

Statistical non-locality of dynamically coherent structures

Andre N. Souza^{1,†}, Tyler Lutz² and Glenn R. Flierl¹

¹Massachusetts Institute of Technology, Cambridge, MA, USA

²Otto-von-Guericke-University Magdeburg, Magdeburg, Saxony-Anhalt, Germany

(Received 4 December 2022; revised 23 April 2023; accepted 29 May 2023)

We analyse a class of stochastic advection problems by conditionally averaging the passive tracer equation with respect to a given flow state. In doing so, we obtain expressions for the turbulent diffusivity as a function of the flow statistics spectrum. When flow statistics are given by a continuous-time Markov process with a finite state space, calculations are amenable to analytic treatment. When the flow statistics are more complex, we show how to approximate turbulent fluxes as hierarchies of finite state space continuous-time Markov processes. The ensemble average turbulent flux is expressed as a linear operator that acts on the ensemble average of the tracer. We recover the classical estimate of turbulent flux as a diffusivity tensor, the components of which are the integrated autocorrelation of the velocity field in the limit that the operator becomes local in space and time.

Key words: turbulence modelling, turbulence theory, turbulent mixing

1. Introduction

The study of passive tracer transport is of fundamental importance in characterizing turbulence. However, predicting a chaotic dynamical trajectory over long times is infeasible (Lorenz 1963), so one must switch to a statistical perspective to make headway on transport properties. From the analysis of dispersion by Taylor (1922), operator notions of mixing from Knobloch (1977), computations of ‘effective diffusivity’ by Avellaneda & Majda (1991), simplified models of turbulence by Pope (2011), rigorous notions of mixing in terms of Sobolev norms by Thiffeault (2012) or upper bounds on transport of Hassanzadeh, Chini & Doering (2014), a variety of different approaches have been developed to elucidate fundamental properties of turbulence.

† Email address for correspondence: andrenogueirasouza@gmail.com

In addition to furthering our understanding of turbulence, there are practical applications for turbulence closures. In particular, Earth Systems Models require closure relations for the transport of unresolved motions (Schneider *et al.* 2017); however, the closure relations are marred by structural and parametric uncertainty, requiring *ad hoc* tuning to compensate for biases. Structural biases are associated with scaling laws and closure assumptions between turbulent fluxes and gradients. Modern studies are bridging the gap by incorporating more complex physics and novel scaling laws, see as examples the works of Tan *et al.* (2018) and Gallet & Ferrari (2020), but the correct functional forms to represent fluxes still need to be discovered.

The multi-scale nature of turbulent flows, coherent structures and the interconnection of reacting chemical species and prognostic fields suggest that fluxes are better modelled using nonlinear and non-local operators in space, time and state. Data-driven methods relying on flexible interpolants can significantly reduce structural bias, but often at the expense of interpretability, generalizability or efficiency. Thus, understanding scaling laws and functional forms of turbulence closures is still necessary to physically constrain data-driven methods and decrease their computational expense. A promising avenue for significant progress, lying at the intersection of theory and practice, is the calculation of closure relations for passive tracers.

The present work addresses a fundamentally different question than that of short-time Lagrangian statistics, such as the works of Taylor (1922), Moffatt (1983), Weeks, Urbach & Swinney (1996) and Falkovich, Gawędzki & Vergassola (2001), since our focus is on Eulerian statistically steady statistics. Consequently, we develop new non-perturbative techniques to address the question at hand. In particular, the use of path integrals or Taylor expansions is stymied since perturbation methods based on local information often yield irredeemably non-convergent expansions for strong flow fields and long time limits. The calculations herein are more akin to calculating a ground state in quantum mechanics rather than short-time particle scattering.

A middle ground is the work of Gorbunova *et al.* (2021), which analyses short-time Eulerian two-point statistics and makes connections to the Lagrangian approach. There they use renormalization groups as the workhorse for calculations. The most similar prior work regarding problem formulation is the upper bound analysis of Hakulinen (2003), in which all moments of a stochastically advected passive tracer were considered. The scope was different because the goal was to understand all moments of the resulting stochastic tracer field advected by the Gaussian model of Kraichnan (1968). Here we focus on exact expressions for the ensemble mean flux caused by a general class of flow fields. Although we only focus on the ensemble mean flux of a passive tracer, we do not make restrictions on the temporal statistics of the stochastic flow field. Furthermore, the methodology we outline extends to other moments of the flow field and short-time statistics, although we emphasize neither.

We make arguments akin to those by Kraichnan (1968) to motivate the operator approach, but our calculation method is fundamentally different. Given that the goal is to construct an operator rather than estimate a diffusivity tensor acting on the ensemble mean gradients or deriving upper bounds, we take a field-theoretic perspective (Hopf 1952). Doing so allows us to derive a coupled set of partial differential equations representing conditional mean tracers where the conditional averages are with respect to different flow states. The flux associated with fluctuations is then a Schur complement of the resulting linear system with respect to statistical ‘perturbation’ variables. Moreover, if the flow statistics are given by a continuous-time Markov process with a small finite state space, the Schur complement becomes tractable to compute analytically. Hereafter, we refer

to the flux associated with fluctuations as the turbulent flux (which is sometimes also called an ‘eddy flux’), and the corresponding operator is the turbulent diffusivity operator (alternatively an ‘effective diffusivity operator’).

The paper is organized as follows. In §2, we formulate the closure problem and recast it as one of solving partial differential equations. In §3, we give two examples: a one-dimensional (1-D) tracer advected by an Ornstein–Uhlenbeck process and a tracer advected by a flow field with an arbitrary spatial structure that switches between three states. Section 4 outlines the general theory, and we apply it to a travelling stochastic wave in a channel in §5. Appendices supplement the body of the manuscript. Appendix A shows how a ‘white-noise’ limit reduces the turbulent diffusivity to a local diffusivity tensor and how the present theory encompasses the class of flow fields of Kraichnan (1968) in the ‘white-noise’ limit. Appendix B provides a direct field-theoretic derivation of arguments in §2, and Appendix C provides a heuristic overview of obtaining continuous-time Markov processes with finite state space and their statistics from deterministic or stochastic dynamical systems.

2. Problem formulation

We consider the advection and diffusion of an ensemble of passive tracers θ_ω

$$\partial_t \theta_\omega + \nabla \cdot (\tilde{\mathbf{u}}_\omega \theta_\omega - \kappa \nabla \theta_\omega) = s(\mathbf{x}) \quad (2.1)$$

by a stochastic flow field $\tilde{\mathbf{u}}_\omega(\mathbf{x}, t)$ where ω labels the ensemble member. Here, s is a deterministic mean zero source term and κ is a diffusivity constant. (For laboratory flows, κ would be the molecular diffusivity; for larger-scale problems, we rely on the fact that away from boundaries, the ensemble mean advective flux (but not necessarily other statistics) may still be much larger than the diffusive flux. Thus, the exact value of κ will not matter.) Our target is to obtain a meaningful equation for the ensemble mean,

$$\partial_t \langle \theta \rangle + \nabla \cdot (\langle \tilde{\mathbf{u}} \theta \rangle - \kappa \nabla \langle \theta \rangle) = s(\mathbf{x}), \quad (2.2)$$

which requires a computationally amenable expression for the mean advective flux, $\langle \tilde{\mathbf{u}} \theta \rangle$, in terms of the statistics of the flow field, $\tilde{\mathbf{u}}_\omega(\mathbf{x}, t)$, and the ensemble average of the tracer, $\langle \theta \rangle$. Thus, the closure problem is to find an operator O that relates the ensemble mean, $\langle \theta \rangle$, to the ensemble mean advective-flux, $\langle \tilde{\mathbf{u}} \theta \rangle$, i.e.

$$O[\langle \theta \rangle] = \langle \tilde{\mathbf{u}} \theta \rangle. \quad (2.3)$$

We show how to define (and solve for) the operator O . We do so by writing down the Fokker–Planck/master equation for the joint Markov process $(\tilde{\mathbf{u}}_\omega, \theta_\omega)$ corresponding to (2.1), integrating with respect to all possible tracer field configurations and manipulating the resulting system of partial differential equations. The operator will be linear with respect to its argument and depend on the statistics of the flow field.

We assume all tracer ensemble members to have the same initial condition and, thus, the ensemble average here is with respect to different flow realizations. The only source of randomness comes from different flow realizations. Throughout the manuscript, we assume homogeneous Neumann boundary conditions for the tracer and zero wall-normal flow for the velocity field when boundaries are present. Combined with the assumption that the source term is mean zero, these restrictions imply that the tracer average is conserved.

For the statistics of the flow field, we consider a continuous-time Markov process with known statistics, as characterized by the generator of the process. When the state space of the flow is finite with N states, we label all possible flow configurations corresponding to

steady flow fields by $\mathbf{u}_n(\mathbf{x})$, where n is the associated state index. We will keep with the convention of using the $\tilde{\mathbf{u}}$ as a dynamic variable and \mathbf{u} as a fixed flow structure. These choices represent the flow as a Markov jump process where the jumps occur between flow fields with a fixed spatial structure.

Physically, we think of these states as representing coherent structures in a turbulent flow. (This is viewed as a finite volume discretization in function space where the states are the ‘cell averages’ of a control volume in function space. Thus, the transition probability matrix becomes an approximated Perron–Frobenius operator of the turbulent flow.) Suri *et al.* (2017) provide the experimental and numerical motivation for this philosophy in the context of an approximate Kolmogorov flow. By its very nature, a turbulent flow that is chaotic and unpredictable over long time horizons is modelled as being semi-unpredictable through our choice. Over short horizons, the probability of remaining in a given state is large. The flow is limited to moving to a subset of likely places in phase space in the medium term. Over long time horizons, the most one can say about the flow is related to the likelihood of being found in the appropriate subset of phase space associated with the statistically steady state.

Thus, we proceed by characterizing the probability, \mathbb{P} , of transitioning from state n to state m by a transition matrix $\mathcal{P}(\tau)$,

$$\mathbb{P}\{\tilde{\mathbf{u}}_\omega(\mathbf{x}, t + \tau) = \mathbf{u}_m(\mathbf{x}) | \tilde{\mathbf{u}}_\omega(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})\} = [\mathcal{P}(\tau)]_{mn}. \tag{2.4}$$

The transition probability is defined through its relation to the generator Q ,

$$\mathcal{P}(\tau) \equiv \exp(Q\tau), \tag{2.5}$$

where $\exp(Q\tau)$ is a matrix exponential. Each entry of $\mathcal{P}(\tau)$ must be positive. Furthermore, for each τ , the columns of $\mathcal{P}(\tau)$ sum to one since the total probability must always sum to one. Similarly, Q ’s off-diagonal terms must be positive and the column sum of Q must be zero. (Indeed, to first order, $\exp(Q dt) = \mathbb{I} + Q dt$. The positivity requirement of the transition probability $\mathcal{P}(dt) = \exp(Q dt)$ necessitates the positivity of Q ’s off-diagonal terms and thus, in turn, the negativity of its diagonal terms.)

We denote the probability of an ensemble member, $\tilde{\mathbf{u}}_\omega$, being found at state m at time t by $\mathcal{P}_m(t)$,

$$\mathcal{P}_m(t) = \mathbb{P}\{\tilde{\mathbf{u}}_\omega(\mathbf{x}, t) = \mathbf{u}_m(\mathbf{x})\}. \tag{2.6}$$

The evolution equation for $\mathcal{P}_m(t)$ is the master equation,

$$\frac{d}{dt}\mathcal{P}_m = \sum_n Q_{mn}\mathcal{P}_n. \tag{2.7}$$

We assume that (2.7) has a unique steady state and denote the components of the steady state by P_m .

We have used several ‘P’ at this stage, and their relation is:

- (i) \mathbb{P} denotes a probability;
- (ii) $\mathcal{P}(\tau)$ denotes the transition probability matrix for a time τ in the future;
- (iii) $\mathcal{P}_m(t)$ denotes the probability of being in state m at time t . The algebraic relation

$$\sum_m \mathcal{P}_m(t + \tau)\hat{\mathbf{e}}_m = \mathcal{P}(\tau) \sum_n \mathcal{P}_n(t)\hat{\mathbf{e}}_n \tag{2.8}$$

holds;

(iv) in addition, we use P_m for the statistically steady probability of being found in state m , in the limit

$$\lim_{t \rightarrow \infty} \mathcal{P}_m(t) = P_m. \tag{2.9}$$

We exploit the information about the flow field to infer the mean statistics of the passive tracer θ_ω . We do so by conditionally averaging the tracer field θ_ω with respect to a given flow state \mathbf{u}_m . More precisely, given the stochastic partial differential equation,

$$\mathbb{P}\{\tilde{\mathbf{u}}_\omega(\mathbf{x}, t + \tau) = \mathbf{u}_m(\mathbf{x}) | \tilde{\mathbf{u}}_\omega(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})\} = [\exp(\mathcal{Q}\tau)]_{mn}, \tag{2.10}$$

$$\partial_t \theta_\omega + \nabla \cdot (\tilde{\mathbf{u}}_\omega \theta_\omega - \kappa \nabla \theta_\omega) = s(\mathbf{x}), \tag{2.11}$$

we shall obtain equations for probability weighted conditional means of θ_ω ,

$$\Theta_m(\mathbf{x}, t) \equiv \langle \theta_\omega \rangle_{\tilde{\mathbf{u}}_\omega(\mathbf{x}, t) = \mathbf{u}_m(\mathbf{x})} \mathcal{P}_m(t). \tag{2.12}$$

Empirically, the probability weighted conditional average is computed by examining all ensemble members at a fixed time step, adding up only the ensemble members that are currently being advected by state \mathbf{u}_m and then dividing by the total number of ensemble members. We will show that the evolution equation for Θ_m is

$$\frac{d}{dt} \mathcal{P}_m = \sum_n \mathcal{Q}_{mn} \mathcal{P}_n, \tag{2.13}$$

$$\partial_t \Theta_m + \nabla \cdot (\mathbf{u}_m \Theta_m - \kappa \nabla \Theta_m) = s(\mathbf{x}) \mathcal{P}_m + \sum_n \mathcal{Q}_{mn} \Theta_n. \tag{2.14}$$

As we shall see, the explicit dependence on the generator in (2.14) yields considerable information. We recover the equation for the tracer ensemble mean, (2.2), by summing (2.14) over the index m , using $\langle \theta \rangle = \sum_m \Theta_m$, $\sum_m \mathcal{Q}_{mn} = \mathbf{0}$, and $\sum_m \mathcal{P}_m = 1$,

$$\partial_t \sum_m \Theta_m + \nabla \cdot \left(\sum_m \mathbf{u}_m \Theta_m - \kappa \nabla \sum_m \Theta_m \right) = s(\mathbf{x}) \tag{2.15}$$

\Leftrightarrow

$$\partial_t \langle \theta \rangle + \nabla \cdot (\langle \tilde{\mathbf{u}} \theta \rangle - \kappa \nabla \langle \theta \rangle) = s(\mathbf{x}). \tag{2.16}$$

The presence of the generator when taking conditional averages is similar to the entrainment hypothesis in the atmospheric literature. See, for example, Tan *et al.* (2018) for its use in motivating a turbulence closure; however, here, we derive the result from the direct statistical representation instead of hypothesizing its presence from a dynamical argument.

We give a brief derivation of (2.13) and (2.14) using a discretization of the advection-diffusion equation here. For an alternative derivation where we forego discretization, see Appendix B, and for a brief overview of the connection between the discrete, continuous and mixed master equation, see Appendix C or, in a simpler context, the work of Hagan, Doering & Levermore (1989). Most of the terms in (2.14) are obtained by applying a conditional average to (2.11), commuting with spatial derivatives when necessary and then multiplying through by \mathcal{P}_m ; however, the primary difficulty lies in proper treatment of the conditional average of the temporal derivative. We circumvent the problem in a roundabout manner: discretize the advection-diffusion equation, write down

the resulting master equation, compute moments of the probability distribution and then take limits to restore the continuum nature of the advection-diffusion equation.

A generic discretization (in any number of dimensions) of (2.11) is of the form

$$\frac{d}{dt}\theta^i + \sum_{jkc} A_{ijk}^c \tilde{u}_\omega^{k,c} \theta^j - \sum_j D_{ij} \theta^j = s^i \tag{2.17}$$

for some tensor A_{ijk}^c , representing advection, and matrix D_{ij} , representing diffusion. Here, each i, j and k corresponds to a spatial location, and the index c corresponds to a component of the velocity field $\tilde{\mathbf{u}}$. The variable θ^i is the value of the tracer at grid location i and $v^{k,c}$ is the value of the c th velocity component and grid location k . The master equation for the joint probability density for each component θ^i and Markov state m , $\rho_m(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta^1, \theta^2, \dots)$ and the m -index denotes a particular Markov state, is a combination of the Liouville equation for (2.17) and the transition rate equation for (2.10),

$$\partial_t \rho_m = \sum_i \frac{\partial}{\partial \theta^i} \left[\left(\sum_{jkc} A_{ijk}^c u_m^{k,c} \theta^j - \sum_j D_{ij} \theta^j - s^i \right) \rho_m \right] + \sum_n Q_{mn} \rho_n. \tag{2.18}$$

Define the following moments:

$$\mathcal{P}_m = \int d\boldsymbol{\theta} \rho_m \quad \text{and} \quad \Theta_m^j = \int d\boldsymbol{\theta} \theta^j \rho_m. \tag{2.19a,b}$$

We obtain an equation for \mathcal{P}_m by integrating (2.18) by $d\boldsymbol{\theta}$ to yield

$$\frac{d}{dt} \mathcal{P}_m = \sum_n Q_{mn} \mathcal{P}_n \tag{2.20}$$

as expected from (2.7). The equation for Θ_m^ℓ is obtained by multiplying (2.18) by θ^ℓ and then integrating with respect to $d\boldsymbol{\theta}$,

$$\frac{d}{dt} \Theta_m^\ell = - \sum_{jkc} A_{ijk}^c u_m^{k,c} \Theta_m^j + \sum_j D_{\ell j} \Theta_m^j + s^\ell \mathcal{P}_m + \sum_n Q_{mn} \Theta_n^\ell, \tag{2.21}$$

where we integrated by parts on the $\int d\boldsymbol{\theta} \theta^\ell \partial_{\theta^i} \bullet$ term. Upon taking limits of (2.21), we arrive at (2.13) and (2.14), repeated here for convenience,

$$\frac{d}{dt} \mathcal{P}_m = \sum_n Q_{mn} \mathcal{P}_n, \tag{2.22}$$

$$\partial_t \Theta_m + \nabla \cdot (\mathbf{u}_m \Theta_m - \kappa \nabla \Theta_m) = s(x) \mathcal{P}_m + \sum_n Q_{mn} \Theta_n. \tag{2.23}$$

We compare (2.23) to the direct application of the conditional average to (2.11) followed by multiplication with \mathcal{P}_m to infer

$$\langle \partial_t \theta_\omega \rangle_{\tilde{\mathbf{u}}_\omega(x,t)=\mathbf{u}_m(x)} \mathcal{P}_m = \partial_t \Theta_m - \sum_n Q_{mn} \Theta_n. \tag{2.24}$$

In summary, for an m -dimensional advection-diffusion equation and N Markov states, (2.13) and (2.14) are a set of N -coupled m -dimensional advection-diffusion equations

with N different steady velocities. When c continuous variables describe the statistics of the flow field, the resulting equation set becomes an $m + c$ dimensional system. Stated differently, if the statistics of $\tilde{\mathbf{u}}_\omega$ are characterized by transitions between a continuum of states associated with variables $\boldsymbol{\alpha} \in \mathbb{R}^c$ and Fokker–Planck operator \mathcal{F}_α , then the conditional averaging procedure yields

$$\partial_t \mathcal{P} = \mathcal{F}_\alpha[\mathcal{P}], \tag{2.25}$$

$$\partial_t \Theta + \nabla \cdot (\mathbf{u}\Theta - \kappa \nabla \Theta) = s(\mathbf{x})\mathcal{P} + \mathcal{F}_\alpha[\Theta], \tag{2.26}$$

where $\mathcal{P} = \mathcal{P}(\boldsymbol{\alpha}, t)$, $\Theta = \Theta(\mathbf{x}, \boldsymbol{\alpha}, t)$ and $\mathbf{u} = \mathbf{u}(\mathbf{x}, \boldsymbol{\alpha})$. Equations (2.22) and (2.23) are thought of as finite volume discretizations of flow statistics in (2.25) and (2.26). We give an explicit example in § 3.1 and another in § 5.

Our primary concern in this work is to use (2.22) and (2.23), or analogously (2.25) and (2.26), to calculate meaningful expressions for $\langle \tilde{\mathbf{u}}\theta \rangle$; however, we shall first take a broader view to understand the general structure of the turbulent fluxes. We attribute the following argument to Weinstock (1969) but use different notation and make additional simplifications.

Applying the Reynolds decomposition

$$\theta_\omega = \langle \theta \rangle + \theta'_\omega \quad \text{and} \quad \tilde{\mathbf{u}}_\omega = \langle \tilde{\mathbf{u}} \rangle + \tilde{\mathbf{u}}'_\omega \tag{2.27}$$

yields

$$\partial_t \langle \theta \rangle + \nabla \cdot (\langle \tilde{\mathbf{u}} \rangle \langle \theta \rangle + \langle \tilde{\mathbf{u}}' \theta' \rangle - \kappa \nabla \langle \theta \rangle) = s, \tag{2.28}$$

$$\partial_t \theta'_\omega + \nabla \cdot (-\langle \tilde{\mathbf{u}} \rangle \theta' - \langle \tilde{\mathbf{u}}' \theta' \rangle + \tilde{\mathbf{u}}_\omega \theta'_\omega - \kappa \nabla \theta'_\omega) = 0. \tag{2.29}$$

The perturbation equation is rewritten as

$$\partial_t \theta'_\omega + \nabla \cdot (\tilde{\mathbf{u}}'_\omega \theta'_\omega - \langle \tilde{\mathbf{u}}' \theta' \rangle) + \langle \tilde{\mathbf{u}} \rangle \theta'_\omega - \kappa \nabla \theta'_\omega = -\nabla \cdot (\tilde{\mathbf{u}}'_\omega \langle \theta \rangle). \tag{2.30}$$

This equation is an infinite system (or finite, depending on the number of ensemble members) of coupled partial differential equations between the different ensemble members. The ensemble members are coupled due to the turbulent flux, $\langle \tilde{\mathbf{u}}' \theta' \rangle$. The key observation is to notice that the terms on the left-hand side involve the perturbation variables and not the ensemble mean of the gradients. Assuming it is possible to find the inverse, the Green’s function for the large linear system is used to yield

$$\theta'_\omega(\mathbf{x}, t) = - \int d\mathbf{x}' dt' d\mu_{\omega'} \mathcal{G}_{\omega\omega'}(\mathbf{x}, t | \mathbf{x}', t') \nabla \cdot (\tilde{\mathbf{u}}'_{\omega'} \langle \theta \rangle), \tag{2.31}$$

where we also have to integrate with respect to the measure defining the different ensemble members through $d\mu_{\omega'}$. In our notation, this implies $\langle \theta \rangle = \int d\mu_\omega \theta_\omega$. We use this expression to rewrite the turbulent flux as

$$\langle \tilde{\mathbf{u}}' \theta' \rangle = - \int d\mathbf{x}' dt' d\mu_\omega d\mu_{\omega'} \tilde{\mathbf{u}}'_{\omega'}(\mathbf{x}, t) \mathcal{G}_{\omega\omega'}(\mathbf{x}, t | \mathbf{x}', t') [\nabla \cdot (\tilde{\mathbf{u}}'_{\omega'}(\mathbf{x}', t') \langle \theta \rangle(\mathbf{x}', t'))]. \tag{2.32}$$

We make two simplifications for illustrative purposes.

- (i) All ensemble averages are independent of time.
- (ii) The flow is incompressible, i.e. $\nabla \cdot \tilde{\mathbf{u}} = 0$.

Equation (2.32) becomes

$$\langle \tilde{u}'\theta' \rangle = - \int dx' dt' d\mu_\omega d\mu_{\omega'} [\tilde{u}'_\omega(x, t) \mathcal{G}_{\omega\omega'}(x, t|x', t') \tilde{u}'_{\omega'}(x', t')] \cdot \nabla \langle \theta \rangle(x'). \quad (2.33)$$

We perform the t' , α , ω integrals first to define the turbulent diffusivity tensor kernel as

$$\langle \tilde{u}'\theta' \rangle = - \int dx' \underbrace{\int dt' d\mu_\omega d\mu_{\omega'} [\tilde{u}'_\omega(x, t) \otimes \tilde{u}'_{\omega'}(x', t') \mathcal{G}_{\omega\omega'}(x, t|x', t')]}_{\mathcal{K}(x|x')} \cdot \nabla \langle \theta \rangle(x'), \quad (2.34)$$

$$= - \int dx' \mathcal{K}(x|x') \cdot \nabla \langle \theta \rangle(x'). \quad (2.35)$$

The independence of \mathcal{K} with respect to t follows from the time-independence of $\langle \tilde{u}'\theta' \rangle$ and $\langle \theta \rangle$. In total, we see

$$\langle \tilde{u}\theta \rangle = \langle \tilde{u} \rangle \langle \theta \rangle - \int dx' \mathcal{K}(x|x') \cdot \nabla \langle \theta \rangle(x'). \quad (2.36)$$

An insight from (2.36) is the dependence of turbulent fluxes $\langle \tilde{u}'\theta' \rangle$ at location x as a weighted sum of gradients of the mean variable $\langle \theta \rangle$ at locations x' . The operator is linear and amenable to computation, even in turbulent flows (Bhamidipati, Souza & Flierl 2020).

We consider the spectrum for the turbulent diffusivity operator $\int dx' \mathcal{K}(x|x') \bullet$ as a characterization of turbulent mixing by the flow field $\tilde{u}_\omega(x, t)$. We comment that the operator $\int dx' \mathcal{K}(x|x') \bullet$ is a mapping from vector fields to vector fields, whereas the kernel $\mathcal{K}(x|x')$ is a mapping from two positions to a tensor.

For example, consider a 1-D problem in a periodic domain $x \in [0, 2\pi)$. If $\mathcal{K}(x|x') = \kappa_e \delta(x - x')$ for some positive constant κ_e , the spectrum of the operator is flat and turbulent mixing remains the same on every length scale. If $\mathcal{K}(x|x') = -\kappa_e \partial_{xx}^2 \delta(x - x')$, then the rate of mixing increases with increasing wavenumber and one gets hyperdiffusion. Lastly, if $\int dx' \mathcal{K}(x|x') \bullet = (\kappa_e - \partial_{xx})^{-1}$, then the kernel is non-local and the rate of mixing decreases at smaller length scales.

In the following section, we calculate $\int dx' \mathcal{K}(x|x') \bullet$ directly from the conditional equations, discuss the general structure in § 4 and then apply the methodology to a wandering wave in a channel in § 5.

3. Examples

We now go through two examples to understand the implications of (2.13) and (2.14). The two examples illustrate different aspects of the conditional averaging procedure. The first aspect is the ability to approximate continuous stochastic processes as one with a finite state space, i.e. a Markov jump process. The second aspect is to obtain closed-form formulae for a Markov jump process. In both cases, we use the methodology to calculate turbulent diffusivities for statistically steady states.

The first example is the advection of a passive tracer in a 1-D periodic domain by an Ornstein–Uhlenbeck process: perhaps the simplest class of problems amenable to detailed analysis. See Pappalettera (2022) for a mathematical treatment of this problem without source terms. The second example further builds intuition for the statistical significance of the conditionally averaging procedure by examining a three-velocity state

approximation to an Ornstein–Uhlenbeck in an abstract n -dimensional setting and a concrete two-dimensional (2-D) setting.

3.1. Ornstein–Uhlenbeck process in one dimension

Consider the advection of a 1-D tracer by an Ornstein–Uhlenbeck (OU) process in a 2π periodic domain. The equations of motion are

$$d\tilde{u}_\omega = -\tilde{u}_\omega dt + \sqrt{2} dW_\omega, \tag{3.1}$$

$$\partial_t \theta_\omega = -\tilde{u}_\omega \partial_x \theta_\omega + \kappa \partial_{xx} \theta_\omega + s(x), \tag{3.2}$$

where s is a mean zero source, ω labels the ensemble member and we choose $\kappa = 0.01$ for the diffusivity constant. We shall calculate the turbulent diffusivity operator for this system in two different ways. The first will be by simulating the equations and the second will be by using numerical approximations of the conditional mean equations.

Since the flow field u is independent of the spatial variable and the domain is periodic, we decompose the tracer equation using Fourier modes and calculate a diffusivity as a function of wavenumber k , wavenumber by wavenumber. Decomposing θ_ω as

$$\theta_\omega(x, t) = \sum_{n=-\infty}^{\infty} \hat{\theta}_\omega^n(t) e^{ik_n x} \quad \text{and} \quad s(x) = \sum_{n=-\infty}^{\infty} \hat{s}^n e^{ik_n x}, \tag{3.3a,b}$$

where $k_n = n$, yields

$$d\tilde{u}_\omega = -\tilde{u}_\omega dt + \sqrt{2} dW_\omega, \tag{3.4}$$

$$\frac{d}{dt} \hat{\theta}_\omega^n = -\tilde{u}_\omega i k_n \hat{\theta}_\omega^n - \kappa k_n^2 \hat{\theta}_\omega^n + \hat{s}^n, \tag{3.5}$$

which are ordinary differential equations for each wavenumber k_n with no coupling between wavenumbers. We define a ‘turbulent diffusivity’ on a wavenumber by wavenumber basis with

$$\kappa_T(k_n) = -\frac{\langle \tilde{u} \hat{\theta}^n \rangle}{\langle i k_n \hat{\theta}^n \rangle}, \tag{3.6}$$

which is a flux divided by a gradient. A local diffusivity would be independent of wavenumber. Here the averaging operator, $\langle \cdot \rangle$, is an average over all ensemble members. This quantity is independent of the source wavenumber amplitude, \hat{s}^n , for non-zero sources and conventions for scaling the Fourier transform. With this observation in place, we take our source term to be $s(x) = \delta(x) - 1$ so that $\hat{s}^n = 1$ for all $n \neq 0$ and $\hat{s}^0 = 0$. We numerically simulate (3.1) and (3.2) in spectral space until a time $t = 25$ using 10^6 ensemble members and then perform the ensemble average at this fixed time. We have finished the description for the first method of calculation. We now move on to using the conditional mean equations directly.

The conditional mean equations recover the same result as the continuous ensemble mean. We show this through numerical discretization. The conditional mean equation

version of (3.1) and (3.2) is

$$\partial_t \mathcal{P} = \partial_u (-\mathcal{P} + \partial_u \mathcal{P}), \tag{3.7}$$

$$\partial_t \Theta = -u \partial_x \Theta + \kappa \partial_{xx} \Theta + \mathcal{P} s + \partial_u (-\Theta + \partial_u \Theta), \tag{3.8}$$

where $\Theta = \Theta(x, u, t)$ and $\mathcal{P} = \mathcal{P}(u, t)$. We make the observation

$$\int_{-\infty}^{\infty} du \Theta = \langle \theta \rangle \quad \text{and} \quad \int_{-\infty}^{\infty} duu \Theta = \langle \tilde{u} \theta \rangle. \tag{3.9a,b}$$

To discretize the u variable, we invoke a finite volume discretization by integrating with respect to $N = N' + 1$ control volumes $\Omega_m = [(2m - 1 - N')/\sqrt{N'}, (2m + 1 - N')/\sqrt{N'}]$ for $m = 0, \dots, N'$, a procedure which is detailed in [Appendix C.2](#). The result is

$$\partial_t \mathcal{P}_m = \sum_{m'} \mathcal{Q}_{mm'} \mathcal{P}_{m'}, \tag{3.10}$$

$$\partial_t \Theta_m = -u_m \partial_x \hat{\Theta}_m + \kappa \partial_{xx} \hat{\Theta}_m^n + \mathcal{P}_m s + \sum_{m'} \mathcal{Q}_{mm'} \Theta_{m'}, \tag{3.11}$$

where

$$\int_{\Omega_m} du \Theta = \Theta_m, \quad \int_{\Omega_m} du \mathcal{P} = \mathcal{P}_m, \quad u_m = \frac{2}{\sqrt{N'}}(m - N'/2), \quad \int_{\Omega_m} duu \Theta \approx u_m \Theta_m, \tag{3.12a-c}$$

and $\mathcal{Q}_{mm'} = (-N' \delta_{mm'} + m' \delta_{(m+1)m'} + (N' - m') \delta_{(m-1)m'})/2$ and $m = 0, \dots, N'$. This approximation is the same as that used by Hagan *et al.* (1989) to represent the Ornstein–Uhlenbeck process by Markov jump processes, but now justified as a finite volume discretization. Observe that for $N = 2$, we have a dichotomous velocity process

$$\partial_t \mathcal{P}_1 = -0.5 \mathcal{P}_1 + 0.5 \mathcal{P}_2, \tag{3.13}$$

$$\partial_t \mathcal{P}_2 = -0.5 \mathcal{P}_2 + 0.5 \mathcal{P}_1, \tag{3.14}$$

$$\partial_t \Theta_1 = -\partial_x \hat{\Theta}_1 + \kappa \partial_{xx} \Theta_1 + \mathcal{P}_1 s - 0.5 \Theta_1 + 0.5 \Theta_2, \tag{3.15}$$

$$\partial_t \Theta_2 = \partial_x \hat{\Theta}_2 + \kappa \partial_{xx} \Theta_1 + \mathcal{P}_2 s - 0.5 \Theta_2 + 0.5 \Theta_1 \tag{3.16}$$

and for $N = 3$, we have a generalization

$$\partial_t \mathcal{P}_1 = -\mathcal{P}_1 + 0.5 \mathcal{P}_2, \tag{3.17}$$

$$\partial_t \mathcal{P}_2 = -\mathcal{P}_2 + \mathcal{P}_1 + \mathcal{P}_3, \tag{3.18}$$

$$\partial_t \mathcal{P}_3 = -\mathcal{P}_3 + 0.5 \mathcal{P}_2, \tag{3.19}$$

$$\partial_t \Theta_1 = -\sqrt{2} \partial_x \hat{\Theta}_1 + \kappa \partial_{xx} \Theta_1 + \mathcal{P}_1 s - \Theta_1 + 0.5 \Theta_2, \tag{3.20}$$

$$\partial_t \Theta_2 = 0 \partial_x \hat{\Theta}_2 + \kappa \partial_{xx} \Theta_1 + \mathcal{P}_2 s - \Theta_2 + \Theta_1 + \Theta_3, \tag{3.21}$$

$$\partial_t \Theta_3 = \sqrt{2} \partial_x \hat{\Theta}_3 + \kappa \partial_{xx} \Theta_1 + \mathcal{P}_3 s - \Theta_3 + 0.5 \Theta_2, \tag{3.22}$$

similar to the models of Davis *et al.* (1991) and Ferrari, Manfroi & Young (2001). The presence of an evolution equation for \mathcal{P}_n accounts for statistically non-stationary states of the flow field. Alternative discretizations are possible. For example, using

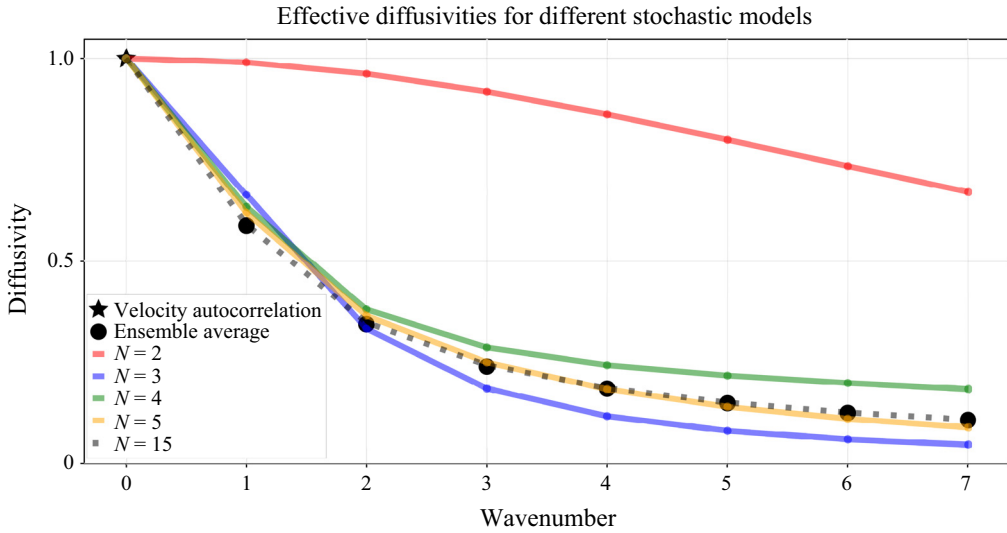


Figure 1. Wavenumber diffusivities. We show the turbulent diffusivity estimate as a function of wavenumber. Here we see that different wavenumbers, or equivalently length scales, produce different estimates of the turbulent diffusivity. Furthermore, we show that the different N -state approximations yield an increasingly better approximation to the Ornstein–Uhlenbeck empirical estimate.

a spectral discretization for the u variable via Hermite polynomials converges to the continuous system at a faster rate; however, the resulting discrete system no longer has an interpretation as a Markov jump process for three or more state variables.

Next, taking the Fourier transform in x yields

$$\frac{d}{dt} \hat{\Theta}_m^n = -u_m i k_n \hat{\Theta}_m^n - \kappa k_n^2 \hat{\Theta}_m^n + \sum_{m'} Q_{mm'} \hat{\Theta}_{m'}^n. \quad (3.23)$$

We solve for the steady-state solution

$$0 = \sum_{m'} Q_{mm'} \mathcal{P}_{m'}, \quad (3.24)$$

$$0 = -u_m i k_n \hat{\Theta}_m^n - \kappa k_n^2 \hat{\Theta}_m^n + \mathcal{P}_m \hat{\delta}^n + \sum_{m'} Q_{mm'} \hat{\Theta}_{m'}^n \quad (3.25)$$

and then compute

$$\kappa_E(k_n) = -\frac{\sum_m u_m \hat{\Theta}_m^n}{\sum_m i k_n \hat{\Theta}_m^n} \Rightarrow -\frac{\int_{-\infty}^{\infty} du \hat{\Theta}^n u}{\int_{-\infty}^{\infty} du (i k_n \hat{\Theta}^n)} \quad (3.26)$$

for various choices of discrete states. The calculations in the numerator and denominator are equivalent to ensemble averages.

We compare the two methods of calculation in [figure 1](#). We see that increasing the number of states from $N = 2$ to $N = 15$ yields, for each wavenumber, ensemble averages similar to those of the Ornstein–Uhlenbeck process. The explicit formula for $N = 3$ is given in the next section. We stop at $N = 15$ states since all wavenumbers

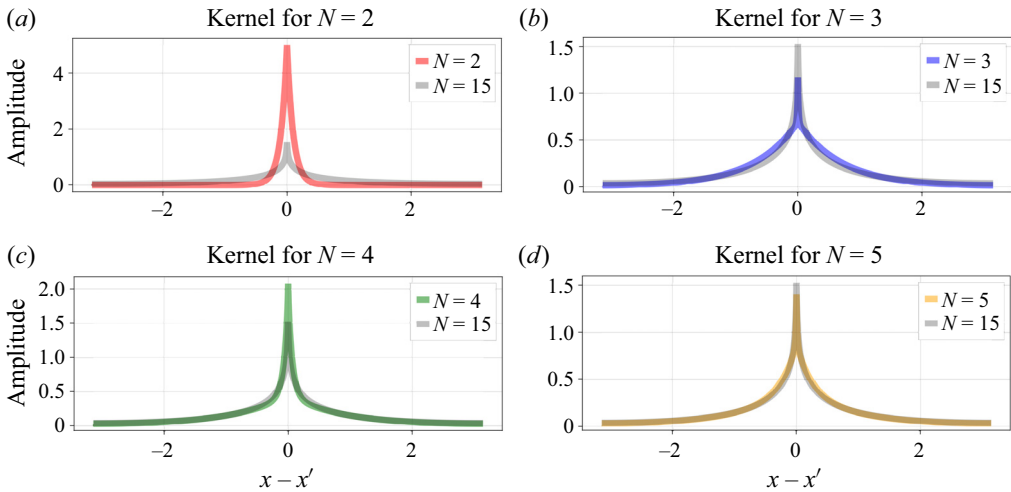


Figure 2. Kernels for N -state systems. Here we show the turbulent diffusivity kernels for several approximations of the Ornstein–Uhlenbeck diffusivity kernel. We compare all kernels to the $N = 15$ state kernel, which is a good approximation to the OU process kernel. We see that the width of the kernel is comparable to the domain. The non-locality of the kernel is a consequence of the turbulent diffusivity estimate differing on a wavenumber per wavenumber basis.

in the plot are within 1% of one another. The $n = 0$ wavenumber is plotted for convenience as well and, in the case of the present process, also happens to correspond to the velocity autocorrelation. The correspondence of all velocity states (including the Ornstein–Uhlenbeck process) at wavenumber $k_n = 0$ is particular to this system and is not a feature that holds in general. The ‘large scale limit’, corresponding to wavenumber $k_n = 0$, can often be considered an appropriate local diffusivity definition.

Although we use the finite volume discretization of velocity states as approximations to the Ornstein–Uhlenbeck process, they also constitute a realizable stochastic process in the form of a Markov jump process. Thus, each finite state case can be simulated from whence the turbulent diffusivity estimate in figure 1 is exact rather than an approximation. We choose a particular example with $N = 3$ in the following section.

Furthermore, each wavenumber k_n yields a different estimate of the turbulent diffusivity. In particular, the decrease of turbulent diffusivity as a function of wavenumber implies non-locality. Choosing a forcing that is purely sinusoidal of a given mode would yield different estimates of turbulent fluxes. For example, forcing with $s(x) = \sin(x)$ would yield a diffusivity estimate of $\kappa_T(1) \approx 0.58$, whereas forcing with $s(x) = \sin(2x)$ would yield a diffusivity estimate of $\kappa_T(2) \approx 0.34$, as implied by figure 1.

In the present case, the inverse Fourier transform of the turbulent diffusivity yields the turbulent diffusivity operator in real space. Since products in wavenumber space are circular convolutions in real space, we naturally guarantee the translation invariance of the kernel for the present case. The turbulent diffusivity kernels are shown in figure 2. Thus, the equation for the ensemble mean is

$$-\partial_x (\mathcal{K} * \partial_x \langle \theta \rangle) + \kappa \partial_x \langle \theta \rangle = s(x), \tag{3.27}$$

where $*$ is a circular convolution with the kernel in figure 2.

To illustrate the action of a non-local operator, we show the implied flux for a gradient that is presumed to have the functional form

$$\partial_x \langle \theta \rangle = \exp(-2(x - \pi)^2) - \exp(-10(x - 9\pi/8)^2). \tag{3.28}$$

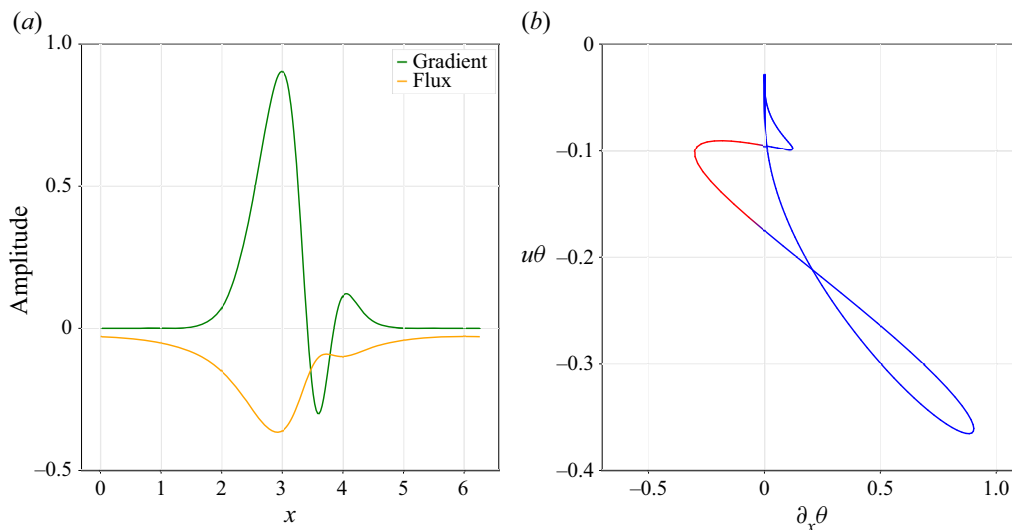


Figure 3. Flux prediction from the kernel. Given the ensemble mean gradient (green), we convolve with the $N = 15$ state kernel to get the flux (yellow) in panel (a). We see that the gradient changes sign while the flux always remains negative. To further illustrate the effect of non-locality, we show flux versus gradient in panel (b). The blue regions are down-gradient and the red regions are up-gradient. A local diffusivity estimate would be a straight line through the origin, whose slope determines the diffusivity constant.

The flux computed from the convolution with the $N = 15$ state kernel

$$\langle \tilde{u}\theta \rangle = -\mathcal{K} * \partial_x\langle \theta \rangle \tag{3.29}$$

is shown in figure 3. Even though the gradient changes sign, we see that the flux remains purely negative, i.e. the flux is up-gradient at particular points of the domain. Furthermore, the figure 3(b) does not resemble a straight line through the origin, as would be the case for a local diffusivity. The red portion highlights the ‘up-gradient’ part of the flux-gradient relation. We point out these features to illustrate the limitations of a purely local flux-gradient relationship.

This section focused on a flow field with a simple spatial and temporal structure. In the next section, we derive the turbulent diffusivity operator for a flow with arbitrary spatial structure but with a simple statistical temporal structure: a three-state velocity field whose amplitude is an approximation to the Ornstein–Uhlenbeck process.

3.2. A generalized three-state system with a 2-D example

For this example, we consider a Markov process that transitions between three incompressible states $\mathbf{u}_1(\mathbf{x}) = \mathbf{u}(\mathbf{x})$, $\mathbf{u}_2(\mathbf{x}) = 0$ and $\mathbf{u}_3(\mathbf{x}) = -\mathbf{u}(\mathbf{x})$. We use the three-state approximation to the Ornstein–Uhlenbeck process, in which case we let the generator be

$$\mathcal{Q} = \gamma \begin{bmatrix} -1 & 1/2 & 0 \\ 1 & -1 & 1 \\ 0 & 1/2 & -1 \end{bmatrix}, \tag{3.30}$$

where $\gamma > 0$. This generator implies that the flow field stays in each state for an exponentially distributed amount of time, with each state’s rate parameter γ . Flow fields \mathbf{u}_1 and \mathbf{u}_3 always transition to state \mathbf{u}_2 and \mathbf{u}_2 transitions to either \mathbf{u}_1 or \mathbf{u}_3 with

50 % probability. This stochastic process is a generalization of the model considered in the previous section and the class of problems considered by Ferrari *et al.* (2001) since the flow field can have spatial structure and is not limited to one dimension.

The three-state system is viewed as a reduced-order statistical model for any flow field that randomly switches between clockwise or counterclockwise advection, such as the numerical and experimental flows of Sugiyama *et al.* (2010). In their work, the numerical computations of 2-D Rayleigh–Bénard convection in a rectangular domain with no-slip boundary conditions along all walls yielded large-scale convection rolls that flipped orientation at random time intervals upon particular choices of Rayleigh and Prandtl numbers. In addition, they showed that the phenomenology could be observed in a laboratory experiment with a setup similar to that of Xia, Sun & Zhou (2003). In general, we expect reducing the statistics of a flow to a continuous-time Markov process with discrete states to be useful and lead to tractable analysis in fluid flows characterized by distinct well-separated flow regimes with seemingly random transitions between behaviours. Consequently, a similar construction is possible for more complex three-dimensional flow fields, such as those of Brown & Ahlers (2007), or geophysical applications (Zhang, Zhang & Tian 2022).

Proceeding with calculations, we note that the eigenvectors of the generator are

$$\mathbf{v}^1 = \begin{bmatrix} 1/4 \\ 1/2 \\ 1/4 \end{bmatrix}, \quad \mathbf{v}^2 = \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \end{bmatrix} \quad \text{and} \quad \mathbf{v}^3 = \begin{bmatrix} 1/4 \\ -1/2 \\ 1/4 \end{bmatrix} \quad (3.31a-c)$$

with respective eigenvalues $\lambda^1 = 0$, $\lambda^2 = -\gamma$ and $\lambda^3 = -2\gamma$.

The statistically steady three-state manifestation of (2.13) and (2.14) is

$$\nabla \cdot (\mathbf{u}\Theta_1) = \kappa \Delta \Theta_1 + s(\mathbf{x})/4 - \gamma \Theta_1 + \gamma \Theta_2/2, \quad (3.32)$$

$$0 = \kappa \Delta \Theta_2 + s(\mathbf{x})/2 - \gamma \Theta_2 + \gamma \Theta_1 + \gamma \Theta_3, \quad (3.33)$$

$$-\nabla \cdot (\mathbf{u}\Theta_3) = \kappa \Delta \Theta_3 + s(\mathbf{x})/4 - \gamma \Theta_3 + \gamma \Theta_2/2. \quad (3.34)$$

We define a transformation using the eigenvectors of the generator \mathcal{Q} ,

$$\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & -1/2 \\ 1/4 & -1/2 & 1/4 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{bmatrix} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \end{bmatrix} \quad (3.35)$$

resulting in the equations

$$\nabla \cdot (\mathbf{u}\varphi_2) = \kappa \Delta \varphi_1 + s(\mathbf{x}), \quad (3.36)$$

$$\frac{1}{2} \mathbf{u} \cdot \nabla \varphi_1 + \frac{1}{2} \mathbf{u} \cdot \nabla \varphi_3 = \kappa \Delta \varphi_2 - \gamma \varphi_2, \quad (3.37)$$

$$\mathbf{u} \cdot \nabla \varphi_2 = \kappa \Delta \varphi_3 - 2\gamma \varphi_3. \quad (3.38)$$

We comment that $\varphi_1 = \langle \theta \rangle$, and that φ_2 and φ_3 are thought of as perturbation variables. Furthermore, the turbulent flux is $\langle \tilde{\mathbf{u}}'\theta' \rangle = \mathbf{u}\varphi_2$. We eliminate dependence on the

perturbation variables φ_2 and φ_3 by first solving for φ_3 in terms of φ_2 ,

$$\varphi_3 = (\kappa \Delta - 2\gamma)^{-1} \mathbf{u} \cdot \nabla \varphi_2, \quad (3.39)$$

and then solving for φ_2 in terms of φ_1 ,

$$\varphi_2 = \left(\kappa \Delta - \gamma - \frac{1}{2} \mathbf{u} \cdot \nabla (\kappa \Delta - 2\gamma)^{-1} \mathbf{u} \cdot \nabla \right)^{-1} \frac{1}{2} \mathbf{u} \cdot \nabla \varphi_1. \quad (3.40)$$

And finally, we write our equation for the ensemble mean as

$$\nabla \cdot \left(\mathbf{u} \left(\kappa \Delta - \gamma - \frac{1}{2} \mathbf{u} \cdot \nabla (\kappa \Delta - 2\gamma)^{-1} \mathbf{u} \cdot \nabla \right)^{-1} \frac{1}{2} \mathbf{u} \cdot \nabla \langle \theta \rangle \right) = \kappa \Delta \langle \theta \rangle + s(\mathbf{x}) \quad (3.41)$$

from whence we extract the turbulent diffusivity operator

$$\int dx' \mathcal{K}(x|x') \bullet = \mathbf{u} \left(\gamma - \kappa \Delta + \frac{1}{2} \mathbf{u} \cdot \nabla (\kappa \Delta - 2\gamma)^{-1} \mathbf{u} \cdot \nabla \right)^{-1} \frac{1}{2} \mathbf{u}. \quad (3.42)$$

We point out a few salient features of (3.42).

For large κ , we have

$$\int dx' \mathcal{K}(x|x') \bullet \rightarrow \mathbf{u} (\gamma - \kappa \Delta)^{-1} \frac{1}{2} \mathbf{u}. \quad (3.43)$$

The inverse Helmholtz operator, $(\gamma - \kappa \Delta)^{-1}$, damps high spatial frequency components of ensemble mean gradients. Thus, the operator's eigenvalues decrease as one examines increasingly fine-scale structure. Intuitively, as one examines a small-scale structure, the presence of diffusivity leads to lower turbulent fluxes, expressing that it is challenging to transport something that immediately diffuses.

A second observation pertains to the presence of the eigenvalue of the generator in the operator. If the flow field changes rapidly, transitioning between the disparate states, then γ is large and one can expect the turbulent diffusivity to be local. In other words, the flow does not stay sufficiently long time near a coherent structure. More concretely, larger transition rates correspond to an enhanced locality in the turbulent diffusivity operator since, as $\gamma \rightarrow \infty$,

$$\int dx' \mathcal{K}(x|x') \bullet \rightarrow \frac{1}{2\gamma} \mathbf{u} \otimes \mathbf{u}. \quad (3.44)$$

Observe that (3.44) is the integrated autocorrelation of the stochastic flow field. Whether or not non-local effects matter depends on the characteristic time scale of the flow field, LU^{-1} as well as the characteristic time scale for diffusion L^2/κ , as it compares to the characteristic time scale for transitioning between states, γ^{-1} .

We see that the $\kappa \rightarrow 0$ limit retains a non-local feature since

$$\int dx' \mathcal{K}(x|x') \bullet \rightarrow \mathbf{u} \left(\gamma - \frac{1}{4\gamma} (\mathbf{u} \cdot \nabla)(\mathbf{u} \cdot \nabla) \right)^{-1} \frac{1}{2} \mathbf{u} \quad (3.45)$$

and the operator $(\mathbf{u} \cdot \nabla)(\mathbf{u} \cdot \nabla)$ can have a significant spatial structure. To further understand this point, we simplify by using the example from the previous section.

Using the 1-D example in Fourier space and using the usual transcriptions $\Delta \rightarrow -k^2$, $\nabla \rightarrow ik$, and with $u > 0$, the kernel is

$$\frac{u^2}{2(\gamma + \kappa k^2) + \frac{u^2 k^2}{(\kappa k^2 + 2\gamma)}} \rightarrow u \frac{\frac{2\gamma}{u}}{\frac{4\gamma^2}{u^2} + k^2} \Rightarrow \mathcal{K}(x, x') = \frac{u}{2} \exp\left(-\frac{2\gamma}{u} |x - x'|\right), \tag{3.46}$$

where in the last line, we used the inverse Fourier transform on the real line to simplify the structure of the kernel. (Accounting for finite domain effects yields the Fourier series of a hyperbolic cosine.) The left-most expression corresponds to the blue dots in [figure 1](#) for $N = 3$, $u = \sqrt{2}$, $\gamma = 1$ and $\kappa = 0.01$ for each wavenumber $k \in \{0, 1, 2, \dots, 7\}$. In the right-most expression, we see the mean free path $\ell \equiv u/(2\gamma)$ as the characteristic width of the non-local kernel. Observe that by rescaling $u \rightarrow \sqrt{2\gamma}u$ and taking the limit as $\gamma \rightarrow \infty$ yields a delta function, a point that we come back to in [§ 4.3](#) and [Appendix A](#). Furthermore, the weak velocity amplitude limit $u \rightarrow 0$ is also local.

To further explicate the conditional average procedure, we compare empirical conditional averages of [\(2.10\)](#) and [\(2.11\)](#) (specialized to the present system) to the direct conditionally averaged equations [\(3.36\)–\(3.38\)](#). Concretely, we use the flow field and source term

$$\mathbf{u}(x, y) = -\sqrt{2} \cos(x) \sin(y) \hat{\mathbf{x}} + \sqrt{2} \sin(x) \cos(y) \hat{\mathbf{y}}, \tag{3.47}$$

$$s(x, y) = \sin(x) \sin(y) \tag{3.48}$$

in a 2π periodic domain and $\gamma = 1$ for the transition rate. We employ a 2-D pseudo-spectral Fourier discretization with 48×48 grid points and evolve both systems to a time $t = 25$, which suffices for the flow and tracer to ‘forget’ its initial condition and reach a statistically steady state. For the stochastic differential equation, we use 10 000 ensemble members. We emphasize that the turbulent diffusivity operator is independent of the source term, and we make a choice for illustrative purposes.

[Figure 4](#) summarizes the result of comparing two different calculation methods. To compute the conditional averages, for example, Θ_1 at time $t = 25$, we proceed as follows. At the time $t = 25$, we tag all fields that are currently being advected by \mathbf{u}_1 , add them up and then divide by the number of ensemble members, which in the present case is 10 000. This sequence of calculations yields Θ_1 , shown in [figure 4\(a\)](#). The conditional means Θ_2 and Θ_3 are obtained similarly. We then calculate the conditional mean equations directly in the bottom row. All fields are plotted with respect to the same colour scale. For convenience, we also show the ensemble average, which here is $\langle \theta \rangle = \varphi_1 = \Theta_1 + \Theta_2 + \Theta_3$. The empirical averages have a slight asymmetry due to finite sampling effects. For reference, the source and stream function of the flow field are shown in the last column of the figure.

When the tracers are being advected by flow field \mathbf{u}_1 , we see that the red in the stream function corresponds to a counterclockwise flow and the blue to a clockwise flow. The negative source concentration corresponding to the top left and bottom right parts of the domain are then transported vertically, whereas the positive source, located in the top right and bottom left of the domain, is transported horizontally. When advected by flow field \mathbf{u}_3 , the clockwise and counterclockwise roles reverse, and positive tracer sources are transported vertically and negative tracer sources are transported horizontally. When advected by $\mathbf{u}_2 = 0$, the conditional averages reflect the source term closely.

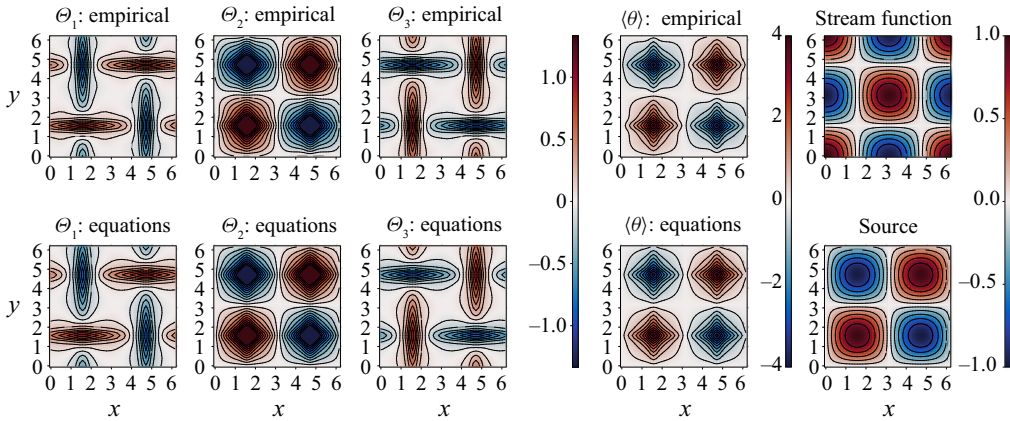


Figure 4. Three-state system. Here, we show the stream function, source term, conditional averages and the ensemble mean obtained using two different calculation methods. The first calculation method uses an empirical average with 10 000 ensemble members, and the second method uses the equations from the text. Both are shown at a final time of $T = 25$ time units where the statistical steady state has been reached.

Now that we have presented two examples, in the next section, we generalize by expressing the turbulent diffusivity operator in terms of the spectrum of the generator.

4. General approach

The previous example had the following pattern:

- (i) compute the eigenvectors of the generator Q ;
- (ii) transform the equations into a basis that diagonalizes Q ;
- (iii) separate the mean equation from the perturbation equations;
- (iv) solve for the perturbation variables in terms of the mean variable.

Here we aim to gather the above procedure in the general case where we have access to the eigenvectors of Q . Furthermore, in the last example, we claimed that the local turbulent diffusivity approximation, as calculated by neglecting the effects of diffusion and perturbation gradients, is equivalent to calculating the integrated auto-correlation of the Markov process. We justify that claim in § 4.3.

4.1. Notation

Let us establish a notation for the general procedure. We again let Q denote the generator with corresponding transition probability matrix $\mathcal{P}(\tau)$ given by the matrix exponential

$$\mathcal{P}(\tau) = \exp(\tau Q). \tag{4.1}$$

The entries of the matrix $[\mathcal{P}(\tau)]_{mn}$ denote the transition probability of state n to the state m . In each column of the transition matrix, the sum of the entries is one. We assume a unique zero eigenvalue for Q with all other eigenvalues negative. We also assume that the eigenvalues can be ordered in such a way that they are decreasing, i.e. $\lambda_1 = 0$, $\lambda_2 < 0$ and $\lambda_i \leq \lambda_j$ for $i > j$ with $j \geq 2$. These choices result in a unique statistical steady state which

we denote by the vector \mathbf{v}_1 with the property

$$\mathcal{Q}\mathbf{v}_1 = 0\mathbf{v}_1 \quad \text{and} \quad \mathcal{P}(\tau)\mathbf{v}_1 = \mathbf{v}_1 \quad \text{for all } \tau, \tag{4.2}$$

and similarly for the left eigenvector, \mathbf{w}_1 . We denote the entries of \mathbf{v}_1 and \mathbf{w}_1 by column vectors

$$\mathbf{v}_1 = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_M \end{bmatrix} \quad \text{and} \quad \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{4.3a,b}$$

where M is the number of states. We assume that the eigenvector \mathbf{v}_1 is normalized such that $\sum_m P_m = 1$. Consequently, $\mathbf{w}_1 \cdot \mathbf{v}_1 = \mathbf{w}_1^T \mathbf{v}_1 = 1$. We introduce unit vectors $\hat{\mathbf{e}}_m$ whose m th entry is zero and all other entries are zero, e.g.

$$\hat{\mathbf{e}}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \hat{\mathbf{e}}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{e}}_M = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{4.4a-c}$$

Thus, $\mathbf{v}_1 = \sum_m P_m \hat{\mathbf{e}}_m$, $\mathbf{w}_1 = \hat{\mathbf{e}}_m$. Furthermore, $\hat{\mathbf{e}}_m \cdot \mathbf{v}_1 = P_m$ and $\hat{\mathbf{e}}_m \cdot \mathbf{w}_1 = 1$ for each m .

For the discussion that follows, we will assume that the matrix \mathcal{Q} has an eigenvalue decomposition. In general, we denote the right eigenvectors of \mathcal{Q} by \mathbf{v}_i for $i = 1, \dots, M$ and the left eigenvectors by \mathbf{w}_i for $i = 1, \dots, M$. These vectors are all associated with eigenvalues λ_i for $i = 1, \dots, M$, where $i = 1$ denotes the unique eigenvalue $\lambda_1 = 0$. We recall that the left eigenvectors can be constructed from the right eigenvectors by placing the right eigenvectors in the columns of a matrix V , computing the inverse V^{-1} and extracting the rows of the inverse. The aforementioned procedure guarantees the normalization $\mathbf{w}_j \cdot \mathbf{v}_i = \mathbf{w}_j^T \mathbf{v}_i = \delta_{ij}$. Thus, we have the relations

$$\mathcal{Q}\mathbf{v}_n = \lambda_n \mathbf{v}_n \quad \text{and} \quad \mathbf{w}_n^T \mathcal{Q} = \lambda_n \mathbf{w}_n^T. \tag{4.5a,b}$$

With notation now in place, we observe that the operators \mathcal{Q} and $\mathcal{P}(\tau)$ are characterized by their spectral decomposition

$$\mathcal{Q} = \sum_i \lambda_i \mathbf{v}_i \mathbf{w}_i^T \quad \text{and} \quad \mathcal{P}(\tau) = \sum_i e^{\tau \lambda_i} \mathbf{v}_i \mathbf{w}_i^T. \tag{4.6a,b}$$

We now introduce our Markov states as steady vector fields. The use of several vector spaces imposes a burden on notation: the vector spaces associated with Markov states, ensemble members and the vector field \mathbf{u} . Instead of using overly decorated notation with an excessive number of indices, we introduce the convention that \mathbf{u} will always belong to the vector space associated with the vector field, and all other vectors are associated with the vector space of Markov states. For example, if we have two flow states \mathbf{u}_1 and \mathbf{u}_2 , then

$$\begin{bmatrix} \mathbf{u}_1(x) & \mathbf{u}_2(x) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1(x) \\ \mathbf{u}_2(x) \end{bmatrix} = \mathbf{u}_1(x) \otimes \mathbf{u}_1(x) + \mathbf{u}_2(x) \otimes \mathbf{u}_2(x) \tag{4.7}$$

$$\begin{bmatrix} \mathbf{u}_1(x) \\ \mathbf{u}_2(x) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1(x) & \mathbf{u}_2(x) \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1(x) \otimes \mathbf{u}_1(x) & \mathbf{u}_1(x) \otimes \mathbf{u}_2(x) \\ \mathbf{u}_2(x) \otimes \mathbf{u}_1(x) & \mathbf{u}_2(x) \otimes \mathbf{u}_2(x) \end{bmatrix} \tag{4.8}$$

4.2. Spectral representation

With this notation now in place, the statistically steady equations

$$\nabla \cdot (\mathbf{u}_m \Theta_m) = \kappa \Delta \Theta_m + P_m s + \sum_n \mathcal{Q}_{mn} \Theta_n \tag{4.9}$$

are represented as the matrix system

$$\sum_m \hat{\mathbf{e}}_m \nabla \cdot (\mathbf{u}_m \Theta_m) = \sum_m \hat{\mathbf{e}}_m \kappa \Delta \Theta_m + s \mathbf{v}_1 + \mathcal{Q} \left(\sum_m \hat{\mathbf{e}}_m \Theta_m \right), \tag{4.10}$$

where we made use of $\sum_m \hat{\mathbf{e}}_m P_m = \mathbf{v}_1$. We now re-express (4.10) in terms of a basis that uses the eigenvectors of the transition matrix. Define components φ_n by the change of basis formula

$$\sum_m \Theta_m \hat{\mathbf{e}}_m = \sum_n \varphi_n \mathbf{v}_n \Leftrightarrow \sum_n \varphi_n \hat{\mathbf{e}}_n = \sum_{mn} (\mathbf{w}_n \cdot \hat{\mathbf{e}}_m) \hat{\mathbf{e}}_n \Theta_m. \tag{4.11}$$

We make the observation $\varphi_1 = \langle \theta \rangle$. We have the following relations based on the general definitions of the left eigenvectors \mathbf{w}_n and right eigenvectors \mathbf{v}_n ,

$$\Theta_n = \sum_i (\hat{\mathbf{e}}_n \cdot \mathbf{v}_i) \varphi_i \quad \text{and} \quad \varphi_n = \sum_m (\mathbf{w}_n \cdot \hat{\mathbf{e}}_m) \Theta_m. \tag{4.12a,b}$$

Multiplying (4.10) by \mathbf{w}_j^T and making use of (4.12a,b), we get

$$\sum_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla \cdot (\mathbf{u}_n \Theta_n) = \kappa \Delta \varphi_j + \delta_{1j} s + \lambda_j \varphi_j \tag{4.13}$$

\Rightarrow

$$\sum_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla \cdot \left(\mathbf{u}_n \left[\sum_i \hat{\mathbf{e}}_n \cdot \mathbf{v}_i \varphi_i \right] \right) = \kappa \Delta \varphi_j + \delta_{1j} s + \lambda_j \varphi_j \tag{4.14}$$

\Rightarrow

$$\sum_{in} (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) (\hat{\mathbf{e}}_n \cdot \mathbf{v}_i) \nabla \cdot (\mathbf{u}_n \varphi_i) = \kappa \Delta \varphi_j + \delta_{1j} s + \lambda_j \varphi_j. \tag{4.15}$$

We now wish to decompose (4.14) into a mean equation, index $j = 1$, and perturbation equations, $j > 1$. For the mean equation, index $j = 1$, we make use of the properties

$$\lambda_1 = 0, \quad \mathbf{w}_1 \cdot \hat{\mathbf{e}}_n = 1, \quad \mathbf{v}_1 \cdot \hat{\mathbf{e}}_n = P_n, \quad \text{and} \quad \sum_n (\mathbf{w}_1 \cdot \hat{\mathbf{e}}_n) (\hat{\mathbf{e}}_n \cdot \mathbf{v}_1) \nabla \cdot (\mathbf{u}_n \varphi_1) = \nabla \cdot (\langle \tilde{\mathbf{u}} \rangle \varphi_1) \tag{4.16}$$

to arrive at (after changing summation index from n to m)

$$\nabla \cdot (\langle \mathbf{u} \rangle \varphi_1) + \nabla \cdot \left[\sum_{(i \neq 1)m} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m \varphi_i \right] = \kappa \Delta \varphi_1 + s. \tag{4.17}$$

We make the observation that the turbulent flux is

$$\langle \tilde{\mathbf{u}}' \theta' \rangle = \sum_{(i \neq 1)m} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m \varphi_i, \tag{4.18}$$

and comment that the flow fields $\mathbf{U}_i(\mathbf{x}) \equiv \sum_m (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m$ are the Koopman mode amplitudes associated with eigenvalue λ_i . These are the relevant statistical spatial

structures that advect the statistical perturbation variables φ_i . This is similar to how the ensemble mean flow field advects the ensemble mean tracer concentration. Stated differently, the perturbation variables are advected by structures associated with non-trivial Koopman modes.

The perturbation equations, indices $j > 1$, are

$$\sum_{in} (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n)(\hat{\mathbf{e}}_n \cdot \mathbf{v}_i) \nabla \cdot (\mathbf{u}_n \varphi_i) = \kappa \Delta \varphi_j + \lambda_j \varphi_j \quad \text{for } j > 1. \quad (4.19)$$

We isolate the dependence on the mean gradients by rearranging the above expression as follows for $j > 1$:

$$\sum_{(i \neq 1)n} (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n)(\hat{\mathbf{e}}_n \cdot \mathbf{v}_i) \nabla \cdot (\mathbf{u}_n \varphi_i) - \kappa \Delta \varphi_j - \lambda_j \varphi_j = - \sum_n P_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla \cdot (\mathbf{u}_n \varphi_1), \quad (4.20)$$

where we used $\hat{\mathbf{e}}_n \cdot \mathbf{v}_1 = P_n$. Assuming that the operator on the left-hand side of (4.20) is invertible, we introduce the Green's function, \mathcal{G}_{ij} , to yield

$$\varphi_i = - \int d\mathbf{x}' \sum_{(j \neq 1)n} \mathcal{G}_{ij}(\mathbf{x}|\mathbf{x}') P_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla_{\mathbf{x}'} \cdot (\mathbf{u}_n(\mathbf{x}') \varphi_1(\mathbf{x}')) \quad \text{for } i \neq 1. \quad (4.21)$$

Thus, we represent our turbulent flux as

$$\langle \tilde{\mathbf{u}}' \theta' \rangle = - \int d\mathbf{x}' \sum_{(i \neq 1)(j \neq 1)mn} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m(\mathbf{x}) \mathcal{G}_{ij}(\mathbf{x}|\mathbf{x}') P_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla_{\mathbf{x}'} \cdot (\mathbf{u}_n(\mathbf{x}') \varphi_1(\mathbf{x}')). \quad (4.22)$$

For compressible flow, the eddy-flux depends on both the ensemble mean gradients and the ensemble mean value; otherwise, when each Markov state is incompressible,

$$\langle \tilde{\mathbf{u}}' \theta' \rangle = - \int d\mathbf{x}' \sum_{(i \neq 1)(j \neq 1)mn} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m(\mathbf{x}) \mathcal{G}_{ij}(\mathbf{x}|\mathbf{x}') P_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \mathbf{u}_n(\mathbf{x}') \cdot \nabla_{\mathbf{x}'} \varphi_1(\mathbf{x}'), \quad (4.23)$$

in which case the turbulent diffusivity kernel is

$$\mathcal{K}(\mathbf{x}|\mathbf{x}') = \sum_{(i \neq 1)(j \neq 1)mn} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m(\mathbf{x}) \mathcal{G}_{ij}(\mathbf{x}|\mathbf{x}') P_n (\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \mathbf{u}_n(\mathbf{x}'). \quad (4.24)$$

The above expression completes the procedure that we enacted for the examples in § 3.

We now discuss local approximations to the turbulent diffusivity operator.

4.3. Local approximation

We start with the same local diffusivity approximation of § 3.2 but using the spectral representation of (2.13) and (2.14). In the perturbation equations, neglect the dissipation operator and perturbation gradients, e.g. only include index $i = 1$, to yield the following

reduction of (4.19),

$$\sum_n (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) P_n (\mathbf{u}_n \cdot \nabla \varphi_1) = \lambda_i \varphi_i \quad \text{for } i > 1, \quad (4.25)$$

where we used $(\hat{\mathbf{e}}_n \cdot \mathbf{v}_1) = P_n$ and have changed indices from j to i . We solve for φ_i for $i > 1$ and focus on the perturbation flux term in (4.17),

$$\langle \tilde{\mathbf{u}}' \theta' \rangle = \sum_{(i \neq 1)m} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m \varphi_i, \quad (4.26)$$

to get the local turbulent diffusivity estimate,

$$\sum_{(i \neq 1)m} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m \varphi_i = \sum_{(i \neq 1)m} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) \mathbf{u}_m \left[\frac{1}{\lambda_i} \sum_n (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) P_n (\mathbf{u}_n \cdot \nabla \varphi_1) \right] \quad (4.27)$$

$$= \underbrace{\left[\sum_{(i \neq 1)mn} \frac{-1}{\lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n \right]}_D \cdot (-\nabla \varphi_1). \quad (4.28)$$

We aim to show that the turbulent diffusivity from (4.28),

$$D = \sum_{(i \neq 1)mn} \frac{-1}{\lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n, \quad (4.29)$$

is equivalent to estimating the diffusivity by calculating the integral of the velocity perturbation autocorrelation in a statistically steady state,

$$D = \int_0^\infty \langle \tilde{\mathbf{u}}'(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}'(\mathbf{x}, t) \rangle d\tau. \quad (4.30)$$

The above turbulent diffusivity is expected to work well in the limit that diffusive effects can be neglected and the velocity field transitions rapidly with respect to the advective time scale, see Appendix A. Under such circumstances, it is not unreasonable to think of velocity fluctuations as analogous to white noise with a given covariance structure. For example, letting ξ be a white-noise process and σ be a variance vector, if

$$\tilde{\mathbf{u}}'(\mathbf{x}, t) \approx \sigma(\mathbf{x}) \xi, \quad \text{where } \langle \xi(t + \tau) \xi(t) \rangle = \delta(\tau), \quad (4.31)$$

then a diffusivity is given by

$$D(\mathbf{x}) = \int_0^\infty \langle \tilde{\mathbf{u}}'(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}'(\mathbf{x}, t) \rangle d\tau = \sigma(\mathbf{x}) \otimes \sigma(\mathbf{x}). \quad (4.32)$$

Indeed, we will show that the intuitive estimate,

$$D(\mathbf{x}) = \int_0^\infty \langle \tilde{\mathbf{u}}'(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}'(\mathbf{x}, t) \rangle d\tau, \quad (4.33)$$

does correspond to (4.29). Using the white-noise limit, one can take the advection term in the tracer equation as a state-dependent noise term, in the Stratonovich sense, from whence standard arguments follow. However, the current approach goes beyond this limit by allowing for temporal correlations of the flow field.

We begin with two observations. First, the statistically steady velocity field satisfies

$$\langle \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle = \sum_m P_m \mathbf{u}_m(\mathbf{x}), \tag{4.34}$$

where $\mathbf{u}_m(\mathbf{x})$ for each m are the states of the Markov process. Second, recall that the vector $\mathcal{P}(\tau)\hat{\mathbf{e}}_n$ is a column vector of probabilities whose entries denote the probability of being found in state m given that at time $\tau = 0$, the probability of being found in state n is one. Thus, the conditional expectation of $\mathbf{u}(\mathbf{x}, t + \tau)$ given $\mathbf{u}(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})$ is

$$\langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \rangle_{\tilde{\mathbf{u}}(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})} = \left(\sum_m \mathbf{u}_m(\mathbf{x}) \hat{\mathbf{e}}_m \right)^T \mathcal{P}(\tau) \hat{\mathbf{e}}_n \tag{4.35}$$

$$= \sum_{im} e^{\tau \lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m(\mathbf{x}). \tag{4.36}$$

Equation (4.35) expresses the conditional expectation as a weighted sum of Markov states $\mathbf{u}_m(\mathbf{x})$.

We are now in a position to characterize the local turbulent diffusivity estimate. The local turbulent diffusivity is computed by taking the long time integral of a statistically steady flow field’s autocorrelation function, i.e.

$$D(\mathbf{x}) = \int_0^\infty R(\mathbf{x}, \tau) d\tau, \tag{4.37}$$

where

$$R(\mathbf{x}, \tau) \equiv \langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle - \langle \mathbf{u}(\mathbf{x}, t + \tau) \rangle \otimes \langle \mathbf{u}(\mathbf{x}, t) \rangle. \tag{4.38}$$

We calculate the second term under the statistically steady assumption of (4.38),

$$\langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \rangle \otimes \langle \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle = \left(\sum_m P_m \mathbf{u}_m(\mathbf{x}) \right) \otimes \left(\sum_n P_n \mathbf{u}_n(\mathbf{x}) \right). \tag{4.39}$$

For the first term of (4.38), we decompose the expectation into conditional expectations,

$$\langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle = \sum_n \langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle_{\tilde{\mathbf{u}}(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})} P_n. \tag{4.40}$$

Given that we are in a statistically steady state, we use (4.35) to establish

$$\langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle = \sum_n \langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle_{\tilde{\mathbf{u}}(\mathbf{x}, t) = \mathbf{u}_n(\mathbf{x})} P_n \tag{4.41}$$

$$= \sum_{imn} e^{\tau \lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n. \tag{4.42}$$

We isolate the $i = 1$ index and use $\lambda_1 = 0$, $\hat{\mathbf{e}}_m \cdot \mathbf{v}_1 = P_m$, and $\mathbf{w}_1 \cdot \hat{\mathbf{e}}_n = 1$ to arrive at

$$\sum_{mn} e^{\tau \lambda_1} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_1) (\mathbf{w}_1 \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n = \left(\sum_m P_m \mathbf{u}_m \right) \otimes \left(\sum_n P_n \mathbf{u}_n \right). \tag{4.43}$$

Equation (4.43) cancels with (4.39) so that in total, we have the following characterization of (4.38):

$$R(\mathbf{x}, \tau) = \sum_{(i \neq 1)mn} e^{\tau \lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n. \tag{4.44}$$

Equation (4.44) is integrated to yield the local turbulent diffusivity,

$$D(\mathbf{x}) = \int_0^\infty \mathbf{R}(\mathbf{x}, \tau) d\tau = \sum_{(i \neq 1)mn} \frac{-1}{\lambda_i} (\hat{\mathbf{e}}_m \cdot \mathbf{v}_i) (\mathbf{w}_i \cdot \hat{\mathbf{e}}_n) \mathbf{u}_m \otimes P_n \mathbf{u}_n, \quad (4.45)$$

where we used $\lambda_i < 0$ for $i > 1$. A comparison of (4.45) and (4.29) reveals the correspondence. The expression on the right-hand side is related to the negative of the Moore–Penrose inverse of generator Q . Thus, we see that estimating the diffusivity through the velocity autocorrelation integral is equivalent to neglecting diffusive effects and perturbation gradients. We further justify the local approximation as an asymptotic ‘white-noise’ limit and show to how connect the present class of stochastic models to that of Kraichnan (1968) in Appendix A.

5. Advanced example: stochastic wave in a channel

Now that we have outlined the general theory, we apply it to a wandering wave in a channel. The application is motivated by Flierl & McGillicuddy (2002), where the wave in question is considered a Rossby wave. Consider the 2-D vorticity equation in a channel

$$\partial_t q_\omega + \{\psi_\omega, q_\omega\} = f_\omega + \nu \Delta q_\omega, \quad (5.1)$$

$$\Delta \psi = q, \quad (5.2)$$

$$\partial_t \theta_\omega + \{\psi_\omega, \theta_\omega\} = s(\mathbf{x}) + \kappa \Delta \theta_\omega \quad (5.3)$$

with a stochastic forcing f_ω and $\{a, b\} = -\partial_y a \partial_x b + \partial_x a \partial_y b$. Here, $x \in [0, 2\pi]$ is periodic and $y \in [0, 1]$ is wall bounded. No-flux boundary conditions for the tracer and stress-free boundary conditions for the flow field on the wall $y = 0, 1$ are imposed. A solution to the vorticity equation is

$$q_\omega = -(1 + \pi^2) \sin(x + \varphi_\omega(t)) \sin(\pi y) \Rightarrow \psi_\omega = \sin(x + \varphi_\omega(t)) \sin(\pi y) \quad (5.4)$$

upon choosing

$$f_\omega = \partial_t q_\omega - \nu \Delta q_\omega \quad \text{and} \quad d\varphi_\omega = c dt + \epsilon \sqrt{2} dW_\omega. \quad (5.5a,b)$$

The nonlinear term in the vorticity equation is zero since $\psi_\omega \propto q_\omega$, i.e. $\{\psi_\omega, q_\omega\} = 0$. The phase φ is a random walk with drift c in a 2π periodic domain. The wave propagates, on average, to the left. For simulation purposes, we choose $c = \epsilon = 1$ and $\kappa = 0.01$.

The conditional mean equations for the passive tracer are

$$\partial_t \mathcal{P} = \partial_\varphi \left(-c\mathcal{P} + \epsilon^2 \partial_\varphi \mathcal{P} \right), \quad (5.6)$$

$$\partial_t \Theta + \{\psi, \Theta\} = \kappa \Delta \Theta + s(\mathbf{x})\mathcal{P} + \partial_\varphi \left(-c\Theta + \epsilon^2 \partial_\varphi \Theta \right), \quad (5.7)$$

where $\Theta = \Theta(x, \varphi, t)$ and $\psi = \sin(x + \varphi) \sin(\pi y)$ with $\varphi \in [0, 2\pi)$. Discretizing the phase variable φ with a finite volume method in φ yields

$$\partial_t \mathcal{P}_m = \sum_{m'} Q_{mm'} \mathcal{P}_{m'}, \quad (5.8)$$

$$\partial_t \Theta_m + \{\psi_m, \Theta_m\} = \kappa \Delta \Theta_m + s(\mathbf{x})\mathcal{P}_m + \sum_{m'} Q_{mm'} \Theta_{m'}, \quad (5.9)$$

where the details of the $Q_{mm'}$ and ψ_m are given in Appendix C.3. Discretizing with a spectral method for φ yields faster convergence to the continuous problem. However, the

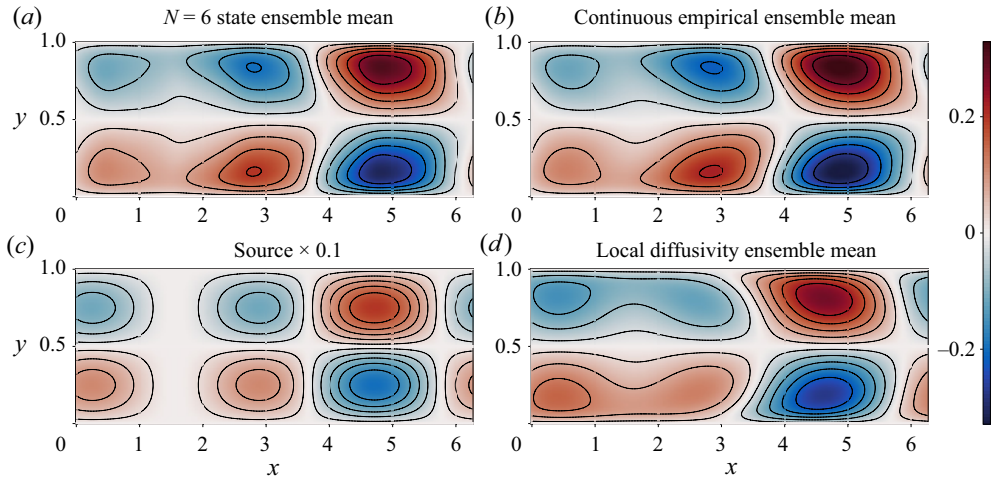


Figure 5. Ensemble mean channel comparison. We show the ensemble mean of the wandering wave as compared to the six-state system and a local diffusivity estimate. For reference, we also show the source term, scaled by a factor of 0.1 so as to be the same scale as the ensemble mean. A six-state system is visually similar to the continuous empirical mean, and the local diffusivity estimate is dissimilar.

finite-volume discretization yields sequences of realizable stochastic processes in the form of a Markov jump process between different flow fields.

For a local diffusivity, we use the auto-correlation of the velocity field,

$$K = \int_0^\infty dt \langle \tilde{\mathbf{u}}(\mathbf{x}, t + \tau) \otimes \tilde{\mathbf{u}}(\mathbf{x}, t) \rangle = \frac{1}{4} \begin{bmatrix} \pi^2 \cos(\pi y)^2 & -\pi \cos(\pi y) \sin(\pi y) \\ \pi \cos(\pi y) \sin(\pi y) & \sin(\pi y)^2 \end{bmatrix} \quad (5.10)$$

and calculate an approximate evolution equation

$$\partial_t \tilde{\varphi}_1 + \nabla \cdot (-K \nabla \tilde{\varphi}_1 - \kappa \nabla \tilde{\varphi}_1) = s(\mathbf{x}). \quad (5.11)$$

We numerically solve both (5.1), the conditional mean equation (5.9) for increasing number of states, and the local diffusivity equation, (5.11), for the particular choice of source term,

$$s(x, y) = \sin(x) \sin(\pi y) + \cos(2x) \sin(\pi y), \quad (5.12)$$

which is shown in figure 5(c). We use a doubly periodic Fourier code with $N = 32$ grid points in a domain $x \in [0, 2\pi)$ and $y \in [0, 2)$, and note that no-flux and stress-free boundary conditions at $y = 0$ and $y = 1$ are satisfied. We evolve all solutions to a final time $t = 10$ and use 10 000 ensemble members for the stochastic evolution of the equations. We choose to show the effect of the turbulent diffusivity indirectly via the choice of source term and the ensemble mean rather than directly examining the structure of the exact non-local tensor kernel for the present circumstance.

In addition, we show the ensemble mean of a six-state approximation to the kernel in figure 5(a), the empirical ensemble mean in figure 5(b) and the local diffusivity tensor results in figure 5(d). In the $x \in [4, 5]$ region of the plot, we see that the shape of the empirical ensemble mean is well captured by both the six-state approximation to the ensemble mean as well as the local diffusivity estimate; however, the magnitude is underestimated by the local diffusivity tensor. Furthermore, the local diffusivity tensor fails to capture the structure and magnitude of the ensemble mean in the $x \in [0, 3]$.

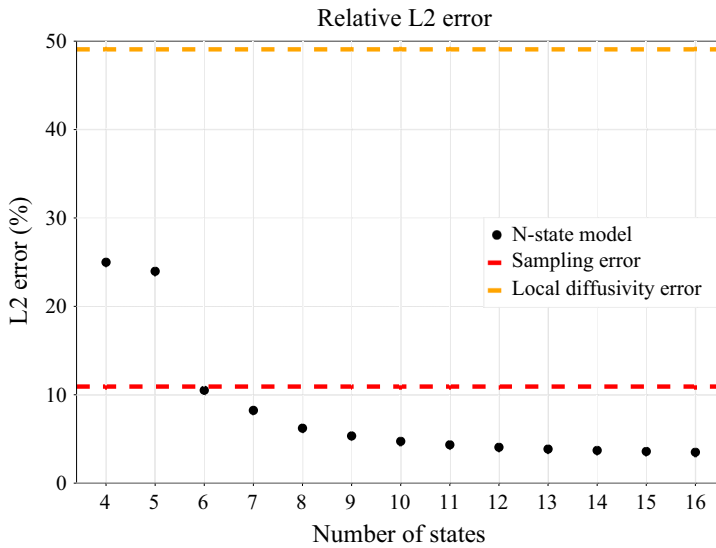


Figure 6. Ensemble mean channel Error. Here we show the relative quantitative error between an N -state approximation to the continuous process and the local diffusivity estimate. A six-state system is within the sampling error for the chosen source term.

In contrast, the $N = 6$ state approximation captures the structure and magnitude throughout the domain.

We quantify these statements with a relative L^2 error metric in figure 6. The sampling error is estimated by computing the difference in the empirical ensemble mean between ensemble members $1 \rightarrow 5000$ and $5001 \rightarrow 10000$, and dividing by the L^2 norm of the ensemble average of all the ensemble members. We see that increasing the number of states increases the fidelity of the discrete representation and that the local diffusivity yields a relative error of approximately 50%. Furthermore, a six-state model (as given by $N = 6$) captures the empirical ensemble mean within sampling error.

These errors are particular to the choice of source term and the resulting ensemble mean. The advantage of using the exact closure relation for the passive tracer is its applicability independent of the choice of source term in the tracer equation; however, a local diffusivity tensor can likely be optimized to reduce the error for a fixed source term. Similarly, using an N -state model could have different errors depending on the choice of source term, and it is more expedient to examine convergence to the exact operator.

6. Conclusions

We have introduced a conditional averaging procedure for analysing tracers advected by a stochastic flow and formulated the problem of finding a turbulence closure into solving a set of partial differential equations. When the flow statistics are modelled as a continuous-time Markov process with finite state space, the resulting system becomes tractable to compute analytically. Furthermore, we show that flow statistics with infinite state space can be approximated by ones with finite state space through a systematic discretization.

The resulting dimensionality of the equations depended on the number of variables required to describe flow statistics and the dimensionality of the flow. A flow characterized by m discrete states leads to a set of m -coupled equations of the same dimensionality as

the original. A flow characterized by a continuum of statistical variables can be discretized and reduced to the former. Eliminating the system's dependence on all but the ensemble mean leads to an operator characterization of the turbulence closure, allowing for an exploration of closures that do not invoke a scale separation hypothesis.

We analysed three examples – an Ornstein–Uhlenbeck process advecting a tracer in one dimension, a three-state system in arbitrary dimensions and a stochastically wandering wave in a 2-D channel – and outlined a general approach to obtaining a closure based on the spectrum of the transition probability operator. In the examples, we examined the role of non-locality in determining a statistically steady turbulence closure. Under general principles, we derived that the small velocity amplitude, weak tracer diffusivity and fast transition rate limit reduce the closure to a spatially heterogeneous tensor acting on ensemble mean gradients. Furthermore, we related this tensor to the time-integrated auto-correlation of the stochastic flow field.

We have not exhausted the number of examples the formulation offers nor simplifications leading to analytically tractable results. Interesting future directions include using Markov states estimated directly from turbulence simulations, analysing scale-separated flows, generalizing the advection-diffusion equation to reaction-advection-diffusion equations, and formulating optimal mixing problems. When the number of Markov states increases, the computational burden of estimating turbulent diffusivity operators becomes demanding; thus, there is a need to develop methods that exploit the structure of the problem as much as possible.

Mathematically, there are many challenges as well. All the arguments provided here are formal calculations, and the necessity for rigorous proofs remains. For example, a direct proof of the conditional averaging procedure is necessary. Ultimately, the goal is to reduce the stochastic-advection turbulence closure problem to one that can leverage theory from partial differential equations.

Supplementary material. Supplementary material is available at <https://github.com/sandrezza/StatisticalNonlocality>.

Acknowledgements. We want to thank the 2018 Geophysical Fluid Dynamics Program, where a significant portion of this research was undertaken. We are also indebted to T. Bischoff, S. Byrne, R. Ferrari and G. Menon for their encouragement and insightful remarks on this manuscript. We are especially indebted to the late Charlie Doering. His expert guidance directed our path towards considering discrete velocity processes, from whence the generalization in the manuscript grew. Charlie's intellectual legacy continues to inspire and drive our work. Finally, we thank the anonymous reviewers for their constructive suggestions.

Funding. Our work is supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. The Geophysical Fluid Dynamics Program is supported by the National Science Foundation, United States and the Office of Naval Research, United States. Glenn Flierl is supported by NSF grant OCE-2124211.

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.

 Andre N. Souza <https://orcid.org/0000-0001-8025-3558>.

Appendix A. White-noise limit and the Kraichnan ensemble

Here we wish to show how the considerations taken in the main text, specifically § 4.3, relate to Kraichnan's work with white noise in time and Gaussian process in space flow fields, see Kraichnan (1968). We will proceed in two steps. In the first step, we show how to take a 'white-noise' limit of an arbitrary set of flow fields \mathbf{u}_n and generators Q so that the time-integrated velocity auto-correlations stay finite. This procedure has been commented

on by both Kraichnan (1968) and Falkovich *et al.* (2001) as a necessary step to remove the stochastic ambiguity of a delta-correlated in time flow field. In the second step, we show how to discretize a Gaussian process to a finite state space.

Introduce a scale parameter γ for each velocity state and the generator Q as

$$\mathbf{u}_n(\mathbf{x}) \mapsto \sqrt{\gamma} \mathbf{u}_n(\mathbf{x}) \quad \text{and} \quad Q \mapsto \gamma Q. \tag{A1a,b}$$

The integrated velocity autocorrelation remains invariant since the characteristic velocity scales like $U = O(\sqrt{\gamma})$, the characteristic time scale scales like $\tau = \mathcal{O}(\gamma^{-1})$ and thus the integrated velocity autocorrelation is $U^2 \tau = \mathcal{O}(1)$ as $\gamma \rightarrow \infty$; however, the mean free path of a particle is now $\ell \propto U \tau = \mathcal{O}(1/\sqrt{\gamma})$ and goes to zero in the limit. This limit yields a delta-correlated flow field in time as $\gamma \rightarrow \infty$.

From (4.20), we have the exact relation as $\gamma \rightarrow \infty$ for $j \neq 1$,

$$\begin{aligned} & \sum_{(i \neq 1)n} \underbrace{(\mathbf{w}_j \cdot \hat{\mathbf{e}}_n)(\hat{\mathbf{e}}_n \cdot \mathbf{v}_i) \nabla \cdot (\sqrt{\gamma} \mathbf{u}_n \varphi_i)}_{\mathcal{O}(\sqrt{\gamma})} - \underbrace{\frac{\kappa \Delta}{\mathcal{O}(1)} \varphi_j}_{\mathcal{O}(1)} - \underbrace{\frac{\gamma \lambda_j}{\mathcal{O}(\gamma)} \varphi_j}_{\mathcal{O}(\gamma)} \\ &= - \sum_n \underbrace{P_n(\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla \cdot (\sqrt{\gamma} \mathbf{u}_n \varphi_1)}_{\mathcal{O}(\sqrt{\gamma})}. \end{aligned} \tag{A2}$$

Since the integrated autocorrelation of velocity is independent of γ , we assume $\varphi_1 = \mathcal{O}(1)$, which yields the leading order balance

$$\underbrace{\frac{\gamma \lambda_j}{\mathcal{O}(\gamma)} \varphi_j}_{\mathcal{O}(\gamma)} \approx \sum_n \underbrace{P_n(\mathbf{w}_j \cdot \hat{\mathbf{e}}_n) \nabla \cdot (\sqrt{\gamma} \mathbf{u}_n \varphi_1)}_{\mathcal{O}(\sqrt{\gamma})}. \tag{A3}$$

The above relation was assumed in (4.25) but is now justified as a particular asymptotic limit. Consequently, $\varphi_j = \mathcal{O}(1/\sqrt{\gamma})$ as $\gamma \rightarrow \infty$ for $j \neq 1$. Hence, the perturbation terms are small, consistent with the requirements of a local diffusivity estimate. We note that the assumption $\varphi_1 = \mathcal{O}(1)$ is self-consistent. These calculations complete the first step.

We have seen how to rescale and take limits to get a local diffusivity estimate from an arbitrary Q matrix and velocity states. The only additional ingredient to connect the present work to the Kraichnan ensemble is to observe that a Gaussian process is a structured matrix version of Q with particular choices for the velocity states. The Karhunen–Loève expansion implies that the Gaussian process velocity field (at a fixed time t) is represented as

$$\mathbf{u}_\omega(\mathbf{x}, t) = \sum_{n=1}^{\infty} A_{\omega_n}(t) \Phi_n(\mathbf{x}), \tag{A4}$$

where each $A_{\omega_n}(t)$ is, for example, an independent Ornstein–Uhlenbeck process (although any process with a Gaussian stationary distribution will suffice) with integrated autocorrelation one and $\Phi_n(\mathbf{x})$ are eigenvectors of the covariance tensor associated with the Gaussian process. They satisfy the relation

$$\int_{\Omega} d\mathbf{x} \Phi_n(\mathbf{x}) \cdot \Phi_m(\mathbf{x}) = \Lambda_n \delta_{nm}, \tag{A5}$$

where Λ_n are the eigenvalues of the covariance tensor.

We now show how to express the above infinite-dimensional state space as a limit of the finite-dimensional ones. Represent each amplitude A_{ω_n} as an N -state model with transition

matrix \tilde{Q} (as was done for the Ornstein–Uhlenbeck process) and observe that each one is independent. The Q matrix becomes a Kronecker sum of infinitely many copies of the same \tilde{Q} . In practice, this infinite Kronecker sum is truncated to a finite number, let us say the first M of them as ordered by the eigenvalues Λ_n of the covariance tensor. The resulting discrete states are then given by the product of N possible modal amplitudes for a fixed basis function Φ with M total number of independent modes, hence, an N^M finite dimensional space. More concretely, if we discretize with three modal amplitudes and each modal amplitude with a two-state system, then each amplitude is either $+1$ or -1 , leading to a total possible $2^3 = 8$ pairings, e.g.

$$u_1 = \Phi_1(x) + \Phi_2(x) + \Phi_3(x), \quad u_2 = -\Phi_1(x) + \Phi_2(x) + \Phi_3(x), \quad (A6a,b)$$

$$u_3 = \Phi_1(x) + \Phi_2(x) - \Phi_3(x), \quad u_4 = -\Phi_1(x) + \Phi_2(x) - \Phi_3(x), \quad (A7a,b)$$

$$u_5 = \Phi_1(x) - \Phi_2(x) + \Phi_3(x), \quad u_6 = -\Phi_1(x) - \Phi_2(x) + \Phi_3(x), \quad (A8a,b)$$

$$u_7 = \Phi_1(x) - \Phi_2(x) - \Phi_3(x), \quad u_8 = -\Phi_1(x) - \Phi_2(x) - \Phi_3(x) \quad (A9a,b)$$

and the Kronecker sum matrix

$$\tilde{Q} = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, \quad (A10)$$

$$Q = \tilde{Q} \oplus \tilde{Q} \oplus \tilde{Q} = \frac{1}{2} \begin{bmatrix} -3 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -3 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -3 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & -3 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -3 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -3 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & -3 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & -3 \end{bmatrix}, \quad (A11)$$

where each transition corresponds to a single sign flip of the Gaussian process basis functions that constitute the velocity states.

Given that our analysis was performed for an arbitrary-sized Q matrix, we use finite truncations and then take limits afterward. Note the order of the limits taken: first, truncate to a finite state space, next, take the white noise limit, then lastly, take the limit to the infinite state space. Those limits, taken in that order, yield the same model as that of Kraichnan (1968).

Appendix B. An alternative formal derivation

We wish to show that one can work directly with the continuous formulation of the advection-diffusion equations for the derivation of the conditional mean equations. Although we consider a finite (but arbitrarily large) number of Markov states here, considering a continuum follows *mutatis mutandi*. In §2, we wrote down the master equation for the discretized stochastic system as

$$\partial_t \rho_m = \sum_i \frac{\partial}{\partial \theta^i} \left[\left(\sum_{jkc} A_{ijk}^c u_m^{k,c} \theta^j - \sum_j D_{ij} \theta^j - s^i \right) \rho_m \right] + \sum_n Q_{mn} \rho_n. \quad (B1)$$

We introduce the (spatial) volume element Δx_i to rewrite (B1) in the evocative manner,

$$\partial_t \rho_m = \sum_i \Delta x_i \frac{1}{\Delta x_i} \frac{\partial}{\partial \theta^i} \left[\left(\sum_{jkc} A_{ijk}^c u_m^{k,c} \theta^j - \sum_j D_{ij} \theta^j - s^i \right) \rho_m \right] + \sum_n Q_{mn} \rho_n. \quad (\text{B2})$$

We now take limits

$$\sum_i \Delta x_i \stackrel{\text{'lim'}}{=} \int d\mathbf{x}, \quad (\text{B3})$$

$$\frac{1}{\Delta x_i} \frac{\partial}{\partial \theta^i} \stackrel{\text{'lim'}}{=} \frac{\delta}{\delta \theta(\mathbf{x})}, \quad (\text{B4})$$

$$\sum_{jkc} A_{ijk}^c u_m^{k,c} \theta^j + \sum_j D_{ij} \theta^j - s^i \stackrel{\text{'lim'}}{=} \mathbf{u}_m \cdot \nabla \theta - \kappa \Delta \theta - s, \quad (\text{B5})$$

to get the functional evolution equation for the probability density,

$$\partial_t \rho_m = \int d\mathbf{x} \frac{\delta}{\delta \theta(\mathbf{x})} ([\mathbf{u}_m \cdot \nabla \theta - \kappa \Delta \theta - s] \rho_m) + \sum_n Q_{mn} \rho_n, \quad (\text{B6})$$

where \mathbf{x} is a continuous index. The notation here is similar to that of Zinn-Justin (2021, § 35). As before, we can derive the conditionally averaged equations directly from the above. To do so, we make the additional correspondence

$$d\theta \stackrel{\text{'lim'}}{=} \mathcal{D}[\theta]. \quad (\text{B7})$$

We now define the same quantities as before, but using the field integral

$$\mathcal{P}_m \equiv \int \mathcal{D}[\theta] \rho_m \quad \text{and} \quad \Theta_m(y) \equiv \int \mathcal{D}[\theta] \theta(y) \rho_m. \quad (\text{B8a,b})$$

The discrete indices i, j, k in the equations from § 2 are replaced by continuous labels \mathbf{x} and \mathbf{y} . We only use a few formal properties of the field integral, with direct correspondence to n -dimensional integrals. We use linearity, i.e. for two mappings with compatible ranges $\mathcal{F}[\theta]$ and $\mathcal{H}[\theta]$,

$$\int \mathcal{D}[\theta] (\mathcal{F}[\theta] + \mathcal{H}[\theta]) = \int \mathcal{D}[\theta] \mathcal{F}[\theta] + \int \mathcal{D}[\theta] \mathcal{H}[\theta]. \quad (\text{B9})$$

We use the analogue to the divergence theorem,

$$\int \mathcal{D}[\theta] \int d\mathbf{x} \frac{\delta}{\delta \theta(\mathbf{x})} ([\mathbf{u}_m \cdot \nabla \theta - \kappa \Delta \theta - s] \rho_m) = 0 \quad (\text{B10})$$

\Leftrightarrow

$$\int d\theta \nabla_{\theta} \cdot (f \rho) = 0 \quad (\text{B11})$$

since the integral of a divergence should be zero if the probabilities vanish at infinity (i.e. that our tracer cannot have infinite values at a given point in space). We also make use of

the integration by parts, i.e. for some functionals \mathcal{F} and \mathcal{H} ,

$$\int \mathcal{D}[\theta] \mathcal{H} \int dx \frac{\delta}{\delta \theta(x)} \mathcal{F} = - \int \mathcal{D}[\theta] \int dx \frac{\delta \mathcal{H}}{\delta \theta(x)} \mathcal{F} \tag{B12}$$

\Leftrightarrow

$$\int d\theta h \nabla_{\theta} \cdot f = - \int d\theta (\nabla_{\theta} h) \cdot f. \tag{B13}$$

And finally, we also interchange sums and integrals,

$$\int \mathcal{D}[\theta] (\Delta \theta) \rho_m = \Delta \int \mathcal{D}[\theta] \theta \rho_m = \Delta \Theta_m \tag{B14}$$

\Leftrightarrow

$$\int d\theta \left(\sum_j D_{\ell j} \theta^j \rho_m \right) = \sum_j D_{\ell j} \int d\theta \theta^j \rho_m = \sum_j D_{\ell j} \Theta_m^j. \tag{B15}$$

We proceed similarly for the $\mathbf{u}_m \cdot \nabla$ term. We also use properties of the variational derivative, such as,

$$\frac{\delta \theta(y)}{\delta \theta(x)} = \delta(x - y) \Leftrightarrow \frac{\partial \theta^{\ell}}{\partial \theta^i} = \delta_{\ell i}. \tag{B16}$$

Taken together, one can directly obtain (2.13) and (2.14) by first integrating (B6) with respect to $\mathcal{D}[\theta]$ to get

$$\partial_t \mathcal{P}_m = \sum_n \mathcal{Q}_{mn} \mathcal{P}_n, \tag{B17}$$

and multiplying (B6) by $\theta(y)$ and then integrating with respect to $\mathcal{D}[\theta]$ to get

$$\partial_t \Theta_m(y, t) + \nabla_y \cdot (\mathbf{u}_m(y) \Theta_m(y, t) - \kappa \nabla_y \Theta_m(y, t)) = s(y) \mathcal{P}_m + \sum_n \mathcal{Q}_{mn} \Theta_n(y, t). \tag{B18}$$

In the above expression, removing explicit dependence of the position variable yields

$$\partial_t \Theta_m + \nabla \cdot (\mathbf{u}_m \Theta_m - \kappa \nabla \Theta_m) = s \mathcal{P}_m + \sum_n \mathcal{Q}_{mn} \Theta_n. \tag{B19}$$

Our reason for mentioning the above methodology is that it allows for expedited computations. There is no need to explicitly discretize, perform usual n -dimensional integral manipulations and then take limits afterward. For example, computing the conditional two-moment equations defined by the variable

$$C_m(y, z, t) \equiv \int \mathcal{D}[\theta] \theta(y) \theta(z) \rho_m \tag{B20}$$

is obtained by multiplying (B6) by $\theta(y)$ and $\theta(z)$, and integrating with respect to $\mathcal{D}[\theta]$,

$$\begin{aligned} \partial_t C_m + \nabla_y \cdot (\mathbf{u}_m(y) C_m - \kappa \nabla_y C_m) + \nabla_z \cdot (\mathbf{u}_m(z) C_m - \kappa \nabla_z C_m) \\ = s(z) \Theta_m(y) + s(y) \Theta_m(z) + \sum_n \mathcal{Q}_{mn} C_n. \end{aligned} \tag{B21}$$

In particular, we note the source term on the right-hand side and the appearance of the first conditional moment. In the derivation, we used the product rule

$$\frac{\delta(\theta(y)\theta(z))}{\delta\theta(x)} = \delta(x-y)\theta(z) + \delta(x-z)\theta(y). \tag{B22}$$

If the advection-diffusion equation is m -dimensional and we have N Markov states, the above equation is made up of N coupled $2m$ dimensional partial differential equations. Indeed the equation for the M th moment is made up of N coupled $M \times m$ dimensional partial differential equations.

Appendix C. A heuristic overview of the master equation and discretizations

In this section, we provide an argument for the form of the master equation in the main text, (2.18) in § 2. Our starting point is § C.1, where we use the Liouville equation for two continuous variables. We then apply the finite volume method to the Fokker–Planck equation of an Ornstein–Uhlenbeck process to derive the transition matrices used in the two-state and three-state systems in § 3. We conclude with a formal argument for the use of discrete Markov states as an approximation to the compressible Euler equations in § C.4.

C.1. Two-variable system

Suppose that we have two variables $x, y \in \mathbb{R}$ governed by the equations

$$\frac{dx}{dt} = f(x) + \sqrt{2}\sigma\xi, \tag{C1}$$

$$\frac{dy}{dt} = g(x, y), \tag{C2}$$

where ξ is white noise. In this context, we think of x as being our flow field \mathbf{u} and y as the tracer θ . The master equation implied by the dynamics is

$$\partial_t \rho = -\partial_x \left(f(x)\rho - \sigma^2 \partial_x \rho \right) - \partial_y (g(x, y)\rho). \tag{C3}$$

We now discretize the equation with respect to the x -variable by partitioning space into non-overlapping cells, characterized by domains Ω_m . First, we start with the Fokker–Planck equation for x , which is independent of the y -variable,

$$\partial_t P = -\partial_x \left(f(x)P - \sigma^2 \partial_x P \right). \tag{C4}$$

Observe the relation $\int \rho(x, y, t) dy = P(x, t)$. Define our coarse grained variable \mathcal{P}_m as

$$\mathcal{P}_m \equiv \int_{\Omega_m} P(x) dx, \tag{C5}$$

which is a probability. Thus, the discretization of (C4) becomes

$$\partial_t \int_{\Omega_m} P dx = - \int_{\Omega_m} \partial_x \left(f(x)P - \sigma^2 \partial_x P \right) dx \tag{C6}$$

\approx

$$\partial_t \mathcal{P}_m = \sum_n Q_{mn} \mathcal{P}_n \tag{C7}$$

for a generator \mathcal{Q} which we derive in (C.2) with respect to a chosen numerical flux. Heuristically, going from (C6) to (C7) is accomplished by observing that \mathcal{P}_m is a probability and the operator $\mathcal{L} \equiv \partial_x(f(x) \bullet - \sigma^2 \partial_x \bullet)$ is linear; thus, upon discretization, the operator is represented a matrix acting on the chosen coarse grained variables \mathcal{P}_n . (It is, of course, possible to approximate using a nonlinear operator, but for simplicity, we only consider the linear case.) The property $\sum_m \mathcal{Q}_{mn} = \mathbf{0}$ is the discrete conservation of probability.

Going back to (C3), defining

$$\rho_m(y) \equiv \int_{\Omega_m} \rho(x, y) \, dx, \tag{C8}$$

introducing $x_m \in \Omega_m$, and performing the same discretization for the joint Markov system yields

$$\partial_t \rho_m = \sum_n \mathcal{Q}_{mn} \rho_n - \partial_y (g(x_m, y) \rho_m), \tag{C9}$$

where we used the approximation

$$\int_{\Omega_m} g(x, y) \rho(x, y) \, dx \approx g(x_m, y) \int_{\Omega_m} \rho(x, y) \, dx = g(x_m, y) \rho_m(y). \tag{C10}$$

The x_m are the Markov states and the \mathcal{Q}_{mn} serves as the specification for transitioning between different states. We also observe that one can simply start with the discrete states for x and continuous variables for y to directly obtain (C9), as was done in the main text.

In what follows, we give a concrete example of deriving a generator \mathcal{Q} from a finite-volume discretization of an Ornstein–Uhlenbeck (OU) process. We explicitly mention the kind of discretization that we use since retaining mimetic properties of the generator \mathcal{Q} is not guaranteed with other discretizations. Furthermore, using a finite volume discretization allows for the resulting discretization to be interpreted as a continuous-time Markov process with a finite state space.

C.2. Example discretization

Consider an Ornstein–Uhlenbeck process and the resulting Fokker–Planck equation,

$$\partial_t \rho = -\partial_x (-x\rho - \partial_x \rho). \tag{C11}$$

We discretize the above equation with $N + 1$ cells, where $N = 1$ and $N = 2$ correspond to the two- and three-state systems, respectively. Using a finite volume discretization, we take our cells to be

$$\Omega_m = [\Delta x (m - 1/2 - N/2), \Delta x (m + 1/2 - N/2)], \tag{C12}$$

$$\Delta x = \frac{2}{\sqrt{N}} \tag{C13}$$

for $m = 0, 1, \dots, N$. Our choice implies that cell centres (the discrete Markov states) are

$$x_m = \Delta x (m - N/2) \tag{C14}$$

for $m = 0, \dots, N$, and cell faces are

$$x_m^f = \Delta x (m - 1/2 - N/2) \tag{C15}$$

for $m = 0, \dots, N + 1$. We define

$$\mathcal{P}_m = \int_{\Omega_m} \rho \, dx \quad \text{and} \quad \bar{\rho}_m \Delta x = \mathcal{P}_m. \tag{C16a,b}$$

Upon integrating with respect to the control volume, we obtain

$$\frac{d}{dt} \mathcal{P}_m = - (-x\rho - \partial_x \rho)|_{x=x_m^f} + (-x\rho - \partial_x \rho)|_{x=x_{m+1}^f}. \tag{C17}$$

The numerical flux is chosen as follows:

$$(-x\rho - \partial_x \rho)|_{x=x_m^f} \approx - \frac{x_{m-1} \bar{\rho}_{m-1} + x_m \bar{\rho}_m}{2} - \frac{\bar{\rho}_m - \bar{\rho}_{m-1}}{\Delta x} \tag{C18}$$

$$= - \frac{x_{m-1} \mathcal{P}_{m-1} + x_m \mathcal{P}_m}{2 \Delta x} - \frac{\mathcal{P}_m - \mathcal{P}_{m-1}}{(\Delta x)^2} \tag{C19}$$

$$= \frac{1}{2} ((N - m + 1) \mathcal{P}_{m-1} - m \mathcal{P}_m), \tag{C20}$$

where we use the convention $\mathcal{P}_{-1} = \mathcal{P}_{N+1} = 0$ so that boundaries, corresponding to indices $m = 0$ and $m = N + 1$, imply no flux conditions. Combining the flux estimates for both cell boundaries, the evolution equation for the probabilities \mathcal{P}_m becomes

$$\partial_t \mathcal{P}_m = \frac{1}{2} [(N - m + 1) \mathcal{P}_{m-1} - N \mathcal{P}_m + (m + 1) \mathcal{P}_{m+1}], \tag{C21}$$

which implies the generator

$$\mathcal{Q}_{mn} = \frac{1}{2} (-N \delta_{mn} + n \delta_{(m+1)n} + (N - n) \delta_{(m-1)n}). \tag{C22}$$

Equation (C21) emphasizes the row structure of the matrix whereas (C22) emphasizes the column structure. The steady state probability distribution is the binomial distribution

$$P_m = 2^{-N} \binom{N}{m}. \tag{C23}$$

(The continuous steady state distribution is a Normal distribution $\rho(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.) Furthermore, the eigenvectors and eigenvalues of the matrix are in correspondence with the eigenfunctions and eigenvalues of the Ornstein–Uhlenbeck process as noted by Hagan *et al.* (1989). In particular, the cell centre vector, (C26), is a left eigenvector of \mathcal{Q}_{mn} with eigenvalue $\lambda = -1$. This relation is useful for calculating the auto-correlation of the Markov process since (4.35) only involves one eigenvalue. We used the generator, (C22), in the construction of the two- and three-state systems.

C.3. Example discretization 2

Consider a random walk with drift in a periodic domain and the resulting Fokker–Planck equation,

$$\partial_t \rho = -\partial_\varphi (\rho - \partial_\varphi \rho). \tag{C24}$$

We use control volumes,

$$\Omega_m = [m \Delta \varphi, (m + 1) \Delta \varphi] \quad \text{and} \quad \Delta \varphi = \frac{2\pi}{N} \tag{C25}$$

for $m = 0, 1, \dots, N - 1$. Our choice implies that cell centres are

$$\varphi_m = \Delta\varphi(m + 1/2) \tag{C26}$$

for $m = 0, \dots, N - 1$, and cell faces are

$$\varphi_m^f = m\Delta\varphi \tag{C27}$$

for $m = 0, \dots, N - 1$. Using a central flux for the advective term, the standard flux for the diffusive term, and accounting for periodicity yields the matrices

$$A_{(i\mp 1)\%N+1, i\%N+1} = \pm 1/2, D_{i,i} = -2 \quad \text{and} \quad D_{(i\pm 1)\%N+1, i\%N+1} = 1 \tag{C28}$$

for $i = 1, \dots, N$ and zero otherwise. Here, % is the modulus operation and accounts for super/sub diagonals of the matrices along with periodicity.

The transition matrix is then

$$Q_{mn} = \frac{1}{\Delta\varphi}A_{mn} + \frac{1}{\Delta\varphi^2}D_{mn}. \tag{C29}$$

With regards to the calculations in § 5, we take our stream function states to be

$$\psi_m = \cos(x + m\Delta\varphi) \sin(\pi y) \tag{C30}$$

for $m = 0, 1, \dots, N - 1$.

C.4. A finite volume approximation in function space

We start with the compressible Euler equations

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{C31}$$

$$\partial_t \rho \mathbf{u} + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p = 0, \tag{C32}$$

$$\partial_t \rho e + \nabla \cdot (\mathbf{u} [\rho e + p]) = 0, \tag{C33}$$

$$p(\rho, \rho \mathbf{u}, \rho e) = p, \tag{C34}$$

where ρ is density, $\rho \mathbf{u}$ is the momentum, $\mathbf{u} = \rho \mathbf{u} / \rho$ is the velocity, ρe is the total energy density and p is a thermodynamic pressure. (For example, one could use the pressure for an ideal gas $p = (\gamma - 1)(\rho e - \rho |\mathbf{u}|^2 / 2)$ with $\gamma = 7/5$.) Here, we introduce Z as a probability density in function space for the state variables $S \equiv (\rho, \rho \mathbf{u}, \rho e)$. In the notation of [Appendix B](#), the evolution equation for the statistics Z are

$$\begin{aligned} \partial_t Z = \int dx & \left[\frac{\delta}{\delta \rho} (\nabla \cdot [\rho \mathbf{u}] Z) + \frac{\delta}{\delta \rho \mathbf{u}} (\nabla \cdot [\rho \mathbf{u} \otimes \mathbf{u}] Z + \nabla p Z) \right. \\ & \left. + \frac{\delta}{\delta \rho e} (\nabla \cdot (\mathbf{u} [\rho e + p]) Z) \right], \end{aligned} \tag{C35}$$

where we have suppressed the index x in the variational derivatives.

Now consider a partition in function space into domains Ω_m and let S_m denote a value of a state within the set S_m . In this case, we define the probability as

$$\mathcal{P}_m \equiv \int_{\Omega_m} \mathcal{D}[\rho] \mathcal{D}[\rho \mathbf{u}] \mathcal{D}[\rho e] \mathcal{Z}. \quad (\text{C36})$$

In analogy with the calculations in § C.2, integrating (C35) with respect to a control volume Ω_m would result in an approximation of the form

$$\partial_t \mathcal{P}_m = \sum_n \mathcal{Q}_{mn} \mathcal{P}_n \quad (\text{C37})$$

for some generator \mathcal{Q}_{mn} . The entries of the generator are functionals of the states $S_m \in \Omega_m$. Performing the necessary integrals and re-expressing it in this finite form is done indirectly through data-driven methods with time series as in Klus, Koltai & Schütte (2016), Fernex, Noack & Semaan (2021) or Maity, Koltai & Schumacher (2022).

The difficulty of performing a discretization from first principles comes from choosing the subsets of function space to partition and carrying out the integrals in function space. Periodic orbits and fixed points of a flow serve as a natural skeleton for function space, but are typically burdensome to compute. We offer no solution, but hope that in the future, such direct calculations are rendered tractable. In the meanwhile, indirect data-driven methods are the most promising avenue for the calculation of the generator \mathcal{Q} .

REFERENCES

- AVELLANEDA, M. & MAJDA, A.J. 1991 An integral representation and bounds on the effective diffusivity in passive advection by laminar and turbulent flows. *Commun. Math. Phys.* **138** (2), 339–391.
- BHAMIDIPATI, N., SOUZA, A.N. & FLIERL, G.R. 2020 Turbulent mixing of a passive scalar in the ocean mixed layer. *Ocean Model.* **149**, 101615.
- BROWN, E. & AHLERS, G. 2007 Large-scale circulation model for turbulent Rayleigh–Bénard convection. *Phys. Rev. Lett.* **98**, 134501.
- DAVIS, C., FLIERL, G., WIEBE, P.H. & FRANKS, P. 1991 Micropatchiness, turbulence and recruitment in plankton. *J. Mar. Res.* **49**, 109–151.
- FALKOVICH, G., GAWĘDZKI, K. & VERGASSOLA, M. 2001 Particles and fields in fluid turbulence. *Rev. Mod. Phys.* **73**, 913–975.
- FERNEX, D., NOACK, B.R. & SEMAAN, R. 2021 Cluster-based network modeling; from snapshots to complex dynamical systems. *Sci. Adv.* **7** (25), eabf5006.
- FERRARI, R., MANFROI, A.J. & YOUNG, W. 2001 Strongly and weakly self-similar diffusion. *Physica D* **154**, 111–137.
- FLIERL, G.R. & MCGILLICUDDY, D.J. 2002 The Sea: Ideas and Observations on Progress in the Study of the Seas, In *Mesoscale and submesoscale physical-biological interactions*, vol. 12, pp. 113–185. John Wiley and Sons.
- GALLET, B. & FERRARI, R. 2020 The vortex gas scaling regime of baroclinic turbulence. *Proc. Natl Acad. Sci.* **117** (9), 4491–4497.
- GORBUNOVA, A., PAGANI, C., BALARAC, G., CANET, L. & ROSSETTO, V. 2021 Eulerian spatiotemporal correlations in passive scalar turbulence. *Phys. Rev. Fluids* **6**, 124606.
- HAGAN, P.S., DOERING, C.R. & LEVERMORE, C.D. 1989 Mean exit times for particles driven by weakly colored noise. *SIAM J. Appl. Maths* **49** (5), 1480–1513.
- HAKULINEN, V. 2003 Passive advection and the degenerate elliptic operators Mn. *Commun. Math. Phys.* **235** (1), 1–45.
- HASSANZADEH, P., CHINI, G.P. & DOERING, C.R. 2014 Wall to wall optimal transport. *J. Fluid Mech.* **751**, 627–662.
- HOPF, E. 1952 Statistical hydromechanics and functional calculus. *Indiana Univ. Math. J.* **1**, 87–123.
- KLUS, S., KOLTAI, P. & SCHÜTTE, C. 2016 On the numerical approximation of the perron-frobenius and koopman operator. *J. Comput. Dyn.* **3**, 51–79.
- KNOBLOCH, E. 1977 The diffusion of scalar and vector fields by homogeneous stationary turbulence. *J. Fluid Mech.* **83** (1), 129–140.

- KRAICHNAN, R.H. 1968 Small-structure of a scalar field convected by turbulence. *Phys. Fluids* **11** (5), 945–953.
- LORENZ, E.N. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.
- MAITY, P., KOLTAI, P. & SCHUMACHER, J. 2022 Large-scale flow in a cubic Rayleigh–Bénard cell: long-term turbulence statistics and markovianity of macrostate transitions. *Phil. Trans. R. Soc. A* **380** (2225), 20210042.
- MOFFATT, H.K. 1983 Transport effects associated with turbulence with particular attention to the influence of helicity. *Rep. Prog. Phys.* **46** (5), 621.
- PAPPALETTERA, U. 2022 Quantitative mixing and dissipation enhancement property of Ornstein–Uhlenbeck flow. *Commun. Part. Diff. Equ.* **47** (12), 2309–2340.
- POPE, S.B. 2011 Simple models of turbulent flows. *Phys. Fluids* **23** (1), 011301.
- SCHNEIDER, T., LAN, S., STUART, A. & TEIXEIRA, J. 2017 Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.* **44** (24), 12,396–12,417.
- SUGIYAMA, K., NI, R., STEVENS, R.J.A.M., CHAN, T.S., ZHOU, S.-Q., XI, H.-D., SUN, C., GROSSMANN, S., XIA, K.-Q. & LOHSE, D. 2010 Flow reversals in thermally driven turbulence. *Phys. Rev. Lett.* **105**, 034503.
- SURI, B., TITHOF, J., GRIGORIEV, R.O. & SCHATZ, M.F. 2017 Forecasting fluid flows using the geometry of turbulence. *Phys. Rev. Lett.* **118**, 114501.
- TAN, Z., KAUL, C.M., PRESSEL, K.G., COHEN, Y., SCHNEIDER, T. & TEIXEIRA, J. 2018 An extended eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection. *J. Adv. Model. Earth Syst.* **10** (3), 770–800.
- TAYLOR, G.I. 1922 Diffusion by continuous movements. *Proc. Lond. Math. Soc.* **s2–20** (1), 196–212.
- THIFFEAULT, J.-L. 2012 Using multiscale norms to quantify mixing and transport. *Nonlinearity* **25** (2), R1–R44.
- WEEKS, E.R., URBACH, J.S. & SWINNEY, H.L. 1996 Anomalous diffusion in asymmetric random walks with a quasi-geostrophic flow example. *Physica D* **97** (1), 291–310.
- WEINSTOCK, J. 1969 Formulation of a statistical theory of strong plasma turbulence. *Phys. Fluids* **12** (5), 1045–1058.
- XIA, K.-Q., SUN, C. & ZHOU, S.-Q. 2003 Particle image velocimetry measurement of the velocity field in turbulent thermal convection. *Phys. Rev. E* **68**, 066303.
- ZHANG, Y., ZHANG, N. & TIAN, M. 2022 Internal dynamics of magma ocean and its linkage to atmospheres. *Acta Geochim.* **41** (4), 568–591.
- ZINN-JUSTIN, J. 2021 *Quantum Field Theory and Critical Phenomena*, 5th edn. Oxford University Press.