), observe that with integration by parts

$$\int_0^t \Psi(s,t)\Sigma_v \mathrm{d}(B_s - B_t) = \Psi(0,t)\Sigma_v B_t + \int_0^t \Psi(s,t)D_s\Sigma_v[B_s - B_t]\mathrm{d}s.$$

It suffices for us to show $\int_0^t \psi(s,t)\|D_s\|\|B_s - B_t\|\mathrm{d}s \to 0$ when $\tau \le t$. This computational verification will be done in Lemma Appendix A.6, part (i) using

$$\begin{aligned}
\|D_s\| &= \|\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t)\| \\
&\le \|\Gamma\| + \sigma_x^{-2}\|R_L\|\|L\mathbf{M} + \tilde{P}_t\| \\
&\le \|\Gamma\| + \sigma_x^{-2}(\|\tilde{R}_s\| + \|R_L\|)(\|L\mathbf{M}\| + c_0) \\
&\le AL\exp(-\mu_L s) + \sqrt{L}B
\end{aligned}$$

for some constants A, B.

To bound the last probability in (4.15), we define $y_t := \sigma_x^{-1}\int_0^t \Psi(s,t)\sqrt{L}R_s\mathrm{d}w_s$. Due to the Chebyshev's inequality, it suffices for us to show $\mathbb{E}_{\vec{v}_{s\ge0}}(|y_t|^2 1_{\tau \le t}) \to 0$. Notice that y_t follows:

$$\mathrm{d}y_t = -(\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t))y_t\mathrm{d}t + \sigma_x^{-1}\sqrt{L}R_t\mathrm{d}w_t.$$

Following the derivation of (4.16), we have the same bound for $\mathrm{Re}(y_t^*(\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t))y_t)$ when $\tau \ge t$; apply it and the bound for the quadratic variation in (4.13) in the following application of Itô formula:

$$\begin{aligned}
\mathrm{d}|y_t|^2 &= -2\mathrm{Re}[y_t^*(\Gamma + \sigma_x^{-2}R_t\mathbf{P}(t))y_t\mathrm{d}t + R_t\langle \mathrm{d}w_t^*, \mathrm{d}w_t^*\rangle R_t + 2y_t^*R_t\mathrm{d}w_t] \\
&\le -2[d_L - LD_3\exp(-2\mu_L t)]|y_t|^2\mathrm{d}t + |\mathbf{K}|\|R_t\|^2\mathrm{d}t + 2\mathrm{Re}(y_t^*R_t\mathrm{d}w_t).
\end{aligned}$$

If we let z_t be the solution with $z_0 = 0$ and

$$\mathrm{d}z_t = -2[d_L - LD_3\exp(-2\mu_L t)]z_t\mathrm{d}t + |\mathbf{K}|\|R_t\|^2\mathrm{d}t + 2\mathrm{Re}(y_t^*R_t\mathrm{d}w_t).$$

This is well defined as one can easily verify that y_s is a L^2-integrable process, and by Lemma Appendix A.4 R_t is uniformly bounded from above. Then by the comparison principle, $z_t \ge |y_s|^2 1_{\{t \le \tau\}}$, $\mathbb{P}_{\vec{v}_{s\ge0}}$-a.s. Moreover, notice that w_t is a martingale, using

$$z_t = \int_0^t \Psi(s,t)|\mathbf{K}|\|R_s\|^2\mathrm{d}s + 2\int_0^t \Psi(s,t)\mathrm{Re}(y_s^*R_s\mathrm{d}w_s),$$

we have:

$$\mathbb{E}_{\vec{v}_{s\ge0}}(|y_t|^2 1_{\tau \ge t}) \le \mathbb{E}z_t \le |\mathbf{K}|\int_0^t \psi(s,t)\|R_s\|^2\mathrm{d}s.$$

It remains to show the quantity above converges to 0; this is rather elementary since $\|R_s\|^2 = \|R_L + \tilde{R}_s\|^2 \le B$, then apply Lemma Appendix A.6 (ii) to verify this. $\quad\square$

4.4. Asymptotic uncertainty reduction

Proof of Corollary 3.4. With the clear knowledge of the asymptotic behavior R_t and m_t, the information theory part is quite elementary.

Part (i) Recall that in (2.13), the dispersion part is

$$-\frac{1}{2}\ln(\det R_t) + \frac{1}{2}\ln(\det R_{att}) - \frac{|\mathbf{K}|}{2} + \frac{1}{2}\mathrm{tr}(R_t R_{att}^{-1}).$$

Since the climatological covariance R_{att} is a constant matrix, $\|R_t\| \to 0$ in $\mathbb{P}_{\vec{v}_{s\geq 0}}$ by Theorem 3.3 (i), so asymptotically the trace part vanishes. Apply the result of Theorem 3.3 (i) on the following decomposition:

$$\ln\det(R_t) = \ln\det(I + R_L^{-1}(R_t - R_L)) + \ln\det(R_L)$$

the first term goes to 0 in $\mathbb{P}_{\vec{v}_{s\geq 0}}$. Hence the dispersion part asymptotically converges to

$$-\frac{1}{2}\ln(\det R_L) + \frac{1}{2}\ln(\det R_{att}) - \frac{|\mathbf{K}|}{2} = \frac{1}{2}\sum_{\mathbf{k}\in\mathbf{K}}\left(\ln\left(\tfrac{1}{2}(1 + \sqrt{1 + L\sigma_x^{-2}d_{\mathbf{k}}^{-2}\sigma_{\mathbf{k}}^2|\vec{r}_{\mathbf{k}}|^2})\right) - 1\right)$$

Part (ii) The signal part of the relative entropy is $\frac{1}{2}(\vec{m}_t - \vec{m}_t^{att})R_{att}^{-1}(\vec{m}_t - \vec{m}_t^{att})$, by our results in (ii) and (iii), $\vec{m}_t - \mathbf{U}_t$ converges to 0 in two different manners. Hence it is straight forward to have our claim.

Part (iii) Since the mutual information is

$$I(\mathbf{U}_t, \mathbf{X}_{s\leq t}) = \frac{1}{2}\ln\det R_{att} - \frac{1}{2}\mathbb{E}\ln\det R_t.$$

Again we decompose $\ln\det R_t = \ln\det(R_L^{-1}R_t) + \ln\det(R_L)$. On one hand by Theorem 3.3 part (i), $|\ln\det(R_L^{-1}R_t)| \to 0$ in $\mathbb{P}$, and notice that according to Lemma Appendix A.4, there is some constant U so that $|\det(R_L^{-1}R_t)| \leq U\ln L$ a.s. Then for any fixed $\xi > 0$,

$$\mathbb{E}|\ln\det(R_L^{-1}R_t)| \leq \xi + \mathbb{P}(|\ln\det(R_L^{-1}R_t)| \geq \xi)U\ln L$$

The right hand side is of order less than $\ln L$. On the other hand,

$$\frac{-\frac{1}{2}\ln\det R_L}{\frac{|\mathbf{K}|}{4}\ln L} \xrightarrow{L\to\infty} 1.$$

Putting the two parts together we see our claim holds. □

Proof of Corollary 3.5. For marginal processes, it suffices to redo the previous derivation but using sub-matrices. Let us add tilde to symbols to indicate sub-vectors of indexes in $\widetilde{\mathbf{K}}$ or sub-matrices of indexes in $\widetilde{\mathbf{K}} \times \widetilde{\mathbf{K}}$. We shift the indexes so the ones in $\widetilde{\mathbf{K}}$ come ahead. As the mean and covariance of $\widetilde{\mathbf{U}}_t$ are these sub-vectors and sub-matrices, the relative entropy is

$$\mathcal{P}(\pi_{t|t}|_{\widetilde{\mathbf{K}}}, \pi_t^{att}|_{\widetilde{\mathbf{K}}}) = \tfrac{1}{2}\left[(\tilde{m}_t - \tilde{m}_t^{att})^*\widetilde{R}_{att}^{-1}(\tilde{m}_t - \tilde{m}_t^{att})\right] \qquad \text{Signal}$$
$$+ \tfrac{1}{2}\left[\mathrm{tr}(\tilde{R}_t\tilde{R}_{att}^{-1}) - |\widetilde{\mathbf{K}}| - \ln\det(\tilde{R}_t\tilde{R}_{att}^{-1})\right] \qquad \text{Dispersion.}$$

The mutual information is

$$\mathcal{I}(\widetilde{\mathbf{U}}_t, \mathbf{X}_{s \leq t}) = \frac{1}{2} \ln \det \widetilde{R}_{att} - \frac{1}{2} \mathbb{E} \ln \det \widetilde{R}_t.$$

The results of Theorem 3.3 indicates their marginal version hold as well, because evidently

$$\|R_t - R_L\| \geq \|\widetilde{R}_t - \widetilde{R}_L\|;$$

as R_{att} is diagonal:

$$(\tilde{m}_t - \tilde{m}_t^{att})^* \widetilde{R}_{att}^{-1}(\tilde{m}_t - \tilde{m}_t^{att}) = \begin{bmatrix} \tilde{m}_t - \tilde{m}_t^{att} \\ 0 \end{bmatrix}^* R_{att}^{-1} \begin{bmatrix} \tilde{m}_t - \tilde{m}_t^{att} \\ 0 \end{bmatrix} \leq (\tilde{m}_t - \tilde{m}_t^{att})^* \widetilde{R}_{att}^{-1}(\tilde{m}_t - \tilde{m}_t^{att});$$

if $\widetilde{R}_t$ has uni-norm eigenvector $\tilde{u}$ for its least eigenvalue, then

$$\|\widetilde{R}_t^{-1}\| = |\widetilde{R}_t \tilde{u}|^{-1} = \left| R_t \begin{bmatrix} \tilde{u} \\ 0 \end{bmatrix} \right|^{-1} \leq [\inf_{|u| \leq 1} |R_t u|]^{-1} = \|R_t^{-1}\|.$$

Using these marginal version of results, the proof of this corollary is identical to the one of Corollary 3.4. □

5. Numerical experiments

In this section we numerically validate the theoretical results in Section 3. We consider a $d = 2$ dimensional set-up, like (2.1) with wavenumbers in $[-2, 2]^2$, the total dimension of the signal process $\mathbf{U}_s$ is $|\mathbf{K}| = 26$.

For system parameters in (1.2), all the damping $d_{\mathbf{k}}$ are set as 0.05, which corresponds to a moderately long decorrelation time $\tau_{corr} = 20$ units as occurs in many geophysical flows. The system is assumed to have no deterministic forcing for simplicity and therefore the climatological mean is zero. Equipartition energy is assumed with all $\sigma_{\mathbf{k}}^v = 0.3$ to mimic turbulence at large scale [15].

The observation noise of the tracers is set to be $\sigma_x = 0.5$ as the observation noise level is roughly one-half of the system noise of each mode. In the following, we study the filtering skill with different number of tracers L, ranging from 2 to 500. One standard measurement of filtering skill is the root-mean-squared (RMS) error between the true signal $\vec{u}_t$ and maximum likely hood filter estimate $\vec{m}_t$,

$$\text{RMSE} = \sqrt{\frac{1}{t} \int_0^t |\vec{m}_s - \mathbf{U}_s|^2 \mathrm{d}s}. \tag{5.1}$$

Comparison of the streamlines in physical space is illustrated in Figure 5.1. The top row displays streamlines associated with the true velocity field at $T = 5, 15$ and 25, followed by the three rows showing the recovered flow field using $L = 2, 10$ and 50 tracers. In all the

panels, the solid and dashed curves represent the counterclockwise and clockwise motion of the flows, respectively. Clearly, the flow field with only 2 tracers succeeds in recovering the general profile of the flow field. Yet, a low recovering skill is observed in certain areas, such as the flipped sign in left top part at $T = 5$. As expected, the recovered flow field approaches the true one with the increase of L. However, despite the significant error reduction in the recovered flow with $L = 50$, some visible difference still exists. Even with L increasing to 500 as shown in Figure 5.1, the recovered streamlines are not perfectly matching the truth, which implies the potential information barriers.

The intuition in Figure 5.1 and 5.2 motivates us to look at the error in the mean state of the estimated flow field. Recall the second and third part of Theorem 3.3, we illustrated the difference between posterior mean and the true signal $\vec{m}_t - \mathbf{U}_t$ converges to 0, which is validated by Figure 5.3 and 5.4 for the real part of mode $[1, 1]$. Yet, the decay rate of the error becomes slower with increasing L. It is worth mentioning that this difference is much smaller than the standard deviation of the climatological distribution even filtering with only 2 tracers, implying that filtering leads to a significant improvement. Furthermore, the RMS error in the filtered velocity field with 10 tracers is only less than 20% of the standard deviation at climatology.

The information barriers are well reflected in the uncertainty reduction, which is included in Figures 5.5 and 5.6. The reduced uncertainty in the signal part converges to its limit value as L increases. On the other hand, although the uncertainty reduction in the dispersion part has a sustained growth, the growth is only at rate $\frac{1}{4}|\mathbf{K}|\ln L$. With $|\mathbf{K}| = 26$, reducing every 15 nats of uncertainty from the dispersion part requires a tenfold increase of tracers. This is illustrated in panel (i) of Figure 5.6. This uncertainty reduction approaches the asymptotic value (3.3) from Corollary 3.4, as indicated by the overlapped solid and dashed lines. In addition, the length of the transient period for the dispersion part to arrive at the equilibrium value goes down with the increase of L.

Finally, Figure 5.7 shows the time evolution of posterior covariance R_t with different L. Based on panels (a)–(h), both the covariance of individual modes and the norm of the full covariance go to zero with increased tracers employed in the flow. As L increases, the posterior covariance R_t converges to R_L as defined in (3.1). R_L scales as $L^{-1/2}$ asymptotically, this is shown by panel (i). Besides, the rescaled deviation,

$$\|R_L^{-1}\|\|R_t - R_L\| \tag{5.2}$$

shown in Figure 5.8, goes to zero with L as expected. These two figures are consistent with Theorem 3.3, part (i).
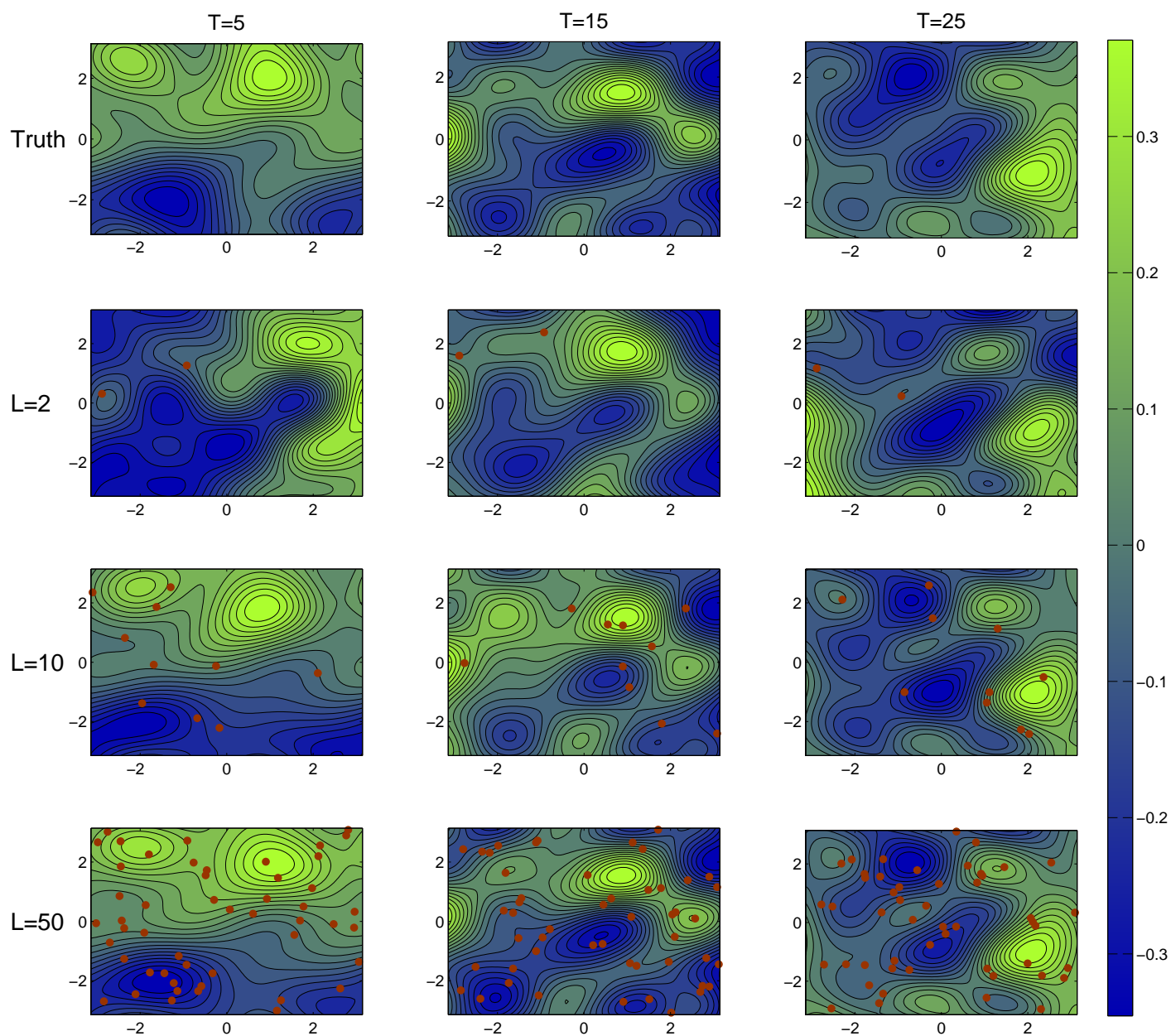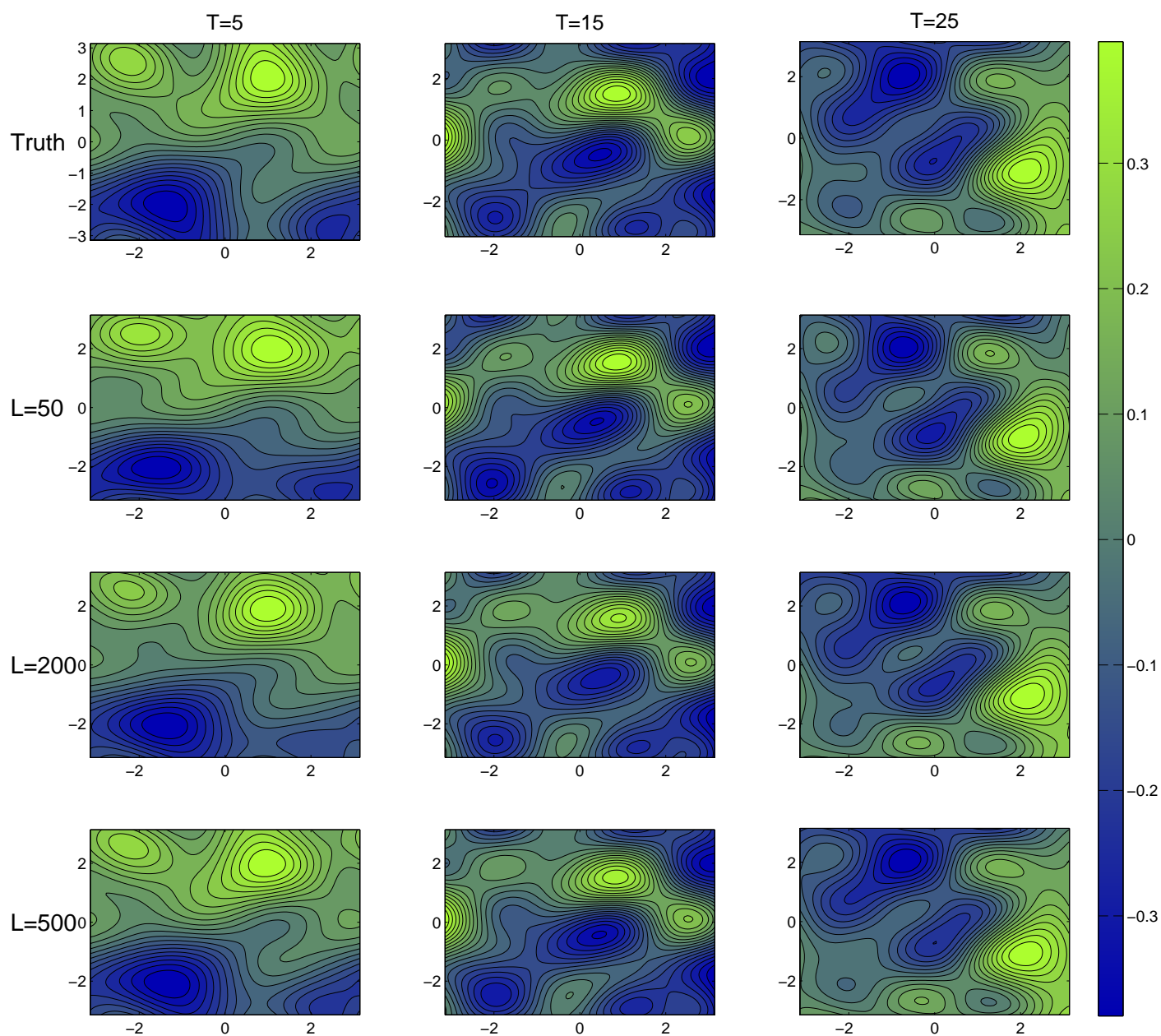
Figure 5.1: Streamlines recovered with different number of tracers L. The first row shows the streamlines of the true realization. The second, third and fourth rows show the recovered streamlines using $L = 2$, 10 and 50 tracers, respectively. The three columns show the comparison of streamlines at time $T = 5, 15$ and 25. Streamlines with positive and negative values representing the counterclockwise and clockwise direction of the velocity field. The circles represent the location of the tracers at current time.

Figure 5.2: Streamlines recovered with different number of tracers L. The first row shows the streamlines of the true realization. The second, third and fourth rows show the recovered streamlines using $L = 50$, 200 and 500 tracers, respectively. The three columns show the comparison of streamlines at time $T = 5, 15$ and 25. Streamlines with positive and negative values representing the counterclockwise and clockwise direction of the velocity field.
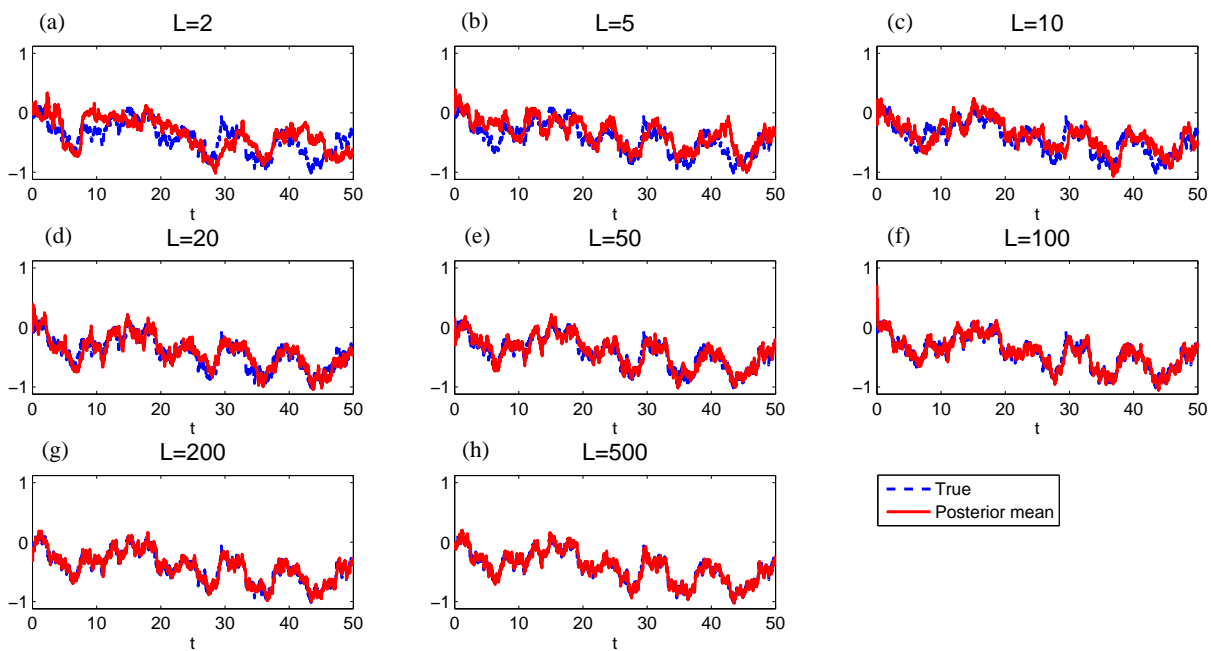
Figure 5.3: Posterior mean (solid line) for the real part of mode $[1, 1]$ as a function of time with different number of tracers L, compared with the true realization (dashed line).
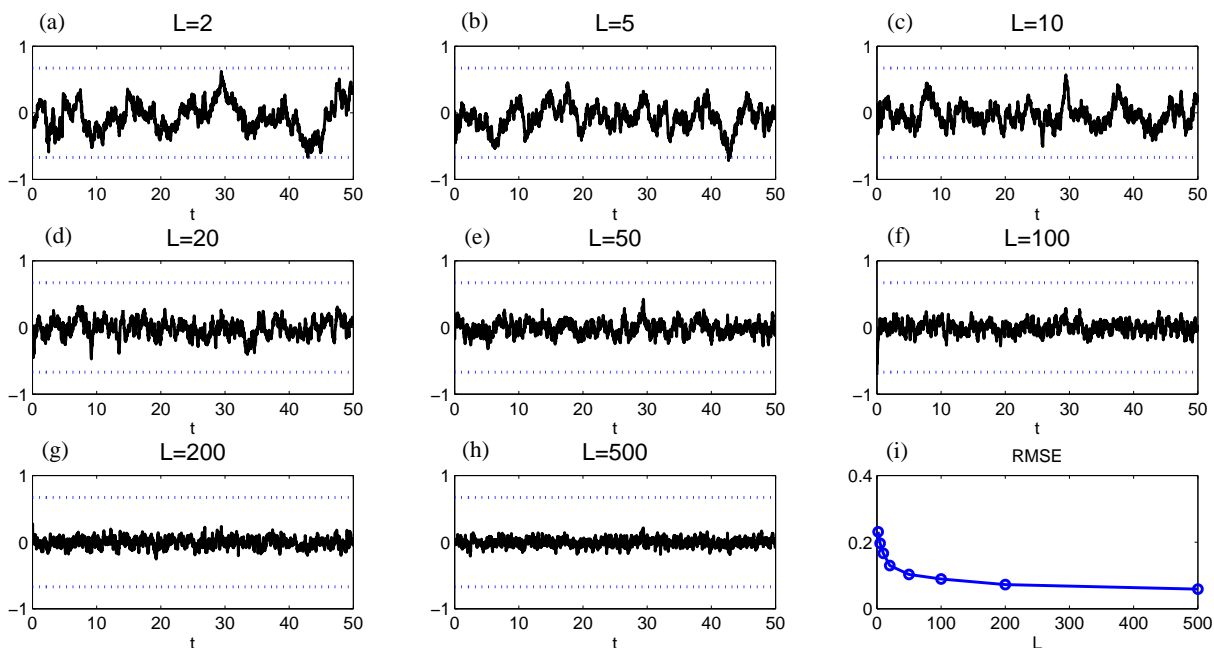
Figure 5.4: Error between posterior mean and the true realization for the real part of mode $[1, 1]$. Panels (a)–(h) show the error as a function of time for each fixed L. The two dot lines indicate the standard deviation of the unfiltered climatological distribution. Panel (i) shows the RMS error across time interval $T \in [0, 50]$ as a function of L.
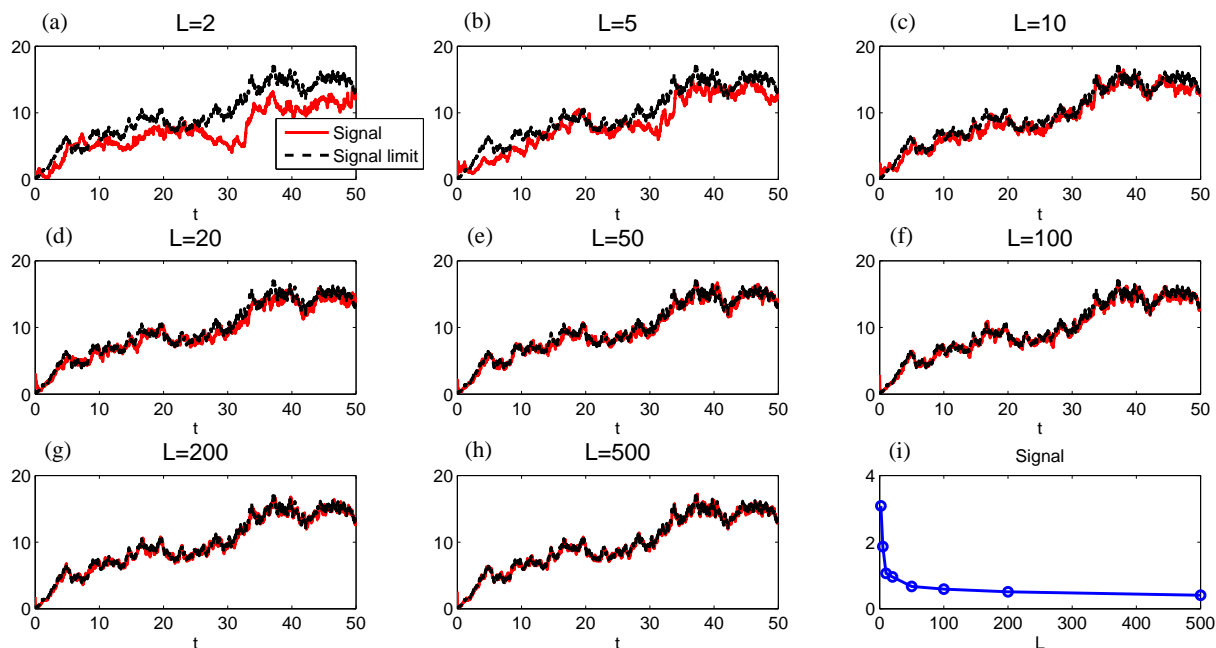
Figure 5.5: Uncertainty reduction in the signal part with different number of tracers L. Panels (a)–(h) show the uncertainty reduction as a function of time (thick solid line) compared with its limit (thick dashed line) for each fixed L. In panel (i), the solid line with circles shows the averaged value of the absolute error in the signal part compared with its limit across time interval $T \in [0, 50]$ as a function of L.
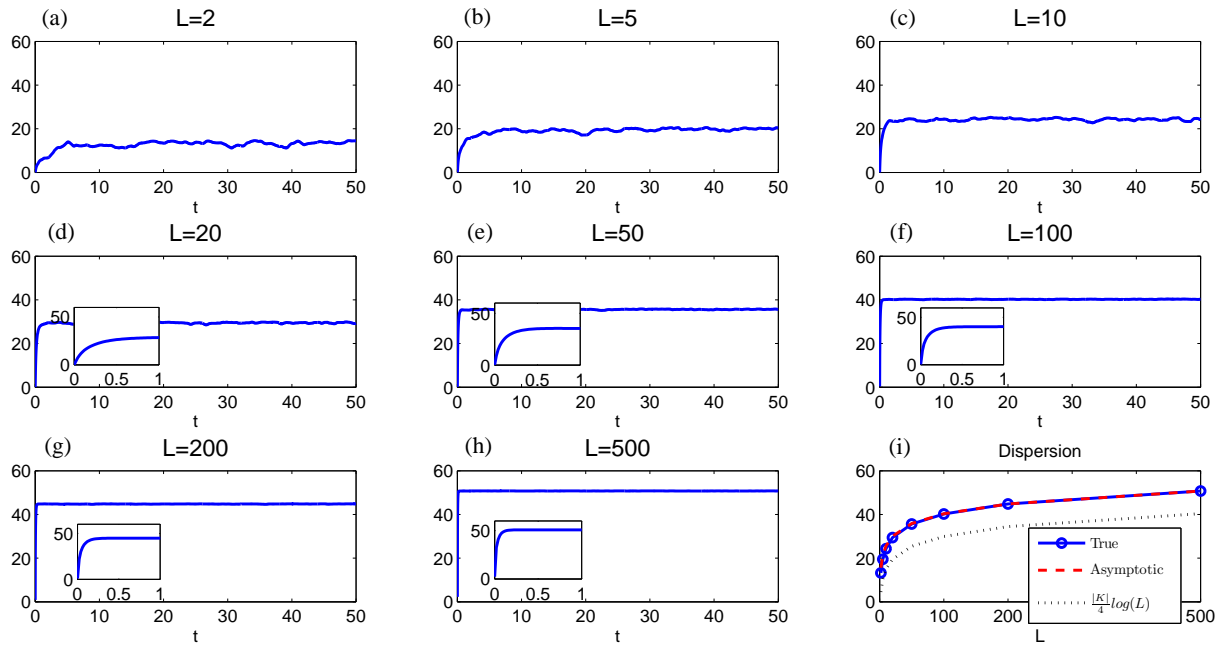
Figure 5.6: Uncertainty reduction in the dispersion part with different number of tracers L. Panels (a)–(h) show the uncertainty reduction as a function of time in the dispersion part for each fixed L. In panel (i), the solid line with circles shows the averaged dispersion part across time interval $T \in [5, 50]$ as a function of L and the dotted line represents the function $\frac{1}{4}|\mathbf{K}| \ln L$. The magnifying blocks show the evolution in the initial one time unit.
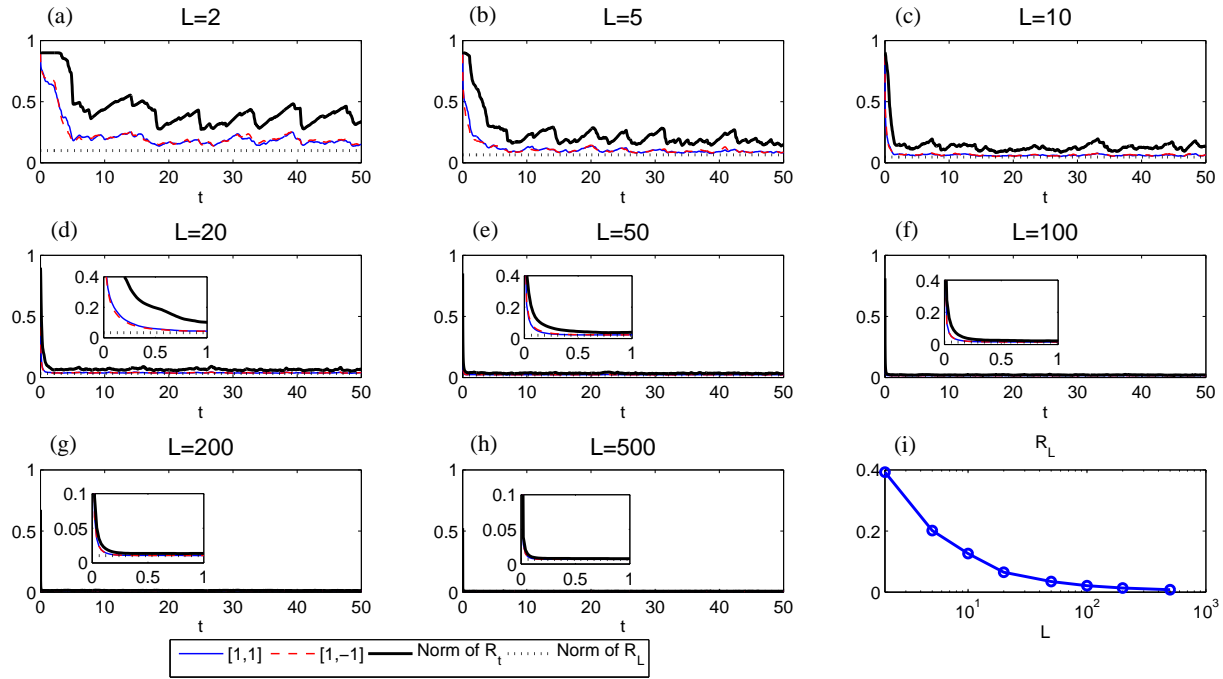
Figure 5.7: Posterior covariance with different number of tracers L. Panels (a)–(h) show the posterior covariance of mode $[1, 1]$ (thin solid line), $[1, -1]$ (thin dashed line) and the norm of the full posterior covariance matrix (thick solid line) as a function of time for each fixed L. The asymptotical limit value R_L is plotted as reference in each panel with dotted line. In panel (i), the solid line with circles shows the averaged posterior covariance's norm across time interval $T \in [5, 50]$ as a function of L.
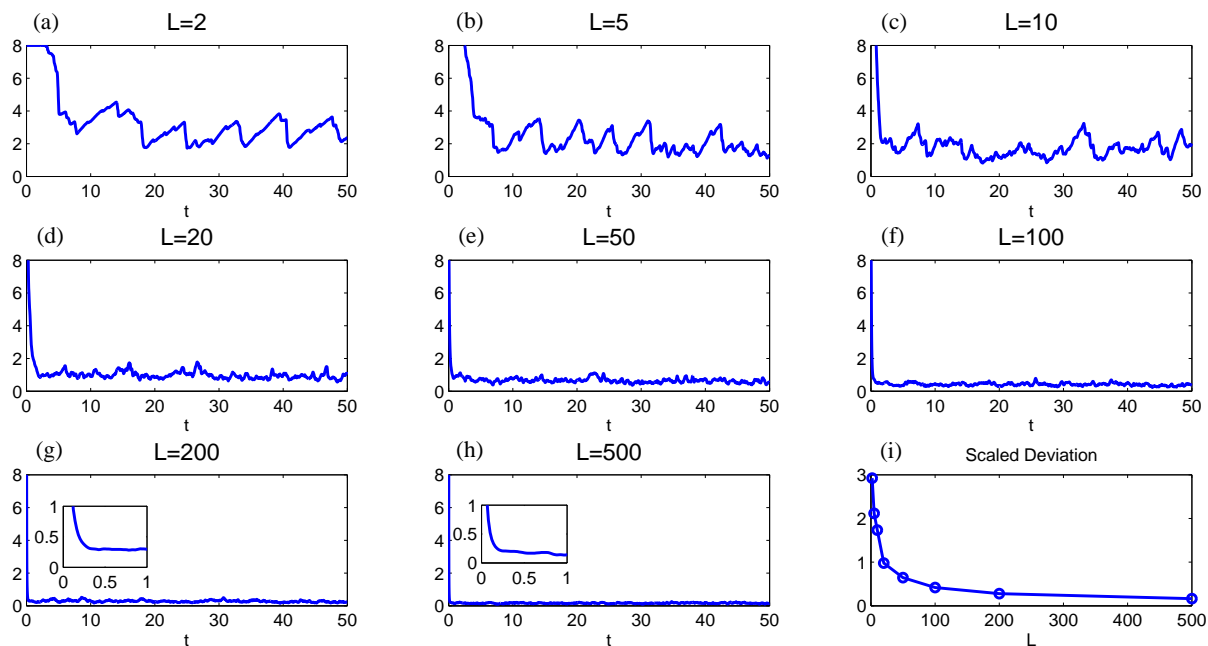
Figure 5.8: Rescaled deviation with different number of tracers L. Panels (a)–(h) show the rescaled deviation (5.2) as a function of time for each fixed L. Panel (i) shows the averaged rescaled deviation across time interval $T \in [5, 50]$ as a function of L. The magnifying blocks show the evolution in the initial one time unit.

6. Conclusion and discussion

In this paper, we set-up an optimal filter of a random incompressible flow using observations from Lagrangian tracers. The derivation of this filter is based on the conditional Gaussian theory in [16]. Despite the nonlinearity of the tracers, the posterior distribution is Gaussian with mean and covariance following elementary differential equations with random coefficients. This may bring insight in building other simple filters for nonlinear systems [15].

We prove the geometric ergodicity of the tracers' projection on $\mathbb{T}^d$. We then analyze the asymptotical behavior of the filter assuming each tracer is distributed uniformly on $\mathbb{T}^d$ when $L \to \infty$. We show the posterior covariance, will approach R_L a deterministic diagonal matrix and the posterior mean will converge to the signal process.

With clear asymptotic characterization of the filter, we can characterize the uncertainty reduction by the observations from L tracers as $L \to \infty$. In terms of both relative entropy and mutual information, the uncertainty reduction grows asymptotically as $\frac{1}{4}|\mathbf{K}|\ln L$. In other words, in order to gain the same amount of additional information, one would need exponentially more tracers. This certainly is an information barrier in practice when only a reasonable amount of tracers are available. As seen in the numerical experiments, Section 5, 50 tracers produce roughly the same result as with 500 tracers. It is worth noticing that [6] has a similar claim, as their result indicates that 36 tracers without any analysis, instead of 64, is optimal under their setting.

Apart from the direct conclusions listed above, we also have the following remarks on our results:

(i) The requirement that the flow is incompressible is not necessary for the derivation of the filter. Indeed, incompressibility does not show up in the filter set-up part, Section 2.2; it is only required in Theorem 3.1 to make the uniform distribution invariant. It is possible to generalize the ideas in this paper to flows that are not incompressible, e.g. a shallow water model. This lies beyond the scope of this paper, and will be addressed by a following up paper of the authors.

(ii) It is worth mentioning that all theoretical results, Theorem 3.1, 3.3 and Corollary 3.4, are stated in the conditional distribution $\mathbb{P}_{\vec{v}_{s \leq t}}$. Generally speaking, proving results in $\mathbb{P}_{\vec{v}_{s \geq 0}}$ for a.s. realization of $\vec{v}_{s \geq 0}$ is more difficult than proving the same results in the super ensemble probability $\mathbb{P}$. However, in actual practical applications, often one is only observing noisy tracers in a single realization of a random flow field, so we emphasize this here. Such a.s. results in the random flow field also allow us to study the filter behavior for rare extreme realization of the flow field.

(iii) As discussed in Section 3, the configuration of the tracers' locations impacts the performance of the filter. Non-ideal situations, e.g. clustering all tracers initially, are ruled out by the Assumption 3.2. As shown in [5, 7, 8], different starting locations of tracers and their trajectories result in very different filter results at least in short period

of time. However, as second part of this paper addresses mainly the asymptotic behavior of the tracers, it is preferable to work under Assumption 3.2.

(iv) Although Theorem 3.1 indicates the distribution of tracers will converge to the equilibrium in a finite amount of time, the length of this transient regime could be decades if all the tracers are disposed initially at the same location and σ_x is insignificant compared to the scale of $\vec{v}$. In practice, in order for filter to behave as Theorem 3.3 describes, the initial location of tracers should be chosen uniformly in the area.

(v) Theorem 3.3 part (i) indicates the posterior covariance is close to a deterministic matrix R_L when L is large. This fact would be used to deploy an efficient imperfect filter with great increase of efficiency: we can simply set R_t to be R_L after a period of time. This resolves a large computational burden when $|\mathbf{K}|$ and L are large, because updating the Riccati equation (2.9) is very stiff while R_L is diagonal. Similar efficient cheap diagonal filters have been developed in the context of turbulent dynamical systems [15].

Acknowledgments

Appendix

Appendix A.1. A detailed formulation of the random incompressible flow model

We consider a random periodic incompressible d-dim flow as a superposition of random plane waves, formulated by (1.1):

$$\vec{v}(\vec{x}, s) = \sum_{\mathbf{k} \in \mathbf{K}} \hat{v}_{\mathbf{k}}(s) \exp(i\vec{k} \cdot \vec{x})\vec{r}_{\mathbf{k}},$$

where $\mathbf{K}$ is some finite set and $\mathbf{k}$ associates with a wavenumber $\vec{k}$. The incompressible condition for a constant density flow gives $\nabla \cdot \vec{v} = 0$, which implies $\vec{k} \cdot \exp(i\vec{k} \cdot \vec{x})\vec{r}_{\mathbf{k}} = 0$ for each Fourier mode $\vec{k}$. Therefore, the eigenvector $\vec{r}_{\mathbf{k}}$ is perpendicular to the wavenumber $\vec{k}$. There can be multiple $\mathbf{k}$s sharing the same wavenumber $\vec{k}$, we require that their $\vec{r}_{\mathbf{k}}$ to be orthogonal to each other so the contribution of each $\vec{v}_{\mathbf{k}}$ to $\vec{v}_s$ is identifiable. We model $\{\hat{v}_{\mathbf{k}}(s)\}_{\mathbf{k} \in \mathbf{K}}$ as a group of independent O.U. processes with damping $d_{\mathbf{k}} > 0$:

$$d\hat{v}_{\mathbf{k}}(s) = -d_{\mathbf{k}}\hat{v}_{\mathbf{k}}(s)ds + f_{\mathbf{k}}(s)ds + \sigma_{\mathbf{k}}dW^v_{\mathbf{k}}(s).$$

When the wavenumbers $\vec{k} = \vec{0}$, the corresponding modes represent a random background mean sweep, with the corresponding $\hat{v}_{\mathbf{k}}(s), \vec{r}_{\mathbf{k}}$ being real-valued.

For nonzero wavenumbers, in order to keep $\vec{v}(\vec{x}, s)$ being real-valued, we assume the components of (1.1) come with their conjugate pairs. We denote $-\mathbf{k}$ as the conjugate index

of $\mathbf{k}$ with corresponding wavenumber being $-\vec{k}$, $\hat{r}_{-\mathbf{k}}$, $\hat{v}_{-\mathbf{k}}(s)$, $f_{-\mathbf{k}}(s)$, $W^v_{-\mathbf{k}}(s)$ being conjugates for the ones of index $\mathbf{k}$ and $d_{\mathbf{k}} = d_{-\mathbf{k}}$, $\sigma_{\mathbf{k}} = \sigma_{-\mathbf{k}}$. One way to ensure this is formulating the complex Wiener processes as follow:

$$W^v_{\mathbf{k}}(s) = \frac{1}{\sqrt{2}}(B_{\mathbf{k}}(s) + iB_{-\mathbf{k}}(s)), \quad W^v_{-\mathbf{k}}(s) = \frac{1}{\sqrt{2}}(B_{\mathbf{k}}(s) - iB_{-\mathbf{k}}(s)),$$

where $\{B_{\mathbf{k}}(s)\}$ are independent 1-dim real-valued Wiener processes. It is usually convenient to discuss these modes in terms of conjugate pairs. This is the reason why the matrix Σ_u in (2.2) holds a particular form as illustrated after (2.2).

Filtering formula for conditional Gaussian process

The following is a simplified version of Theorem 12.7 [16] in our context:

Theorem Appendix A.1. *Let (u_t, x_t) be a continuous diffusion type process with*

$$\mathrm{d}u_t = [-\Gamma_t u_t + \mathbf{F}_t]\mathrm{d}t + \Sigma_u \mathrm{d}W_1(t),$$

$$\mathrm{d}x_t = [P_t u_t + Q_t]\mathrm{d}t + \Sigma_x \mathrm{d}W_2(t),$$

with bounded Γ_t, F_t, P_t, Q_t processes being functions of x_t. If $\mathbb{P}(u_0 \in \cdot | x_0)$ is $\mathcal{N}(m_0, R_0)$, then conditioned on $x_{s \leq t}$, $\mathbb{P}(u_t \in \cdot | x_{s \leq t})$ is $\mathcal{N}(m_t, R_t)$, with m_t, R_t being solutions to the following with initial value m_0, R_0:

$$\mathrm{d}m_t = [-\Gamma_t m_t + F_t]\mathrm{d}t + R_t P_t^*(\Sigma_x \Sigma_x^*)^{-1}[\mathrm{d}x_t - (P_t m_t + Q_t)\mathrm{d}t],$$

$$\mathrm{d}R_t = [-\Gamma_t R_t - R_t \Gamma_t^* + \Sigma_u \Sigma_u^* - R_t P_t^*(\Sigma_x \Sigma_x^*)^{-1} P_t R_t]\mathrm{d}t.$$

Appendix A.2. Information theory formulae for complex Gaussian random vectors

We say a complex-valued random vector $X + Y\mathrm{i}$ is Gaussian distributed, if (X, Y) as a real-valued vector is Gaussian distributed. The relative entropy of $X + Y\mathrm{i}$ between probability density p and q can be naturally defined as relative entropy of (X, Y) under these two probabilities.

One caveat in dealing with complex random vector $X + Y\mathrm{i}$ is its covariance matrix does not contain sufficient second moments to induce the density of (X, Y). For example X and $\frac{1}{\sqrt{2}}(X + Y\mathrm{i})$ have the same mean and covariance if X, Y are independent $\mathcal{N}(0, 1)$ random variables. Of course one can look for the covariance matrix for the joint vector (X, Y), but in many cases that will be very inconvenient. For instance in our filtering model, we will need additional formulas other than (2.10) and (2.9).

The common practice in engineering community, e.g. [32], to overcome this setback is consider the pair $(X + Y\mathrm{i}, X - Y\mathrm{i})$, since one can show its covariance matrix is sufficient to induce the law of (X, Y). Moreover, if one has the $(X + Y\mathrm{i}, X - Y\mathrm{i}) \sim \mathcal{N}(\vec{m}_p, R_p)$ and $\mathcal{N}(\vec{m}_q, R_q)$, the relative entropy of (X, Y) under two measures can also be computed by (2.12). As the authors cannot find a reference clearly stating this fact, we prove it using elementary algebra.

Lemma Appendix A.2. *Suppose under measures $\mathbb{P}$ and $\mathbb{Q}$, X, Y, Z are real-valued joint Gaussian vectors with Y and Z being of the same dimension, suppose also that $[X, Y + Zi, Y - Zi]'$ has mean $\vec{m}_p, \vec{m}_q$ and covariance R_p, R_q under each measure respectively, then the relative entropy between $\mathbb{P}$ and $\mathbb{Q}$ of $[X, Y, Z]'$ is given by*

$$\mathcal{P}(\mathbb{P}, \mathbb{Q}) = \left[\tfrac{1}{2}(\vec{m}_p - \vec{m}_q)^T R_q^{-1}(\vec{m}_p - \vec{m}_q) \right] + \tfrac{1}{2}\left[tr(R_p R_q^{-1}) - N - \ln det(R_p R_q^{-1}) \right].$$

where $N = \dim X + \dim Y + \dim Z$.

Proof. Simply notice that

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & \frac{1}{2}I & \frac{1}{2}I \\ 0 & -\frac{i}{2}I & \frac{i}{2}I \end{bmatrix} \begin{bmatrix} X \\ Y + Zi \\ Y - Zi \end{bmatrix}.$$

Denote M as the matrix on the right hand side. Then $[X, Y, Z]'$ are distributed as $\mathcal{N}(M\vec{m}_p, MR_pM^*), \mathcal{N}(M\vec{m}_q, MR_qM^*)$ respectively under $\mathbb{P}, \mathbb{Q}$. So applying the formula for the real-valued Gaussian vector $[X, Y, Z]'$ and cancel out the appearance of M, one can show our claim elementarily. □

As application to our filtering purpose, X consists of the real modes. Pick one complex mode out of each conjugate pairs, and use their real parts to form Y and imaginary part to form Z. The conjugate counterpart of these modes are naturally represented by $Y - Zi$.

This is the first reason why we prefer to talk of the complex modes in conjugate pairs like the statement of Corollary 3.4: their joint mean and covariance is necessary and sufficient to induce the law of them. On the other hand, in the sense of uncertainty reduction, the information regarding $\hat{v}_{\mathbf{k}}$, or $\hat{v}_{-\mathbf{k}}$, or both of them as a pair, are the same: $\hat{v}_{\mathbf{k}}$'s value determine $\hat{v}_{-\mathbf{k}}$ and vice versa. So it is more reasonable to think of complex modes in pairs.

Lemma Appendix A.3. *Under the set-up of Section 2, the mutual information is given by*

$$\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s\leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_{att}))$$
$$= \frac{1}{2}\ln \det R_{att} - \frac{1}{2}\mathbb{E}\ln \det R_t$$

Proof. The first line come from an equivalent definition of mutual information, cf. [23] and [33] equation (3.2). That is if we write the law of X conditioned on $Y = y$ as $p_{X|Y=y}$, then by disintegration $p(\mathrm{d}x, \mathrm{d}y) = p_{X|Y=y}(\mathrm{d}x)p_Y(\mathrm{d}y)$, hence

$$\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s\leq t}) = \int p(\mathrm{d}x, \mathrm{d}y) \ln \frac{p_{X|Y=y}(\mathrm{d}x)p_Y(\mathrm{d}y)}{p_X(\mathrm{d}x)p_Y(\mathrm{d}y)}$$
$$= \int p_Y(\mathrm{d}y) \int p_{X|Y=y}(\mathrm{d}x) \ln \frac{p_{X|Y=y}(\mathrm{d}x)}{p_X(\mathrm{d}x)}$$
$$= \int p_Y(\mathrm{d}y)\mathcal{P}(p_{X|Y=y}, p_X).$$

Let $X = \mathbf{U}_t$, $Y = \mathbf{X}_{s \leq t}$, notice $\pi_{t|t}$ is the law of $\mathbf{U}_t$ conditioned on $\mathbf{X}_{s \leq t}$ and $\mathbf{U}_t \sim \pi_{att}$ under $\mathbb{P}$, we have $\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s \leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_{att}))$. Next, plug in the formula (2.13):

$$\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s \leq t}) = \tfrac{1}{2} \ln \det R_{att} - \tfrac{1}{2} \mathbb{E} \ln \det R_t + \tfrac{1}{2} \mathrm{tr}(\mathbb{E} R_t R_{att}^{-1}) - \tfrac{|\mathbf{K}|}{2}$$
$$+ \tfrac{1}{2} \mathrm{tr}\big((\mathbb{E}(\vec{m}_t - \vec{m}_t^{att})(\vec{m}_t - \vec{m}_t^{att})^*) R_{att}^{-1} \big)$$

Using that $\vec{m}_t$ is the posterior mean, so by $\mathbb{E}_{\mathbf{X}_{s \leq t}} \mathbf{U}_t = \vec{m}_t$ and $\vec{m}_t, \vec{m}_t^{att}$ is $\sigma(\mathbf{X}_{s \leq t})$-adapted:

$$\mathbb{E}_{\mathbf{X}_{s \leq t}}(\vec{m}_t - \mathbf{U}_t)(\vec{m}_t - \vec{m}_t^{att})^* = [\mathbb{E}_{\mathbf{X}_{s \leq t}}(\vec{m}_t - \mathbf{U}_t)](\vec{m}_t - \vec{m}_t^{att})^* = 0.$$

Hence

$$R_{att} = \mathbb{E}(\mathbf{U}_t - \vec{m}_t^{att})(\mathbf{U}_t - \vec{m}_t^{att})^*$$
$$= \mathbb{E}(\mathbf{U}_t - \vec{m}_t + \vec{m}_t - \vec{m}_t^{att})(\mathbf{U}_t - \vec{m}_t + \vec{m}_t - \vec{m}_t^{att})^*$$
$$= \mathbb{E}(\mathbf{U}_t - \vec{m}_t)(\mathbf{U}_t - \vec{m}_t)^* + \mathbb{E}(\vec{m}_t - \vec{m}_t^{att})(\vec{m}_t - \vec{m}_t^{att})^*$$
$$= \mathbb{E} R_t + \mathbb{E}(\vec{m}_t - \vec{m}_t^{att})(\vec{m}_t - \vec{m}_t^{att})^*$$

In other word, $\mathbb{E}(\vec{m}_t - \vec{m}_t^{att})(\vec{m}_t - \vec{m}_t^{att})^* = R_{att} - \mathbb{E} R_t$, so

$$\mathrm{tr}\big((\mathbb{E}(\vec{m}_t - \vec{m}_t^{att})(\vec{m}_t - \vec{m}_t^{att})^*) R_{att}^{-1} \big) = \mathrm{tr}\big(I_{|\mathbf{K}|} \big) - \mathrm{tr}\big(\mathbb{E} R_t R_{att}^{-1} \big).$$

Put this back to the equation of $\mathcal{I}(\mathbf{U}_t, \mathbf{X}_{s \leq t})$, we have shown the identity. $\qquad \square$

Appendix A.3. Verifications of various claims

Lemma Appendix A.4. *There exists two diagonal matrices R_L^-, R^+, $R_L^- \geq mL^{-1/2}I$ for some constant $m > 0$, such that if $R_L^- \leq R_0 \leq MI$, $M \geq \|R^+\|$ then the solution to (2.9) is uniformly bounded in time by $R_L^- \leq R_t \leq MI$ a.s. Hence $\mathbb{E}|\vec{m}_t|^2 < \infty$.*

Proof. Consider the following deterministic processes. Let $R_{+,t}$ be the solution to the following:

$$\dot{R}_{+,t} = -\Gamma R_{+,t} - R_{+,t} \Gamma^* + \mathbf{\Sigma}, \quad R_{+,0} = MI$$

and let R^+ be the stable point of the dynamics above. Therefore if $R_{+,0} = MI \geq R_+$, $R_{+,t}$ will be decreasing in time, so $R_{+,t} \leq MI$.

Next let $R_{-,t} = R_L$ be the stable point of following

$$\dot{R}_{-,t} = -\Gamma R_{-,t} - R_{-,t} \Gamma^* + \mathbf{\Sigma} - \sigma_x^{-2} L |\mathbf{K}| R_{-,t}^2$$

Since this is a diagonalizable matrix equation, one can easily verify that each entry of R_L is of order $\sqrt{L}^{-1/2}$.

Using the comparison theorem of differential Riccati equation, Theorem 4.1.4 of [30], because $0 \leq \mathbf{P}(t) \leq |\mathbf{K}|I$, we have $R_{-,t} \leq R(t) \leq R_{+,t}$ a.s. This concludes our claim of R_t.

Then notice that by Young's inequality

$$\mathbb{E}|\vec{m}_t|^2 \leq 2\mathbb{E}|\vec{m}_t - \mathbf{U}_t|^2 + 2\mathbb{E}|\mathbf{U}_t|^2 = 2\mathbb{E}\mathrm{tr}(R_t) + 2\mathrm{tr}(R_{att}) < \infty.$$

$\qquad \square$

Lemma Appendix A.5. *For any fixed* $\mathbf{j} \neq \mathbf{k}$, $t > 0, q > \frac{1}{2}$, $z_s^L := L^{-q}([\tilde{P}_s]_{\mathbf{j},\mathbf{k}})$ *is family of process that converges to 0 in the uniform norm of* $C([0, t])$ *under* $\mathbb{P}_{\vec{v}_{s \geq 0}}$ *when* $L \to \infty$.

Proof. Recall that if $\vec{j} = \vec{k}$, then $z_s^L \equiv 0$, this Lemma trivially holds. So we assume $\vec{j} \neq \vec{k}$ in the following.

By theorem 7.5 of [29], it suffices for us to check $z_s^L \to 0$ in $\mathbb{P}_{\vec{v}_{s \geq 0}}$ for each s, and also show the modulus of continuity condition for the family $\{z_s^L\}$.

Fix s, as $\{Y_l = \exp(i(\vec{k} - \vec{j}) \cdot \vec{X}_l(s))\}$ is a sequence of i.i.d. random complex variables in $\mathbb{P}_{\vec{v}_{s \geq 0}}$ with

$$\mathbb{E}_{\vec{v}_{s \geq 0}} Y_i = 0, \quad \mathrm{Var}_{\vec{v}_{s \geq 0}} Y_i = 1, \quad \mathbb{E}_{\vec{v}_{s \geq 0}} Y_l^2 = \frac{1}{|\mathbb{T}^2|} \int_{\mathbb{T}^2} \exp(2i(\vec{j} - \vec{k}) \cdot x) dx = 0.$$

By the central limit theorem for complex random variables,

$$\frac{1}{\sqrt{L}} [\tilde{P}_s]_{\mathbf{j},\mathbf{k}} = (\vec{r}_{\mathbf{j}}^* \vec{r}_{\mathbf{k}}) \frac{1}{\sqrt{L}} \sum_{l=1}^{L} Y_l \to_p (\vec{r}_{\mathbf{j}}^* \vec{r}_{\mathbf{k}}) \cdot \mathcal{N}(0, 1, 0) \quad \text{for a.s. } \vec{v}_{s \leq t}.$$

Where $X + Yi \sim \mathcal{N}(0, 1, 0)$ iff $X \perp Y, X, Y \sim \mathcal{N}(0, \frac{1}{2})$, see [32] for more details of complex Gaussian distribution. So for any $q > 1/2$, $L^{-q}[\tilde{P}_s]_{\mathbf{j},\mathbf{k}} \to 0$ in $\mathbb{P}_{\vec{v}_{s \geq 0}}$ for any fixed $s \leq t$.

Next we check the modulus continuity condition for the processes $\{z_s^L\}$. By the corollary following Theorem 7.4 of [29], it suffices for us to show for any $\epsilon, \eta > 0$, there are $\delta > 0$ and L_0 such that for any $0 \leq r \leq t - \delta$, and $L \geq L_0$

$$\frac{1}{\delta} \mathbb{P}_{\vec{v}_{s \geq 0}} \left(\sup_{r \leq s \leq r + \delta} |z_s^L - z_r^L| \geq 2\epsilon \right) \leq \eta.$$

Applying the Itô's formula we have

$$z_s^L - z_r^L = \int_r^s L^{-q} \sum_{l=1}^{L} \exp(i(\vec{k} - \vec{j}) \cdot \vec{X}_l(u))[i(\vec{k} - \vec{j}) \cdot \vec{v}(\vec{X}_l(u), u) + \sigma_x^2 |\vec{k} - \vec{j}|^2] du$$

$$+ \sigma_x \int_r^s L^{-q} \sum_{l=1}^{L} \exp(i(\vec{k} - \vec{j}) \cdot \vec{X}_l(u)) i(\vec{k} - \vec{j}) \cdot dW_l^x(u)$$

$$= \int_r^s L^{-q} \sum_{l=1}^{L} [f_u(\vec{X}_l(u)) du + g_u(\vec{X}_l(u)) dW_l^x(u)]$$

with f_u, g_u representing the corresponding functions. So it suffices to show

$$\frac{1}{\delta} \mathbb{P}_{\vec{v}_{s \geq 0}} \left(\sup_{s \leq r + \delta} \left| \int_r^s L^{-q} \sum_{l=1}^{L} g_u(\vec{X}_l(u)) dW_l^x(u) \right| + \sup_{s \leq r + \delta} \left| \int_r^s L^{-q} \sum_{l=1}^{L} f_u(\vec{X}_l(u)) du \right| \geq 2\epsilon \right) \leq \eta.$$

$$\text{(A.1)}$$

Apply Doob's inequality and Itô's isometry to the martingale part of (A.1):

$$\frac{1}{\delta}\mathbb{P}_{\vec{v}_{s\geq0}}\left(\sup_{s\leq r+\delta}\left|\int_r^s L^{-q}\sum_{l=1}^L g_u(\vec{X}_l(u))\mathrm{d}W_l^x(u)\right|\geq\epsilon\right)$$

$$\leq\frac{L^{-2q}}{\epsilon^2}\mathbb{E}_{\vec{v}_{s\geq0}}\int_r^{r+\delta}\sum_{l=1}^L|g_u(\vec{X}_l(t))|^2\mathrm{d}u\leq\frac{L^{1-2q}\sigma_x^2|\vec{k}-\vec{j}|^2}{\epsilon^2}.$$

We can find L_0 so the right hand side is sufficiently small when $L\geq L_0$.

Apply the Cauchy Schwarz inequality to the Riemann integral part in (A.1):

$$\sup_{s\leq r+\delta}\left|\int_r^s L^{-q}\sum_{l=1}^L f_u(\vec{X}_l(u))\mathrm{d}u\right|^2\leq\delta^2 L^{-2q}\int_r^{r+\delta}\left|\sum_{l=1}^L f_u(\vec{X}_l(u))\right|^2\mathrm{d}u.$$

Using the Chebychev's inequality,

$$\frac{1}{\delta}\mathbb{P}_{\vec{v}_{s\geq0}}\left(\sup_{s\leq r+\delta}|\int_r^s L^{-q}\sum_{l=1}^L f_u(\vec{X}_l(u))\mathrm{d}u|\geq\epsilon\right)\leq\frac{\delta}{\epsilon^2}L^{-2q}\int_r^{r+\delta}\mathbb{E}_{\vec{v}_{s\geq0}}\left|\sum_{l=1}^L f_u(\vec{X}_l(u))\right|^2\mathrm{d}u.$$

$$(\text{A.2})$$

Next notice that

$$\mathbb{E}_{\vec{v}_{s\geq0}}f_u(\vec{X}_l(u))=\frac{1}{|\mathbb{T}^d|}\int_{\mathbb{T}^d}\exp(\mathrm{i}(\vec{k}-\vec{j})\cdot x)[\mathrm{i}(\vec{k}-\vec{j})\cdot\vec{v}(x,u)+\sigma_x^2|\vec{k}-\vec{j}|^2]\mathrm{d}x$$

$$=\mathrm{i}\sum_{\vec{n}=\vec{j}-\vec{k}}(\vec{k}-\vec{j})\cdot\vec{r}_\mathbf{n}\hat{v}_\mathbf{n}(u)=0,$$

the second equality invokes the decomposition (1.1) where all modes with wavenumbers not being $\vec{j}-\vec{k}$ vanishes in the periodic integral; the third equality holds because $(\vec{k}-\vec{j})\cdot\vec{r}_\mathbf{n}=0$ by definition of $\vec{r}_\mathbf{n}$.

Since $\vec{X}_l(u)$ are conditionally independent,

$$\mathbb{E}_{\vec{v}_{s\geq0}}\left|\sum_{l=1}^L f_u(\vec{X}_l(u))\right|^2=L\mathbb{E}_{\vec{v}_{s\geq0}}|f_u(\vec{X}_l(u))|^2\leq LC(|\vec{v}_u|_\infty^2+1).$$

Using this inequality, the right hand side of (A.2) is bounded by

$$\frac{\delta}{\epsilon^2}L^{1-2q}C\left(\delta+\int_r^{r+\delta}|\vec{v}_u|_\infty^2\mathrm{d}u\right)$$

which goes to 0 for large L and small δ. Hence we complete the verification of continuous modulus condition. □

Lemma Appendix A.6. *If $|\psi(s,t)|\leq C\exp(-d_L(t-s)+D\sqrt{L}\exp(-\mu_L s))$, suppose s_0 is a constant such that $d_L s\geq D\sqrt{L}, D\sqrt{L}\exp(-\mu_L s)\leq 1$ for $s\geq\frac{1}{2}s_0$ and L sufficiently large, then for any $t>s_0$,*

(i) *With any fixed continuous process v_s, if $v_t = 0$ and $|d(s)| \leq AL \exp(-\mu_L s) + \sqrt{L}B$, then as $L \to \infty$,*

$$\int_0^t \psi(s,t)d(s)v_s \mathrm{d}s \to 0.$$

(ii) *With $|d(s)| \leq B$, $\int_0^t \psi(s,t)d(s)\mathrm{d}s \to 0$ as $L \to \infty$.*

Proof. For the claim (i), note

$$AL \int_0^t \psi(s,t)\exp(-\mu_L s)v_s \mathrm{d}s \leq ACL \sup_{s \leq t}|v_s|t \exp(-d_L t + D\sqrt{L})$$

If $t \geq s_0$, when $L \to \infty$, the right hand side tends to 0.

And for $\sqrt{L}\int_0^t \phi(s,t)v_s$, because $v_t = 0$ and v_s is continuous, for any fixed $\xi > 0$, there exists a constant $\delta > 0$ such that $|v_s| \leq \xi$ for $s \geq t - \delta$. Hence we can break the integration into three parts and bound each using $d_L s \geq D\sqrt{L}, D\sqrt{L}\exp(-\mu_L s) \leq 1, \forall s \geq \frac{1}{2}s_0$:

$$\sqrt{L}\int_{t-\delta}^t \psi(s,t)v_s \mathrm{d}s \leq C\sqrt{L}\xi \int_{t-\delta}^t \exp(-d_L(t-s)+1)\mathrm{d}s \leq Cd_L^{-1}\sqrt{L}e\xi.$$

$$\sqrt{L}\int_{\frac{1}{2}s_0}^{t-\delta} \psi(s,t)v_s \mathrm{d}s \leq \sqrt{L}\sup_{s \leq t}|v_s|\int_{\frac{1}{2}s_0}^{t-\delta} \exp(-d_L(t-s)+1)\mathrm{d}s \leq ed_L^{-1}\sqrt{L}\sup_{s \leq t}|v_s|\exp(-d_L\delta)$$

$$\sqrt{L}\int_0^{\frac{1}{2}s_0} \psi(s,t)v_s \mathrm{d}s \leq \sqrt{L}\sup_{s \leq t}|v_s|\int_0^{\frac{1}{2}s_0} \exp(-d_L(t-s)+D\sqrt{L})\mathrm{d}s$$
$$\leq \sqrt{L}d_L^{-1}\sup_{s \leq t}|v_s|\exp(-d_L(t-s_0))\mathrm{d}s.$$

Recall d_L is of order $\sqrt{L}$, the last two terms go to 0 when $L \to \infty$ if $t > s_0$. The first is of order ξ. Since ξ is an arbitrary number, our claim is true.

For (ii), just notice

$$\int_0^t \psi(s,t)\mathrm{d}s = \int_{\frac{1}{2}s_0}^t \psi(s,t)\mathrm{d}s + \int_0^{\frac{1}{2}s_0} \psi(s,t)\mathrm{d}s \leq ed_L^{-1} + \sqrt{L}d_L^{-1}\exp(-d_L(t-s_0)),$$

the right hand side goes to 0 as L increases. $\qquad\square$

References

[1] Griffa A, Kirwan A D, Mariano A J, Özgökmen T M, and Rossby T. *Lagrangian analysis and prediction of coastal and ocean dynamics.* Cambridge University Press, 2007.

[2] Gould J and et al. Argo profiling floats bring new era of in situ ocean observations. *Eos Trans. AGU,* 85(19):185–191, 2004.

[3] Molcard A, Piterbarg L I, Griffa A, zgkmen T M, and Mariano A J. Assimilation of drifter observations for the reconstruction of the eulerian circulation field. *Journal of Geophysical Research: Oceans,* 108(C3):3056, 2003.

[4] Barreiro A, Liu S, Namachchivaya N S, Sauer P W, and Sower R B. Data assimilation in the detection of vortices. In *Applications of Nonlinear Dynamics*, pages 47–59. Springer Berlin Heidelberg, 2009.

[5] Kuznetsov L, Ide K, and Jones C K R T. A method for assimilation of lagrangian data. *Mon. Wea. Rev.*, 131(10):22472260, 2003.

[6] Salman H, Kuznetsov L, Jones C K R T, and Ide K. A method for assimilating lagrangian data into a shallow-water-equation ocean model. *Mon. Wea. Rev.*, 134:1081–1101, 2006.

[7] Salman H, Ide K, and Jones C K R T. Using flow geometry for drifter deployment in lagrangian data assimilation. *Tellus A*, 60:321335, 2008.

[8] Apte A, Jones C K R T, and Stuart A M. A bayesian approach to lagrangian data assimilation. *Dynamic Meteorology and Oceanography*, 60(2):336–347, 2008.

[9] Apte A, Jones C K R T, Stuart A M, and Voss J. Data assimilation: Mathematical and statistical perspectives. *Int. J. Numer. Meth. Fluids.*, 56:1033–1046, 2008.

[10] Majda A J. Challenges in climate science and contemporary applied mathematics. *Comm. Pure Appl. Math*, 65(7):920–948, 2012.

[11] Majda A J and Branicki M. Lessons in uncertainty quantification for turbulent dynamical system. *Discrete and Continuous Dynamical Systems*, 32(9):3133–3221, 2013.

[12] Majda A J and Gershgorin B. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proc. Natl. Acad. Sci.*, 108(31):12599–12604, 2011.

[13] Branicki M, Chen N, and Majda A J. Non-gaussian test models for prediction and state estimation with model errors. *Chin. Ann. Math.*, 34(1):29–64, 2013.

[14] Elliott Jr F W, Horntrop D J, and Majda A J. A fouriercwavelet monte carlo method for fractal random fields. *J. Comput. Phys*, 132(2):384–408, 1997.

[15] Majda A J and Harlim J. *Filtering complex turbulent systems*. Cambridge University Press, Cambridge, UK, 2012.

[16] Liptser R S and Shiryaev A N. *Statistics of random processes. I, II,*, volume 5 of *Applications of Mathematics*. Springer-Verlag, 2001.

[17] Kallenberg O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, 1997.

[18] Branicki M and Majda A J. Quantifying uncertainty for predictions with model error in non-guassian systems with intermittency. *Nonlinearity*, 25(9):2543–2578, 2012.

[19] Giannakis D, Majda A J, and Horenko I. Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems. *Physica D*, 241(20):1735–1752, 2012.

[20] Kleeman R. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, 59:2057–2072, 2002.

[21] Majda A J and Wang X. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, Cambridge, UK, 2006.

[22] Branicki M and Majda A J. Quantifying filter performance for turbulent dynamical systems through information theory. *Comm. Math. Sci.*, in press.

[23] DelSole T. Predictability and information theory. part i: Measures of predictability. *J. Atmos. Sci.*, 61:2425–2440, 2004.

[24] Giannakis D and Majda A J. Quantifying the predictive skill in long-range forecasting. part i: Coarse-grained predictions in a simple ocean model. *J. Climate.*, 25:4523–4548, 2012.

[25] Schneider T and Griffies S M. A conceptual framework for predictability studies. *J. Climate.*, 12:3133–3155, 1999.

[26] Lindvall T. *Lectures on the coupling methods*. Dover Publications, Inc, 2002.

[27] Chirikjian G S. *Stochastic Models, Information Theory, and Lie Groups*, volume 1 of *Applied and Numerical Harmonic Analysis*. Springer-Verlag, 2009.

[28] Karatzas I and Shreve S. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer-Verlag, 1991.

[29] Billingsley P. *Convergence of Probability Measure*. Wiley Series in Probability and Statistics. Wiley-

Interscience, 1999.

[30] Abou-Kandil H, Freiling G, Ionescu V, and Jank G. *Matrix Riccati Equations in Control and Systems Theory*. Systems & Control: Foundations & Applications. Birkhauser Verlag, 2003.

[31] Lawson J and Lim Y. A birkhoff contraction formula with applications to riccati equations. *SIAM J. Control Optim*, 46:930–951, 2007.

[32] Picinbono B. Second-order complex random vectors and normal distributions. *IEEE Trans on Signal Processing*, 44, 1996.

[33] Csiszár I and Köruer J. *Information theory: coding theorems for discrete memoryless systems, 2nd edition*. Cambridge University Press, 2011.