J. Fluid Mech. (2024), *vol.* 992, A6, doi:10.1017/jfm.2024.537

JFM PAPERS

Mixing as a correlated aggregation process

J. Heyman[1],[†], E. Villermaux[2], P. Davy[1] and T. Le Borgne[1]

[1]Géosciences Rennes, Université de Rennes, UMR CNRS 6118, 263 Avenue du Général Leclerc,
F-35042 Rennes, France

[2]Aix Marseille Université, CNRS, Centrale Marseille, IRPHE UMR 7342, 13384 Marseille, France

(Received 26 September 2023; revised 24 May 2024; accepted 25 May 2024)

Mixing describes the process by which solutes evolve from an initial heterogeneous state to uniformity under the stirring action of a fluid flow. Fluid stretching forms thin scalar lamellae that coalesce due to molecular diffusion. Owing to the linearity of the advection–diffusion equation, coalescence can be envisioned as an aggregation process. Here, we demonstrate that in smooth two-dimensional chaotic flows, mixing obeys a correlated aggregation process, where the spatial distribution of the number of lamellae in aggregates is highly correlated with their elongation, and is set by the fractal properties of the advected material lines. We show that the presence of correlations makes mixing less efficient than a completely random aggregation process because lamellae with similar elongations and scalar levels tend to remain isolated from each other. We show that correlated aggregation is uniquely determined by a single exponent that quantifies the effective number of random aggregation events. These findings expand aggregation theories to a larger class of systems, which have relevance to various fundamental and applied mixing problems.

Key words: porous media, chaotic advection, laminar mixing

1. Introduction

Mixing of solutes by the stirring action of a fluid flow is ubiquitous to many natural and industrial processes (Ottino 1990). The evolution of the solute concentration c in an incompressible velocity field $\boldsymbol{v}$ is governed by the conservation equation

$$\partial_t c + \boldsymbol{v} \cdot \nabla c = \kappa \, \nabla^2 c, \tag{1.1}$$

[†] Email address for correspondence: joris.heyman@univ-rennes.fr

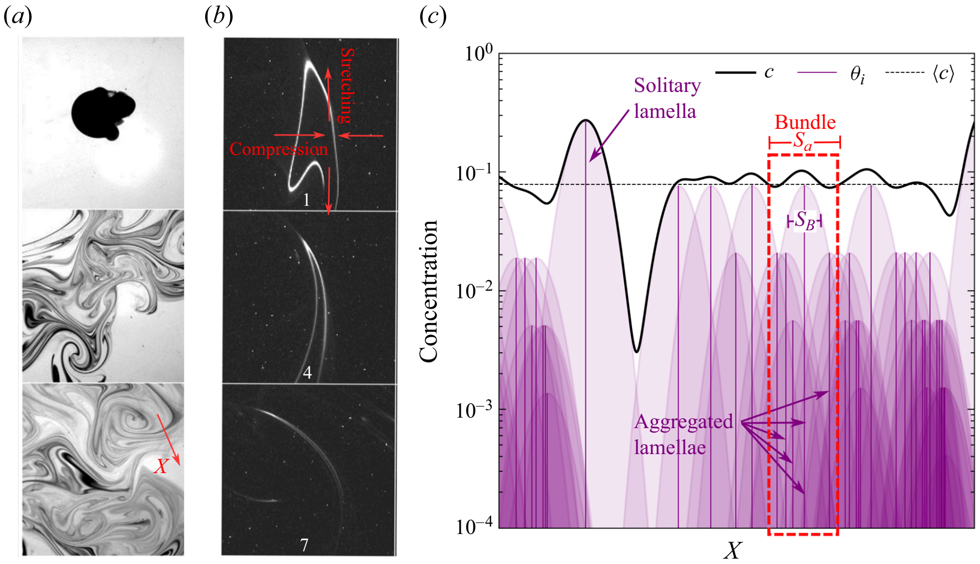

Figure 1. (*a*) Mixing of a diffusive scalar by a random stirring protocol (time sequence top to bottom), evidencing the apparition of stretched scalar filaments (adapted from Villermaux 2012). (*b*) Blow-up on the coalescence of neighbouring filaments under the action of compression (adapted from Duplat & Villermaux 2008). (*c*) Concentration profile of a scalar field showing the coexistence of solitary filaments and bundles of filaments. The scalar concentration c is obtained by the superposition of individual lamellae in a bundle of size s_a. All lamellae have a Gaussian shape with decaying maximum concentration θ_i (see (2.14)) and width tending to s_B.

with κ the molecular diffusivity. While advected by the flow, the solute also mixes with its surrounding due to the irreversible effect of molecular diffusion, and concentration tends to homogenize.

Although fully linear, (1.1) is characterized by rich spatio-temporal behaviours that have garnered considerable attention over the years (Rothstein, Henry & Gollub 1999; Warhaft 2000; Falkovich, Gaweęzdki & Vergassola 2001). As illustrated in figure 1(*a*), an initial blob of passive solute repeatedly stirred in a chaotic flow produces scalar filaments that stretch exponentially in time (Aref 1984). Their thickness is blocked at the so-called Batchelor scale, at which molecular diffusion compensates the effect of stretching on scalar gradients (Batchelor 1959). Owing to flow incompressibility, these filaments are also exponentially compressed onto each other, leading to their diffusive coalescence (figure 1*b*). These three processes – stretching, diffusion and coalescence – are key to quantifying and predicting the statistics of c in the homogenizing mixture.

Filament stretching and diffusion can be quantified effectively in a Lagrangian frame aligned with the local directions of elongation and compression of the flow field (Ranz 1979; Balkovsky & Fouxon 1999). Such an approach, referred to as Lagrangian stretching or lamellar theory, provides a theoretical link between the stretching statistics of the flow (its advective properties) and the evolution of concentration in isolated lamellae. It yields good estimates of the decay of the scalar concentration variance (Haynes & Vanneste 2005; Tsang, Antonsen & Ott 2005) and of the shape of the scalar probability density function (p.d.f.) in a range of chaotic flows (Meunier & Villermaux 2010). However, it is *a priori* limited to the early stage of mixing, when scalar filaments evolve far from each other, well before coalescence (Fereday *et al.* 2002; Villermaux & Duplat 2003).

Interestingly, the Lagrangian stretching framework has been shown (Haynes & Vanneste 2005; Tsang *et al.* 2005) to provide correct mixing rate predictions even at asymptotic time, for smooth flow in the so-called Batchelor regime, that is, when scalar gradients persist only at small scales. Indeed, in periodic domains, asymptotic mixing rates have been shown to be controlled either 'locally' by the stretching of material lines, or 'globally' by macro-dispersive flow properties, depending on the domain size with respect to the velocity correlation length. In the 'local' case, it is still unknown why the Lagrangian stretching framework would remain accurate after coalescence time, when stretched filaments are not isolated from each other. In addition, it is unknown how stretching statistics control the shape of the asymptotic scalar p.d.f. Some theoretical results exist on the limiting shape of the scalar p.d.f. (Sinai & Yakhot 1989), but owing to closures that are not trivial to relate to physical properties of the flow (Sukhatme 2004). In turn, Pierrehumbert (1994) and followers (Rothstein *et al.* 1999; Sukhatme & Pierrehumbert 2002) have documented the self-similarity of asymptotic scalar p.d.f.s through the emergence of a 'strange eigenmode', but a physically based modelling of the shape of the scalar p.d.f. is still an open question.

Recognizing the additive nature of scalar coalescence in bounded domains, Villermaux & Duplat (2003) have proposed describing the late-time evolution of concentration p.d.f.s by a random aggregation mechanism, whereby the concentration field results from a sum of independent filaments that individually obey Lagrangian stretching dynamics, and that overlap by diffusion. This random aggregation scenario accurately captures the shape of asymptotic scalar p.d.f.s (Duplat & Villermaux 2008; Meunier & Villermaux 2010) if one uses the rate of filament aggregation as a fitting parameter. However, no independent measure of filament aggregation in chaotic flows has been obtained so far to validate the random aggregation hypothesis. In addition, it is still unclear if aggregation processes tend to accelerate or decelerate mixing compared to stretching.

Recently, Heyman, Lester & Le Borgne (2021) have proposed that scalar aggregation could obey a correlated process, where poorly stretched regions of the flow containing weakly mixed solutes remain weakly aggregated in proportion. Such intimate correlation between stretching and aggregation has the advantage of explaining the persistent role of Lagrangian stretching statistics on asymptotic scalar mixing that is largely observed in the Batchelor regime (Tsang *et al.* 2005).

Here, we develop a general aggregation theory for scalar mixing in smooth, two-dimensional and fully chaotic flows in the Batchelor regime. The theory is based on the description of the spatial clustering of advected material lines, and its correlation with local elongation rates. The theory allows us to describe the asymptotic evolution of the scalar p.d.f. and its moments at any Péclet number based on the sole knowledge of Lagrangian stretching statistics and fractal dimensions of the advective material lines. It may be generalized to a wider range of flows and boundary conditions.

The paper is organized as follows. In § 2, we recall important concepts and variables pertaining to Lagrangian stretching and aggregation frameworks. In § 3, we present the two-dimensional incompressible chaotic flows used to illustrate and validate the general aggregation theory, and their numerical resolution. In § 4, we focus on the geometry of material lines advected by a chaotic flow, and we derive the evolution of the p.d.f. of the number of advective filaments in aggregates, as well as the statistics of elongation in aggregates. In § 5, we describe how the scalar concentration p.d.f. and its moment can be obtained from the knowledge of both the distribution of aggregate size and filaments statistics in aggregates.

2. Lagrangian stretching and aggregation

2.1. *Stretching and dilution*

We focus on smooth, two-dimensional, incompressible and time-varying flows, for which there exist two Lyapunov exponents of opposite signs. Advection thus creates elongated one-dimensional structures, called lamellae or filaments, that are the backbone of scalar mixing.

The temporal evolution of single scalar filaments is trackable in a Lagrangian frame with a coordinate system (x, y) advected with the flow and aligned with the directions of compression (x) and elongation (y) (Ranz 1979; Villermaux 2019). Because of their elongated shape, the concentration of lamellae is approximately constant in the y direction. Thus $\partial_y c$ is negligible compared to $\partial_x c$, and the two-dimensional advection–diffusion problem (1.1) simplifies to a one-dimensional advection–diffusion equation

$$\partial_t c + u(x)\,\partial_x c = \kappa\,\partial_x^2 c, \tag{2.1}$$

with $u = -x\,\lambda(t)$ the velocity at which solute particles are compressed in the x direction, and $\lambda(t) \geq 0$ the stretching rate. In two-dimensional incompressible flows, the stretching rate $\lambda(t)$ in the y direction leads to a compression rate $-\lambda(t)$ in the x direction. In chaotic flows, the mean stretching rate μ_λ is positive and equal to the Lyapunov exponent. Note that the notation

$$\mu_\bullet = \langle \bullet \rangle \tag{2.2}$$

is used throughout the paper for a mean quantity. We define the total lamellar elongation as

$$\rho(t) = \exp\left(\int_0^t \lambda(t')\,\mathrm{d}t'\right). \tag{2.3}$$

Ranz (1979) showed that (2.1) transforms to the simple diffusion equation

$$\partial_\tau c = \partial_\xi^2 c \tag{2.4}$$

if time is rescaled by

$$\tau(t) = \frac{\kappa}{s_0^2}\int_0^t \rho(t')^2\,\mathrm{d}t', \tag{2.5}$$

and space is rescaled by $\xi = x\rho/s_0$, with s_0 the initial lamellar width. For a Gaussian initial condition $c(\xi, 0) = \theta_0 \exp(-\xi^2)$, the solution of (2.3) is

$$c(\xi, \tau) = \frac{\theta_0}{\sqrt{1+4\tau}} \exp(-(\xi/\sqrt{1+4\tau})^2). \tag{2.6}$$

Equivalently, in the original Lagrangian coordinate system (x, y),

$$\Theta(x, t) = \theta(t) \exp(-(x/s(t))^2), \tag{2.7}$$

where θ is the maximum concentration of the lamella, following

$$\theta(t) = \frac{\theta_0}{\sqrt{1+4\tau(t)}}, \tag{2.8}$$

and s is the unit lamellar thickness, following

$$s(t) = \frac{s_0\sqrt{1+4\tau(t)}}{\rho(t)}. \tag{2.9}$$

Note that the scalar mass per unit length under a lamella simplifies to

$$m(t) = \sqrt{\pi}\,\theta\, s(t) = \sqrt{\pi}\, s_0\theta_0\, \rho^{-1}(t). \tag{2.10}$$

Since in random flows the stretching rate is a random variable of time, the elongation of lamellae and the rescaled time are also randomly distributed. An approximation of the statistics of τ was proposed (Meunier & Villermaux 2010; Lester, Dentz & Le Borgne 2016) as

$$\tau \approx \frac{\kappa}{2s_0^2}\,\frac{t}{\log \rho}\,(\rho^2 - 1), \tag{2.11}$$

recognizing the fact that the last stretching events have a predominant weight in the stochastic integral (2.4). At large times, $\rho \gg 1$ and $\log \rho / t \to \mu_\lambda$. Thus

$$\tau \to \frac{\kappa}{2\mu_\lambda s_0^2}\,\rho^2 = \frac{1}{4}\left(\frac{s_B}{s_0}\,\rho\right)^2, \tag{2.12}$$

with the so-called Batchelor scale

$$s_B = \sqrt{2\kappa/\mu_\lambda}. \tag{2.13}$$

Note that in his seminal paper, Batchelor (1959) uses the mean turbulent strain rate instead of the Lyapunov exponent μ_λ to quantify fluid stretching. Thus filaments dilute at large times, with a peak concentration following

$$\theta(t) \to \frac{\theta_0 s_0}{s_B}\,\rho^{-1}(t), \tag{2.14}$$

while their thickness tends to the Batchelor scale $s \to s_B$. In contrast, the length of filaments grows as

$$L(t) = \ell_0\, \mu_\rho(t), \tag{2.15}$$

where ℓ_0 is the initial filament length. Equation (2.14) shows that the concentration of single diffusing filaments is inversely proportional to their elongation. Thus describing the advective stretching statistics is sufficient to describe the mixing of passive scalars. Indeed, the spatial variance of an elongated material line in a flow domain of area $\mathcal{A}$ can be obtained by integration of the lamellar concentration (2.14) squared along the filament path (Meunier & Villermaux 2010), giving

$$\sigma_c^2 \sim \frac{1}{\mathcal{A}}\, L(t)\, \mu_\rho^{-2} = \frac{1}{\mathcal{A}}\, \mu_\rho^{-1}. \tag{2.16}$$

Assuming a normal distribution of stretching with mean $\mu_\lambda t$ and variance $\sigma_\lambda^2 t$, the scalar variance decays can be estimated from the stretching statistics. For flows with moderate stretching variability, where $\sigma_\lambda^2 \leq \mu_\lambda$, we have

$$\sigma_c^2 \sim \exp(-(\mu_\lambda - \sigma_\lambda^2/2)t). \tag{2.17}$$

In contrast, for flows with larger stretching variability where $\sigma_\lambda^2 \geq \mu_\lambda$, the value of μ_ρ^{-1} is entirely controlled by small elongations with $\rho \leq 1$ (Balkovsky & Fouxon 1999). In that case, $\mu_\rho^{-1} \sim \exp(-\mu_\lambda^2/(2\sigma_\lambda^2)t)$.

It is important to note that the Lagrangian stretching approximation is correct in the limit of one-dimensional filaments, that is, when the curvature radius is much larger than the diffusing filament width s_B. In bounded flow domains, there exist regions where high curvatures develop (Tang & Boozer 1996), which are known (Thiffeault 2004) to be associated with weak stretching rates (low elongations) and thus low mixing (high concentrations). This questions the ability of the Lagrangian stretching framework to capture the tails of the scalar concentration p.d.f. However, at small s_B (small κ), the impact of these regions on global mixing rates should remain limited. This is confirmed by the good agreement of the Lagrangian stretching theory with observed mixing rates (Haynes & Vanneste 2005).

2.2. *Aggregation*

The production of curvature tends to produce folds in the material line (figure 1*b*). The distance between separated parts of the elongated material line is thus reduced exponentially via fluid compression. This mechanism creates a highly foliated structure at late time (figure 1*a*). Individual filaments are thus no longer isolated, and start to coalesce on length scales comparable to their width s_B. This so-called aggregation regime (Villermaux & Duplat 2003) has two properties.

First, filaments tend to accumulate locally due to exponential flow compression. Bundles of aggregated lamellae are thus formed by individual filaments sharing the same diffusive neighbourhood of size $\sim s_B$. The time at which aggregation initiates is when the total area of a lamella of length $L(t)$ and width s_B is greater than the available area of the flow domain $\mathcal{A}$ (Garrett 1983), e.g.

$$L(t)\, s_B \gtrsim \mathcal{A}. \tag{2.18}$$

The mean number of lamellae n in each bundle is thus

$$\mu_n(t) \sim \frac{L(t)\, s_B}{\mathcal{A}}. \tag{2.19}$$

Assuming a constant stretching rate λ, the length of the lamella is $L(t) = \ell_0 \exp(\lambda t)$, and the coalescence time t_c at which (2.18) is first fulfilled is

$$t_c \sim \frac{1}{\lambda} \log\left(\frac{\mathcal{A}}{\ell_0 s_B}\right). \tag{2.20}$$

Second, the linearity of the advection–diffusion equation implies that scalar concentration fields can be decomposed into a sum over the concentration profiles of solitary lamellae (Le Borgne *et al.* 2017). The scalar concentration c at a given position can be constructed as the sum of concentrations θ_i of individual lamellae i (and defined by (2.14)) present in a small neighbourhood around this position, of size s_a comparable to the asymptotical size of the lamellae, s_B, that is,

$$c(t) \sim \sum_{i=1}^{n(t)} \theta_i(t) \sim \sum_{i=1}^{n(t)} \rho_i^{-1}(t), \tag{2.21}$$

where the asymptotic relation (2.14) between θ and ρ was used. Through (2.21), the aggregation process offers an appealing way to model the statistics of the scalar concentration field c via knowledge of the aggregated number of filaments n at the aggregation scale s_a, their elongations ρ_i, and the possible correlations between these two variables.

Assuming no correlations between n and ρ_i, and among ρ_i, leads to a random aggregation scenario. The scalar concentration c in a bundle is thus formed by the sum of independent and identically distributed random variables, following the solitary filament concentration p.d.f. The scalar concentration p.d.f. $P_c(c, t)$ thus results from the n-convolution of the isolated lamella concentration p.d.f. $P_{1/\rho}(\rho^{-1}, t)$, with the mean number of aggregations μ_n given by (2.19) (Duplat & Villermaux 2008). If $P_{1/\rho}$ is exponential or gamma distributed, then P_c is a gamma distribution. Note that if $P_{1/\rho}$ is log-normal, then the n-convolution p.d.f. is not explicit, although its moments are given by the central limit theorem (Schwartz & Yeh 1982). This random conjecture was shown (Duplat & Villermaux 2008) to correctly describe mixing in turbulent flows, for which the velocity cascade over a large range of scales favours the decorrelation of filament stretching histories.

Below the characteristic velocity length scale (the Batchelor regime), however, filament aggregation seems not to obey completely random dynamics. To see that, it is useful to consider the decay of scalar variance. In the random hypothesis, by the central limit theorem, scalar variance decays as the inverse of the aggregation number, that is, the inverse of the material length (2.19):

$$\sigma_c^2 = \frac{\mu_c^2}{\mu_n} \sim 1/L(t). \tag{2.22}$$

Assuming a normal distribution of stretching rates, $L \sim \exp((\mu_\lambda + \sigma_\lambda^2/2)t)$ and thus the asymptotic scalar variance decay exponent is $\mu_\lambda + \sigma_\lambda^2/2$, larger than the decay exponent of solitary strips ($\mu_\lambda - \sigma_\lambda^2/2$, (2.17)). This result contradicts observations suggesting the same decay exponent before and after aggregation time in the Batchelor regime (Fereday *et al.* 2002; Tsang *et al.* 2005).

We attribute this failure to the existence of correlations in the aggregation process below the characteristic velocity length scale. Indeed, in incompressible flows, lamellar elongation is always balanced with transverse compression (figure 1*b*), which attracts neighbouring lamellae onto each other. Thus highly stretched lamellae tend to be highly aggregated. Conversely, weakly stretched lamellae have also experienced little compression, thus remaining isolated from the bulk, and their evolution well described by the isolated lamellar theory. Since these lamellae bear high concentration levels, they must dominate the statistics of scalar fluctuations at late time. Such correlations may explain why the Lagrangian stretching prediction for the scalar dissipation remains accurate long after aggregation time.

Evidence of correlated aggregation was also observed experimentally during the chaotic mixing of two dye blobs (Duplat, Jouary & Villermaux 2010). If injected in a concentric manner, the blobs locally experience similar stretching rates and aggregate in a correlated manner. In contrast, when placed at a certain distance from each other (larger than the characteristic length scale of the velocity field), their concentrations obey random aggregation rules. This observation suggests that stretching and aggregation are strongly correlated below the characteristic velocity length scale, while they become uncorrelated above this scale.

The remainder of the paper aims to establish the laws governing correlated aggregation for scalar mixing by smooth chaotic flows, and to deduce their impact on the evolution of scalar concentration p.d.f.

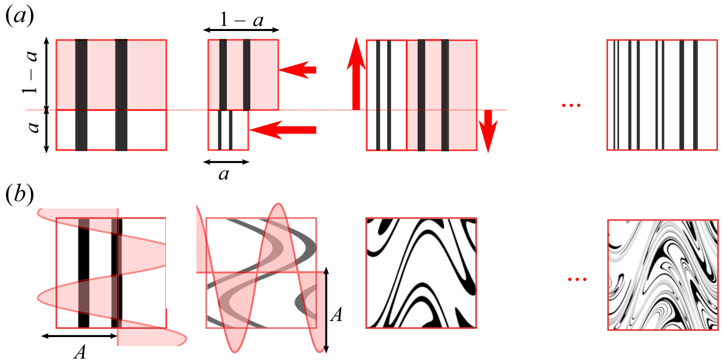

Figure 2. (*a*) Transformations operated by the incompressible baker map with parameter a. First, the domain is cut horizontally at $y = a$, where a is a constant between 0 and 0.5. Uniform fluid compression operates on the domain parts $y < a$ and $y > a$ with a and $1 - a$, respectively. Then vertical stretching occurs with a factor $1 - a$ and a in these two regions, preserving the total area. (*b*) Transformations operated by the sine flow with amplitude A. The flow is an alternation of horizontal and vertical sinusoidal velocity waves with amplitude A and period 2π. Random phases are chosen at each time period so that the flow is fully chaotic.

3. Numerical simulation of synthetic chaotic flows

We focus on scalar mixing in the so-called Batchelor regime (Haynes & Vanneste 2005) where the velocity field is smooth, and present a single length scale, much greater than the typical scalar fluctuation scale. The domain size and the velocity length scale are of comparable magnitude, $\mathcal{L} \sim \mathcal{L}_v \sim 1$ – the so-called 'local' mixing regime described in Haynes & Vanneste (2005). Thus no scalar gradients can persist at scales larger than the velocity scale.

To illustrate the geometrical features of aggregation, we use two synthetic chaotic transformations that comply with these conditions, namely the baker map and the random sine flow (figure 2). These are simple sequential advective maps that have been widely used to investigate the properties of chaos (Finn & Ott 1988; Ott & Antonsen 1989; Giona, Cerbelli & Adrover 2001; Tsang *et al.* 2005; Meunier & Villermaux 2010, 2022). We recall their definitions below.

3.1. *Incompressible baker map*

The incompressible baker map (Ott & Antonsen 1989; Wonhas & Vassilicos 2002) is a discontinuous transformation that operates on a two-dimensional periodic domain $[0, 1] \times [0, 1]$. The transformation is

$$\left.\begin{aligned} x_{t+1} &= \begin{cases} a x_t & \text{if } y_t < a, \\ 1 - (1-a) x_t & \text{if } y_t > a, \end{cases} \\ y_{t+1} &= \begin{cases} y_t / a & \text{if } y_t < a, \\ (1 - y_t)/(1-a) & \text{if } y_t > a, \end{cases} \end{aligned}\right\} \tag{3.1}$$

where $a \in [0, 0.5]$ is a parameter controlling the heterogeneity of the map. A visual sketch of the map operation is shown in figure 2*a*.

An advantage of the baker map is that purely vertical scalar patterns (for which $c(x, y) = f(x)$) remain one-dimensional after application of the map, thus simplifying the problem to a single dimension. This simplicity allows for the analytical derivation of many features of the map, as we will show later. Another advantage is that it is possible to explore a wide range of stretching heterogeneity by varying a between 0 and 0.5. Indeed, the first two moments of stretching rate in the baker map are

$$\mu_\lambda/t = -a\log(a) - (1-a)\log(1-a), \tag{3.2}$$

$$\sigma_\lambda^2/t = a(1-a)(\log(1-a) - \log(a))^2. \tag{3.3}$$

Thus for $a = 0.01$, $\sigma_\lambda^2/\mu_\lambda = 3.7$, while for $a = 0.49$, $\sigma_\lambda^2/\mu_\lambda = 5.7 \times 10^{-4}$. It is important to note that this map involves discontinuous transformations, or ‘cuts’, that are absent in continuous flows such as turbulence but are common in flows through porous media (Lester, Metcalfe & Trefry 2013). This map also prevents the formation of folds and cusps, where the one-dimensional hypothesis of the Lagrangian stretching framework is violated. The correspondence between (1.1) and (2.1) is thus exact in this map.

3.2. *Random sine flow*

The random sine flow (Pierrehumbert 1994; Haynes & Vanneste 2005) is a continuous transformation operating on a periodic domain [0, 1] × [0, 1] (figure 2*b*). The flow is periodic in time and space, and reads for a given time period t as

$$\left.\begin{aligned} y_{t'+\delta t} &= y_{t'} + A\,\delta t \begin{cases} \sin(2\pi x'_t + \phi_t) & \text{for } t < t' < t+1/2, \\ 0 & \text{for } t+1/2 < t' < t+1, \end{cases} \\ x_{t'+\delta t} &= x_{t'} + A\,\delta t \begin{cases} 0 & \text{for } t < t' < t+1/2, \\ \sin(2\pi y'_t + \psi_t) & \text{for } t+1/2 < t' < t+1, \end{cases} \end{aligned}\right\} \tag{3.4}$$

where the amplitude A is a positive constant, ϕ_t, ψ_t are random phases that change at each time period t, and δt is the time step. The flow velocity having a single component, incompressibility is enforced. Scalar transport is continuous and considered on a periodic domain [0, 1] × [0, 1], ensuring a local control of mixing rates (Haynes & Vanneste 2005).

As most random chaotic flows, the elongation of material lines in sine flows approximately follows a log-normal distribution with parameters $\mu_\lambda t$ and $\sigma_\lambda^2 t$ that depend on the amplitude A. The stretching heterogeneity is much less variable than in the baker map, with ratio $\sigma_\lambda^2/\mu_\lambda$ ranging from 1 when $A \to 0$ to $\sigma_\lambda^2/\mu_\lambda \approx 0.6$ for $A = 1.8$. The stretching statistics of random sine flows can be found in Meunier & Villermaux (2022). In Appendix A, we recall useful results concerning the distribution of filament elongation and its moments in the sine flow and the baker map.

3.3. *Numerical methods*

The hypothesis used throughout the paper is that the concentration field obeying (1.1) is well approximated by a local summation over elementary lamellar concentrations, each of them individually following (2.1). This implies that the statistics of concentrations and their temporal evolution can in principle be inferred from the statistics of lamellar elongation and aggregation. Thus instead of directly solving the two-dimensional advection–diffusion equation (1.1), our approach consists in computing elongation of a deforming Lagrangian support, advected by the flow (figure 3).

Numerically, we advect a material line of initial length ℓ_0 in the velocity field. In practice, the material line is defined by a series of particles $\boldsymbol{x}_i$ linked by segments. In the alternated random sine wave, the positions of particles are tracked through time t via

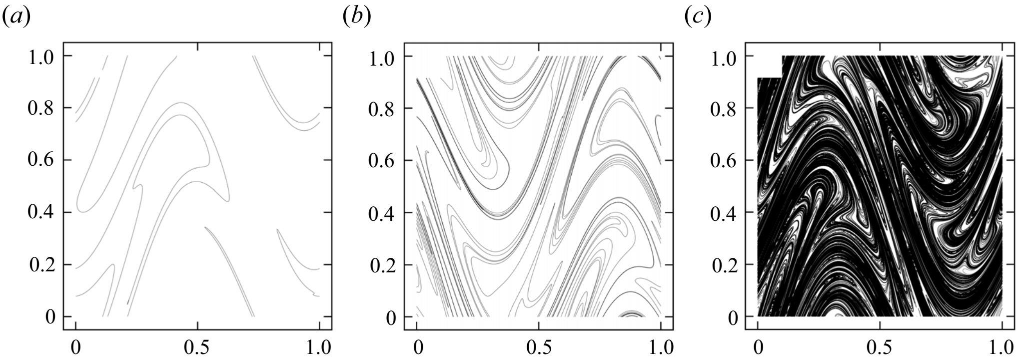

Figure 3. Advection of a material filament in the sine flow with $A = 0.8$, for (*a*) $t = 3$, (*b*) $t = 5$ and (*c*) $t = 10$.

an explicit Euler scheme (figure 3). The particles have constant velocities during a time step $\mathrm{d}t = 1/2$, so this scheme is exact. The local elongation of segments is estimated as

$$\rho_i(t) = \rho_i(t - \mathrm{d}t)\, \frac{\|\boldsymbol{x}_i(t + \mathrm{d}t) - \boldsymbol{x}_{i+1}(t + \mathrm{d}t)\|}{\|\boldsymbol{x}_i(t) - \boldsymbol{x}_{i+1}(t)\|}. \tag{3.5}$$

The material line is refined after each time step to maintain a maximum local distance between points of $\mathrm{d}x \leq 10^{-3}$.

In the baker map, we start with a single filament aligned with the y direction, so that we need to track only its coordinate x. Each operation of the map doubles the number of filaments, with the first half being elongated by a factor $1/a$, and the second half by a factor $1/(1 - a)$. We thus keep track of the growing number of filament positions x_i and their elongations ρ_i.

The elongated and folded filament is tracked up to the advection time where the total length is $L = 10^7 \ell_0$, a limit corresponding to our computer memory. Local elongation of the advected material line can then be used to compute its local concentration, via (2.14). Aggregated scalar fields can then be estimated by a local summation of individual lamellar concentrations, via (2.21).

3.4. *Reconstruction of scalar fields*

As shown by Meunier & Villermaux (2010), it is possible to reconstruct the aggregated scalar field via the superposition of elementary lamellae, represented as Gaussian ellipsoids of short axis s, that swipe along the advected material line. This reconstruction is exact in the baker map since no cusps are forming. In the sine flow, it is only approximate in the regions of high curvature, where diffusion is genuinely two-dimensional.

Here, we take an alternative reconstruction approach, which has the advantage of allowing analytical treatment. At late time, lamella widths tend to s_B (2.13), which is the minimum scale of fluctuations of the scalar field (Batchelor 1959). Thus we construct the aggregated scalar field by binning lamellar concentrations on a regular grid of size $s_a \sim s_B$. Consider a box of surface s_a^2 centred in position $\boldsymbol{x}$. The aggregated concentration level in this box can be constructed from the sum of the masses per unit length m_i of the $n(\boldsymbol{x})$

individual lamellae of length ℓ_i present in this box:

$$c(\boldsymbol{x}) \approx \frac{1}{s_a^2}\sum_{i=1}^{n(\boldsymbol{x})} s_a m_i = \frac{\sqrt{\pi}\,\theta_0 s_0}{s_a}\sum_{i=1}^{n(\boldsymbol{x})} \rho_i^{-1}, \tag{3.6}$$

where we used (2.9) for the evolution of the solute mass per unit length carried by an individual lamella at a given location. Equation (3.6) forms the base of the statistical description of aggregation intended in this paper.

In practice, we choose

$$s_a = \sqrt{2\pi}\, s_B \tag{3.7}$$

for the variance of the regular grid reconstruction to match the variance of the true field (i.e. the one obtained by aggregating Gaussian lamellae). The factor $\sqrt{2\pi}$ is obtained so that both reconstruction methods provide the same mean-squared concentration in the case of a single lamella.

Note that the aggregated concentration field can be deduced at various Péclet numbers by varying the aggregation scale s_a (3.7), without recomputing the advective filament position. We define the Péclet number as the ratio of diffusive to stretching time scales, e.g.

$$Pe = \mu_\lambda \mathcal{L}_v^2/\kappa. \tag{3.8}$$

Through (2.13), $s_B = \mathcal{L}_v\sqrt{2/Pe}$ and thus $s_a = 2\sqrt{\pi}\,\mathcal{L}_v/Pe$.

This regular grid approximation of the aggregated scalar field disregards the exact Gaussian shape of lamellar concentrations and the variations of their width s (2.5). However, classical box counting methods and fractal dimensions can be used to describe the geometrical and statistical features of the aggregated scalar field. Our goal is to describe the joint statistics of the variables ρ_i and n arising in (3.6), to infer the statistics of c.

The approximation (3.6) is compared to direct numerical simulations (DNS) of the advection–diffusion equation for different Péclet numbers in figure 4. The DNS are obtained with the spectral method described in Meunier & Villermaux (2022), with a 2048^2 grid, and time step 0.1. There is a good qualitative match between the aggregated field and the DNS, a match confirmed by comparing the distributions of concentrations. Thus we conclude that the construction of the aggregated scalar field via (3.6) is able to capture the essential statistical features of the two-dimensional advection–diffusion problem, and offers a convenient way to explore its statistics.

4. Clustering properties of advected material lines

In incompressible flows, the stretching of material elements by velocity gradients is compensated by transverse compression, which causes distances between lamellar elements to decrease exponentially over time. Smaller and smaller scales are thus produced continuously by flow compression. Furthermore, in smooth chaotic flows, the typical scale of variation of velocity gradients is fixed and produces a heterogeneous stretching field for material lines. Dense (black) or diluted (white) regions of material lines are thus created at large scale in the chaotic flow (figure 5). Such heterogeneous structures then cascade to smaller scales under the action of net compression, thus creating a fractal set of one-dimensional objects (lines) clustered around their transverse direction. In two-dimensional incompressible flows, the Hausdorff dimension of this fractal set

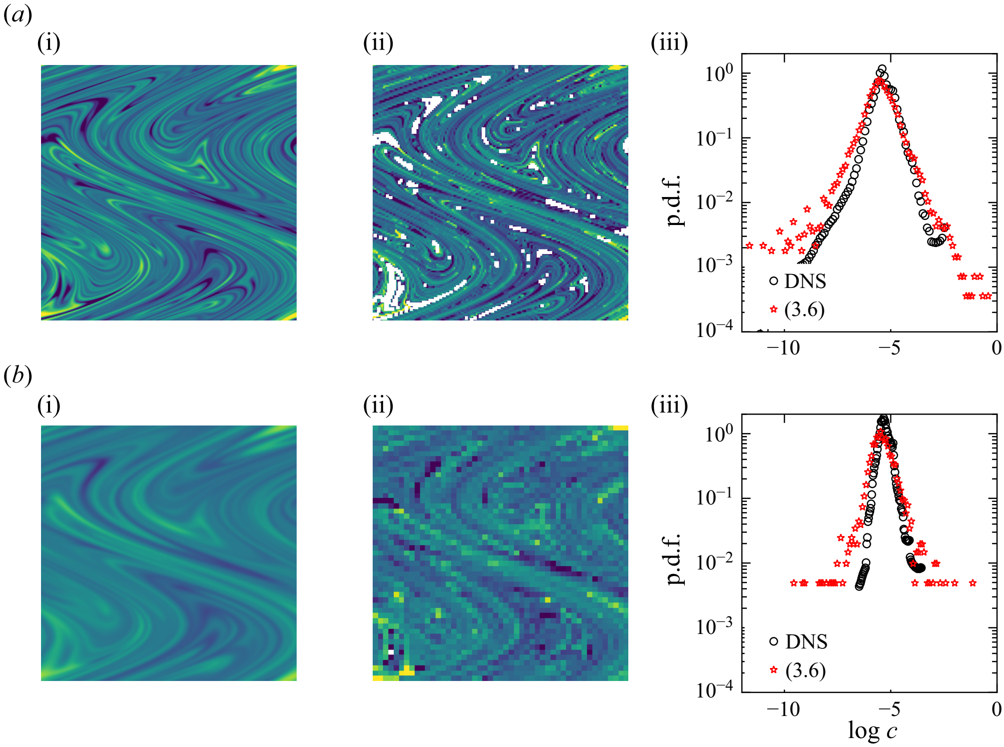

Figure 4. Comparison of log-concentration fields obtained at time $t = 10$ (fully aggregated regime): (i) with the aggregation framework (3.6), (ii) with direct numerical simulations (DNS) of (1.1), and (ii) with their p.d.f. The comparison is made for the sine flow ($A = 0.8$) for two Péclet numbers, corresponding to (*a*) $s_a = 1/150$ and (*b*) $s_a = 1/50$.

is necessarily $D_0 = 2$, as per the Kaplan–Yorke result (Farmer, Ott & Yorke 1983). Higher-order dimensions can be smaller than 2 if stretching is heterogeneous. To illustrate this, let us define a normalized measure p_k with $k = 1, \ldots, N$ defining a regular grid of bin size $\epsilon = \mathcal{L}/N$, where $\mathcal{L}$ is the domain size. For instance, p_k may be defined as the local density of lamellae in the bin, e.g. $p_k = n_k/n$, where n is the total number of lamellae.

The fractal dimension of order q of the measure p is obtained with (Grassberger 1983)

$$D_q - 1 = \lim_{\epsilon \to 0} \frac{1}{q-1} \frac{\log I_q(\epsilon)}{\log \epsilon}, \quad I_q(\epsilon) \equiv \sum_{k}^{N=\mathcal{L}/\epsilon} p_k^q, \tag{4.1a,b}$$

where the subtraction of 1 on the left-hand side accounts for the counting of one-dimensional structures (lamellae) in a two-dimensional domain. This definition implies the following spatial scaling of the integral of the measure:

$$I_q(\epsilon) \sim \epsilon^{(q-1)(D_q-1)}. \tag{4.2}$$

In simple flows such as the baker map, D_q can be obtained (Finn & Ott 1988) by observing the similarity properties of the map, which transfer at small scales the heterogeneity of the measure produced at large scales by a single operation of the map. Characterizing the result of one elementary operation of the map on the measure thus also informs us about the spectrum of fractal dimensions.

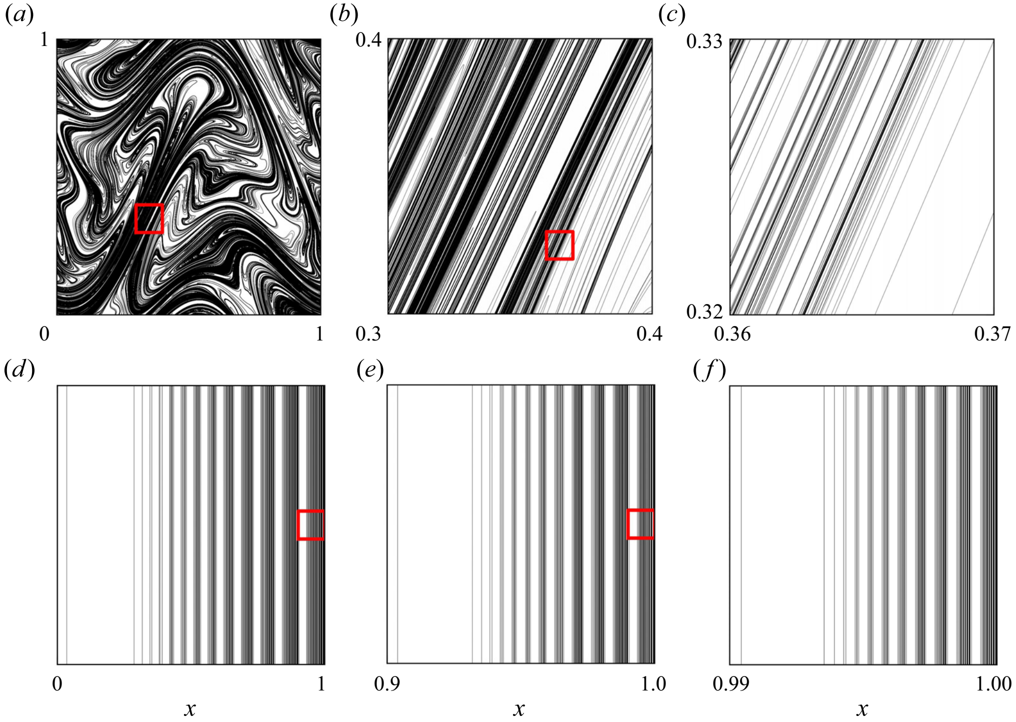

Figure 5. Fractal geometry of material lines in (*a–c*) the sine flow ($A = 0.8$) and (*d–f*) the baker map ($a = 0.1$), observed at different scales. The red square indicates the area selected for zooming.

We consider the measure of the local number of lamellae in bin k, $p_k = n_k/n$. As shown in figure 2(*a*), an operation of the baker map doubles the total number of these lamellae, while maintaining the same local distribution of lamellae on smaller bins of sizes $a\epsilon$ for $x < a$, and $(1 - a)\epsilon$ for $x > a$. This similarity allows one to compute the first fractal dimension explicitly (see Appendix B and Finn & Ott 1988):

$$D_0 = 2, \quad D_1 = 1 + \frac{2 \log 2}{\log(a^{-1} + (1 - a)^{-1})}. \tag{4.3a,b}$$

For random flows such as the sine flow, Ott & Antonsen (1989) argue that there exists a general relationship between stretching rate statistics and fractal dimensions as $D_q = f_q(\sigma^2_\lambda, \mu_\lambda)$, although a closed-form solution is not always as trivial as for the baker map. We show in figure 6 that the ratio $\sigma^2_\lambda/\mu_\lambda$ is directly related to the fractal dimension D_1. This suggests that the fractal clustering of material lines is closely linked to the large-scale heterogeneity of stretching rates. Since the flow is smooth, the heterogeneity created at large scales cascades to smaller scales, conserving its geometrical structure and creating a fractal geometry.

As suggested by figure 6, the function f_1 is different for the baker map and the sine flow. Indeed, the ratio $\sigma^2_\lambda/\mu_\lambda$ tends to a positive constant in the sine flow when $A \to \infty$, while $\sigma^2_\lambda/\mu_\lambda \to 0$ in the baker map when $a \to 0.5$. This finite limit comes from the fact that the sine flow is a continuous transformation with no cutting, thus does not tend to a uniform stretching rate. On the contrary, when $A \to 0$, $\sigma^2_\lambda/\mu_\lambda \to 1$, which is a maximum bound for the ratio in the sine flow (Meunier & Villermaux 2022), thus limiting the possible range

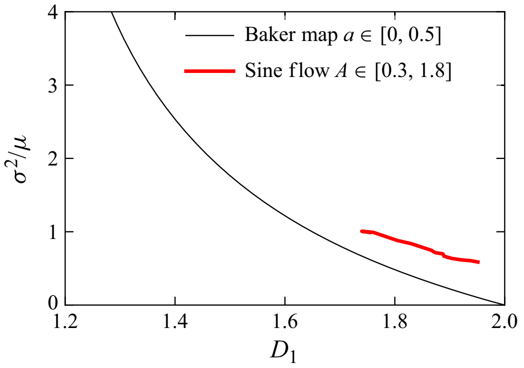

Figure 6. Relation between stretching rate mean μ_λ and variance σ_λ^2, and fractal dimension D_1, in the baker map and sine flow with varying parameters a and A.

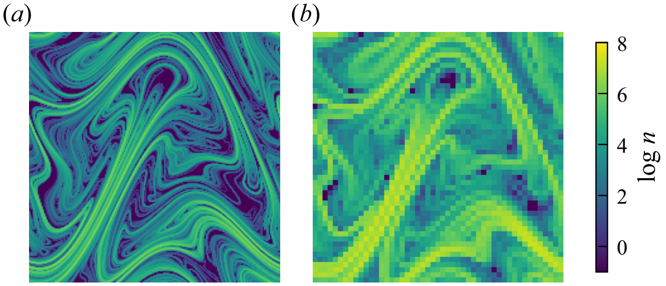

Figure 7. Spatial distribution of the number n of lamellae in bundles defined by a regular grid of size (*a*) $s_a = 1/200$ and (*b*) $s_a = 1/50$, in the sine flow with parameter $A = 0.5$.

of fractal dimensions produced by continuous chaotic flows, compared to discontinuous maps.

4.1. *Spatial distribution of n*

The spatial distribution of the number n of lamella per bin size $\mathcal{A}$ (figure 7) can be obtained as follows. Comparing the mean area occupied by a filament of length $L(t)$ and aggregation scale s_a to the domain surface $\mathcal{A}$, we get an estimate of the mean number of lamellae μ_n in bundles:

$$\mu_n \sim L(t)\, s_a/\mathcal{A}. \tag{4.4}$$

Higher moments can be obtained from a study of the fractal structure of material lines. To this end, we consider the spatial measure corresponding to the local number of lamellae in each bundle:

$$p_k \equiv \frac{n_k}{\sum\limits_k n_k} \tag{4.5}$$

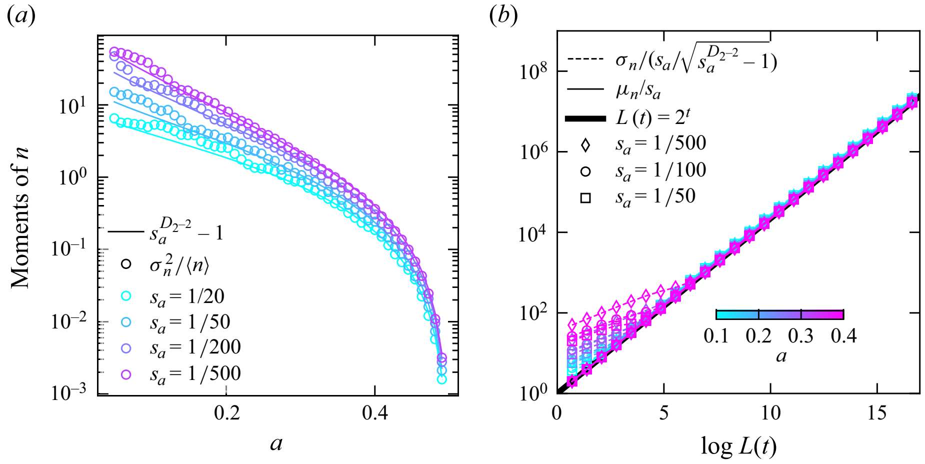

Figure 8. (*a*) Scaling of the spatial variance of $\log n$ as a function of a in the baker map and theoretical prediction, (4.10). (*b*) First two moments of P_n through time compared to theoretical predictions, (4.4) and (4.11) in the baker map ($\mathcal{A} = 1$).

The Renyi definition (Grassberger 1983) of the fractal dimension of order 2 of this measure is

$$D_2 - 1 \approx \frac{\log \sum_k p_k^2}{\log s_a}, \tag{4.6}$$

assuming $s_a \to 0$. Replacing (4.5) in the last expression provides

$$\sum_k \left(\frac{n_k}{\sum_k n_k} \right)^2 = s_a^{D_2-1}. \tag{4.7}$$

Since $\sum_k n_k = N\mu_n$, we have

$$\sum_i \left(\frac{n_k}{\sum_k n_k} \right)^2 = \frac{1}{N} \langle (n/\mu_n)^2 \rangle, \tag{4.8}$$

with $N \approx \sqrt{\mathcal{A}}/s_a$ the number of bundles in the flow domain. Since

$$\sigma^2_{n/\mu_n} = \langle (n/\mu_n)^2 \rangle - \langle n/\mu_n \rangle^2, \tag{4.9}$$

we have

$$\sigma^2_{n/\mu_n} = \sqrt{\mathcal{A}}\, s_a^{D_2-2} - 1. \tag{4.10}$$

Thus the variance of n/μ_n reaches a constant at asymptotic times, which is given by the fractal dimension of order 2. The spatial variance of n is then

$$\sigma_n^2 = \mu_n^2 \left(\sqrt{\mathcal{A}}\, s_a^{D_2-2} - 1 \right). \tag{4.11}$$

The predictions of (4.4)–(4.11) are plotted against time in figures 8 and 9 showing good agreement with simulations for a large range of aggregation scales s_a and flow

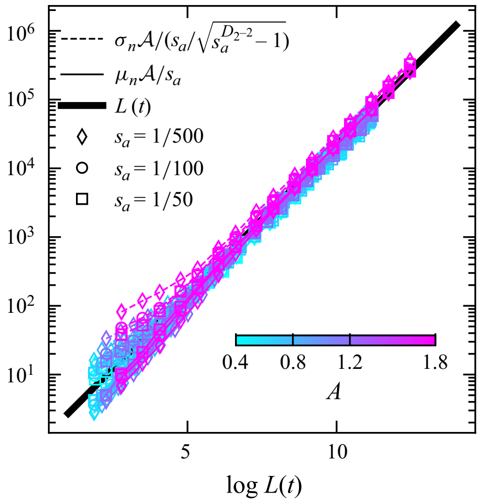

Figure 9. First two moments of P_n through time compared to theoretical predictions, (4.4) and (4.11) in the sine flow ($\mathcal{A} = 1$).

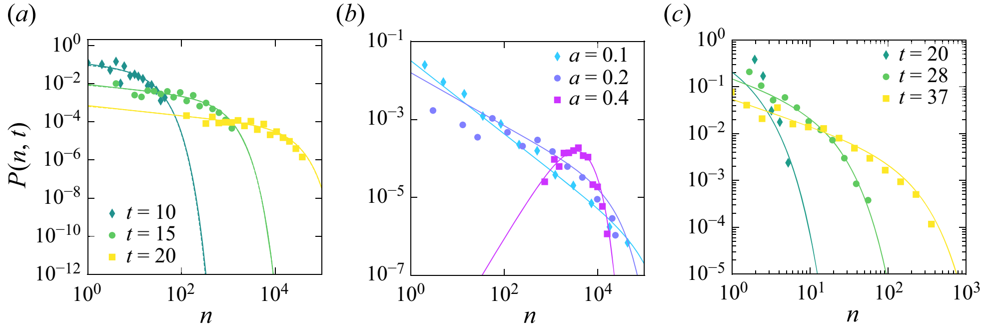

Figure 10. Plots of $P(n, t)$ for the baker map and sine flow for $s_a = 1/100$. Solid lines stand for the gamma p.d.f. with theoretical moments given by (4.14), and symbols stand for numerical simulations. (*a*) Baker map $a = 0.3$ and variable t. (*b*) Baker map for $t = 20$ and variable a. (*c*) Sine flow for $s_a = 1/100$ and $A = 0.4$.

heterogeneity, characterized by the parameters a for the baker map, and A for the sine flow. Higher moments of n can be obtained with similar scaling arguments and linked to fractal dimension of higher order. This is, however, out of the scope of the present study.

Instead, we observed that the p.d.f. of lamellar aggregation number $P_n(n)$ is well fitted by a gamma distribution (figure 10)

$$P_n(n) = \frac{1}{\Gamma(k_n)\,\theta_n^k}\, n^{k_n-1}\exp(-n/\theta_n), \tag{4.12}$$

with $n \geq 0$, and k_n, θ_n defined by the moments of the distribution of n:

$$k_n = \left(\sqrt{\mathcal{A}}\, s_a^{D_2-2} - 1\right)^{-1}, \tag{4.13}$$

$$\theta_n = \mu_n(t)\left(\sqrt{\mathcal{A}}\, s_a^{D_2-2} - 1\right), \tag{4.14}$$

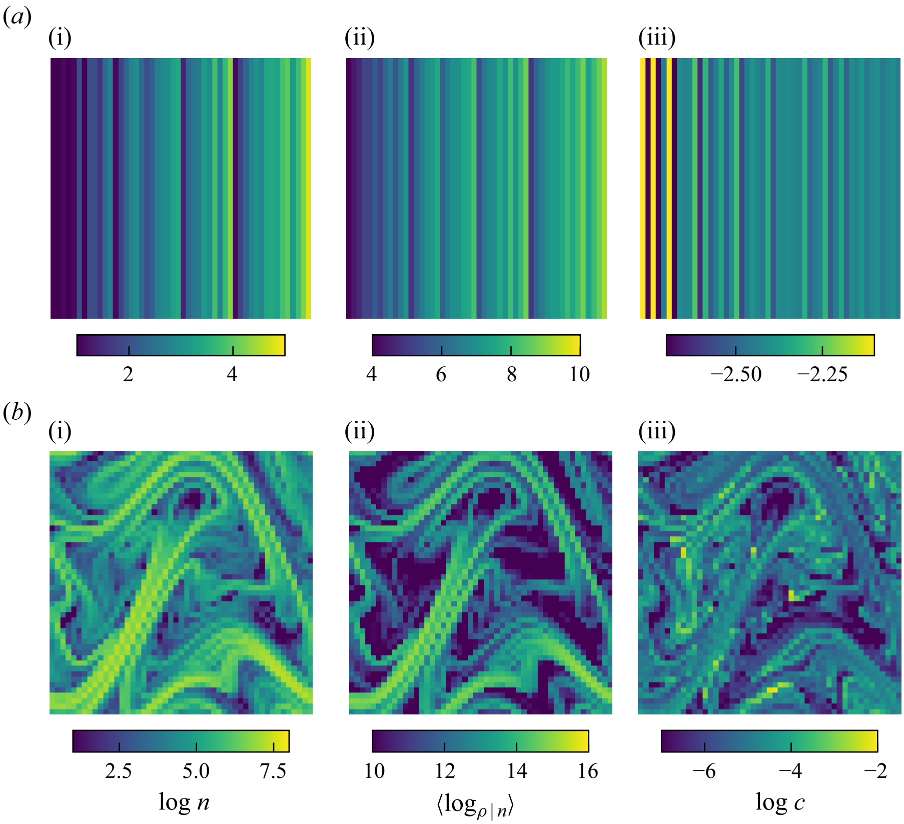

Figure 11. Simulation of aggregation statistics in the baker map ($a = 0.3$) and sine flow ($A = 0.5$) for $s_a = 1/50$: (i) number of lamellae n in bundles, (ii) mean of log-elongation in bundles, and (iii) sum of lamellar concentrations in bundles.

where $\mu_n = L(t)\, \mathcal{A}/s_a$. Note that the gamma distribution yields a power-law distribution at small n, with exponent $k_n - 1$ going from -1 to infinity with increasing D_2. For small D_2 (strong stretching heterogeneity), a significant part of the probability is concentrated at small n, thus in non-aggregated regions of the flow. We will show later that this can affects the value of negative moments of n.

4.2. *Local correlations between n and ρ*

Having given a precis of the statistics of the bundle size n, we now explore the local statistics of lamellar elongations in these bundles. To this end, we define the conditional averaging operator acting in lamellae located in the local neighbourhood of size s_a by

$$\mu_{X\,|\,n} = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{4.15}$$

where X is a Lagrangian variable transported by lamellae and n is the number of lamellae aggregated in the bundle. The spatial variability of averaged variables is shown in figure 11 for simulations in the baker map and sine flow. The remainder of this section is dedicated to uncovering the behaviour of conditional moments of elongation conditioned to n (e.g.

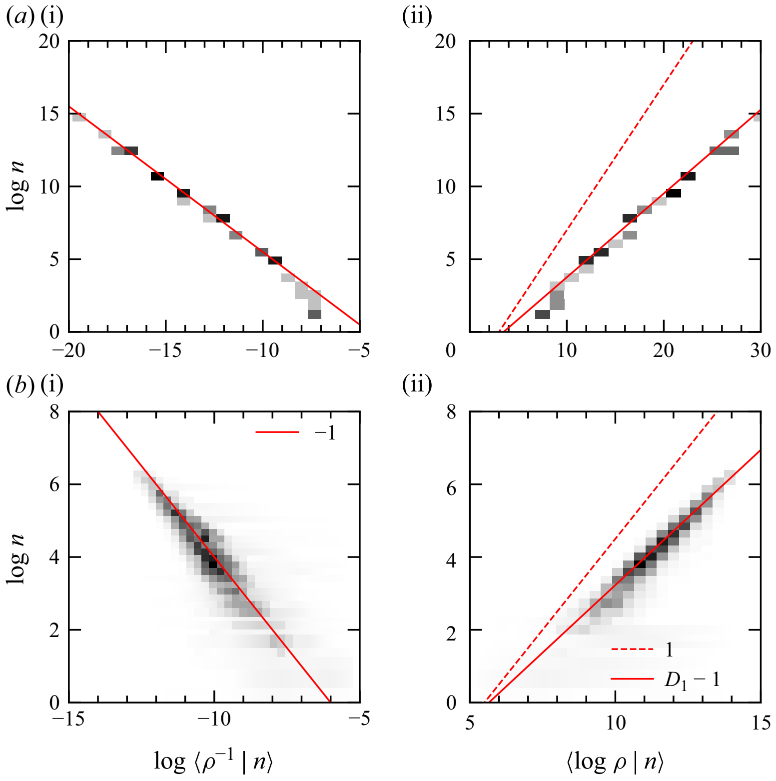

Figure 12. Joint p.d.f. (grey scale) of the number of lamellae in a bundle of size $s_a = 1/200$, and (i) their mean inverse elongation, (ii) their mean log-elongation, for (*a*) baker map ($a = 0.1$, $t = 24$, $D_1 = 1.57$) and (*b*) sine flow ($A = 0.8$, $t = 10$, $D_1 = 1.74$). The theoretical scalings of the measures (i) and (ii), given by (4.16) and (4.18), respectively, are plotted as solid red lines with the slope indicated in the legend. Dashed red lines are guides for the eye.

$\mu_{\rho^{-q}\,|\,n}$). Section 5 will then be dedicated to deriving unconditional probabilities by averaging on the distribution of n.

We plot in figure 12 the joint probability $P(n, \mu_{X\,|\,n})$ obtained in the baker map and the sine flow for $X = \rho^{-1}$ and $X = \log \rho$, the inverse of elongation and the log-elongation of lamella, respectively. Figure 12 suggests that the following scaling holds in both flows:

$$\log n \sim -\log \mu_{\rho^{-1}\,|\,n}, \tag{4.16}$$

which confirms the strong correlation between the number of lamellae in aggregates and their elongation. Indeed, for large times, c must tend to the conserved average scalar concentration $c \to \mu_c$. Thus, owing to (3.6), we must have

$$n \sim 1/\mu_{\rho^{-1}\,|\,n}, \tag{4.17}$$

a scaling that we confirm numerically (figure 12).

Figure 12 also suggests that

$$\log n \sim (D_1 - 1)\, \mu_{\log \rho\,|\,n}, \tag{4.18}$$

where D_1 is the information dimension (Ott & Antonsen 1989) of the measure n, given by (4.3*a,b*) for the baker map. Equation (4.18) can be derived analytically in the case of the baker map. Indeed, by the action of the map, the total number of lamellae increases as $\log n = t \log 2$, while the mean log-elongation of these lamellae is $\mu_{\log \rho} = t(-\log a - \log(1-a))/2$, leading to a constant ratio

$$\frac{\log n}{\mu_{\log \rho}} = \frac{2 \log 2}{\log a + \log(1-a)}, \tag{4.19}$$

which is exactly the value of $D_1 - 1$ (4.3*a*,*b*). Assuming that the partition between $\log n$ and $\mu_{\log \rho}$ is preserved at small scales in each bundle, we have

$$\log n = (D_1 - 1)(\mu_{\log \rho \,|\, n} - \mu_{\log \rho_c}), \tag{4.20}$$

with $\mu_{\log \rho_c}$ a constant standing for the mean elongation at coalescence time.

The constant $\mu_{\log \rho_c}$ can be estimated by comparing the surface covered by the filament at the aggregation scale, $\mathcal{S} = s_a \rho_c L_0$, with the domain area $\mathcal{A}$. The first aggregation event occurs when $\mathcal{S} \approx \mathcal{A}$, that is, when the elongation is $\rho_c \approx \mathcal{A}/(s_a \ell_0)$. Equation (4.20) is indeed verified for the baker map and sine flow with various parameters a and $\mathcal{A}$, with $\mu_{\log \rho_c} = -\log s_a$.

4.3. *Distribution of* $\log \rho$ *in a bundle of size* n

The two scaling laws $n \sim \mu_{\rho^{-1} \,|\, n}$ and $\log n \sim (D_1 - 1)\, \mu_{\log \rho \,|\, n}$ provide key information about the heterogeneity of lamellar elongations inside bundles. Since the ensemble distribution of elongation $P_\rho(\rho)$ has a log-normal shape, we assume that the distribution of elongations inside bundles, denoted $P_{\rho \,|\, n}$, is also log-normally distributed. This implies that $\log \rho$ is normally distributed in bundles, with mean

$$\mu_{\log \rho \,|\, n} \sim (D_1 - 1)^{-1} \log n. \tag{4.21}$$

Since $\log \mu_{\rho^{-1} \,|\, n} = -\mu_{\log \rho \,|\, n} + \sigma^2_{\log \rho \,|\, n}/2 \sim -\log n(\boldsymbol{x})$, the variance of log-elongation in bundles at large n must be

$$\sigma^2_{\log \rho \,|\, n} \sim \frac{2(2 - D_1)}{D_1 - 1} \log n. \tag{4.22}$$

We report in figure 13 the simulated scaling $\mu_{\log \rho \,|\, n}/\log n$ and $\sigma^2_{\log \rho \,|\, n}/\log n$ obtained asymptotically at large times. When $D_1 \to 2$, $\mu_{\log \rho \,|\, n}/\log n \to 1$ while $\sigma^2_{\log \rho \,|\, n}/\log n \to 0$, meaning that bundles are formed by lamellae of identical elongations. In contrast, when $D_1 \to 1$, both $\mu_{\log \rho \,|\, n}/\log n$ and $\sigma^2_B/\log n$ become infinite, while their ratio is $\sigma^2_{\log \rho \,|\, n}/\mu_{\log \rho \,|\, n} = 2(2 - D_1) \to 0.5$. This limit suggests that the aggregation of lamellae remains correlated to their average elongation, although a fixed amount of stretching variability arises in bundles. Good agreement is found between theoretical prediction (4.21)–(4.22) and numerical simulations of aggregation in the baker map (figure 13). In contrast, the theory captures only qualitatively the behaviour of the random sine flow. This may be due to the continuity of the sine flow, which produces curved lamellar structures that are not exactly one-dimensional.

The stretching variability in bundles is directly linked to the heterogeneity of the chaotic flow, because of the intimate relationship existing between the fractal geometry of the material line and the stretching statistics of fluid elements (Ott & Antonsen 1989). As such, it is impossible to have a single stretching rate per bundle as soon as the chaotic flow is heterogeneous and exhibits a distribution of stretching rates. The absence

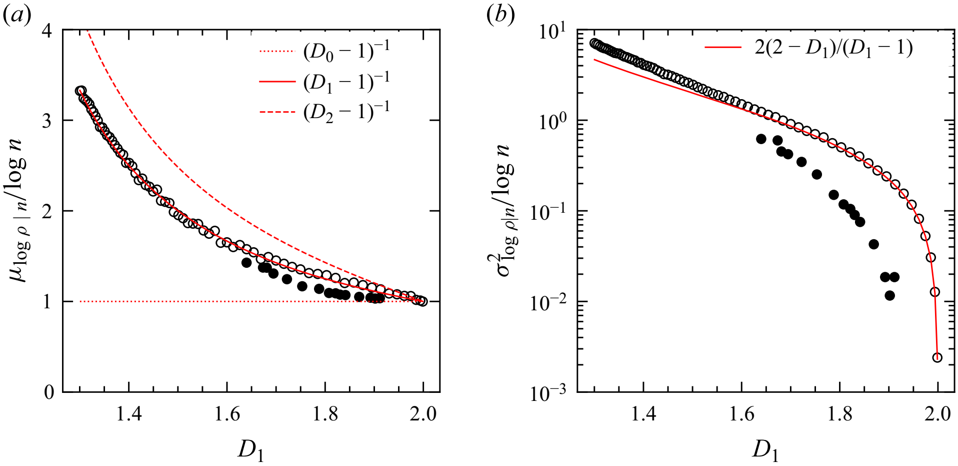

Figure 13. Scaling of the (*a*) mean and (*b*) variance of log-elongation in bundles as a function of the information dimension D_1. Circles stands for numerical simulations in baker maps (open circles) and sine flow (filled circles). Solid lines stand for theoretical prediction of the mean (4.21) and variance (4.22). Dashed and dotted lines are plotted to compare the mean with fractal dimensions of other orders.

of stretching variability in bundles ($\sigma^2_{\log\rho\,|\,n}=0$) implies the uniformity of stretching at large scale ($\sigma^2_{\rho}=0$). This uniform case is reached when $D_1\to D_0=2$, for instance, in the baker map when $a\to 0.5$. In continuous flow maps such as the sine flow, regions of high and low stretching always coexist, and $\sigma^2_{\log\rho\,|\,n}>0$.

4.4. *Moments of* $1/\rho$ *in a bundle of size n*

Having described the first two moments of the distribution of lamella elongation in bundles (4.21)–(4.22), we now assume that the distribution is of log-normal shape. This choice is justified by the fact that elongation is a multiplicative process, thus usually leading to log-normal distributions (Le Borgne, Dentz & Villermaux 2015; Souzy *et al.* 2020). This allows us to compute the scaling of the q moments of lamella concentrations in bundles, $\theta\,|\,n$. Owing to (2.14), we have

$$\begin{aligned}\mu_{\theta^q\,|\,n} &\sim \mu_{\rho^{-q}\,|\,n}\\ &= \int_1^\infty \rho^{-q} P_{\rho\,|\,n}(\rho)\,\mathrm{d}\rho\\ &\approx \int_1^\infty \exp(-(\log\rho-\mu_{\log\rho\,|\,n})^2/(2\sigma^2_{\log\rho\,|\,n})-q\log\rho)\,\mathrm{d}\rho. \end{aligned}\tag{4.23}$$

The minimum bound for the integral is taken at $\rho=1$ and not 0, taking into account the fact that lamellar structures cannot be compressed in their longitudinal direction. As a consequence, $P_{\rho\,|\,n}(\rho)$ is truncated for $\rho<1$. Denoting $\varLambda=\log\rho/\log n$, $\tilde{\mu}=\mu_{\log\rho\,|\,n}/\log n$ and $\tilde{\sigma}^2=\sigma^2_{\log\rho\,|\,n}/\log n$, this expression becomes

$$\mu_{\rho^{-q}\,|\,n}\approx\int_1^\infty \exp(H(\varLambda)\log n)\,\mathrm{d}\rho,\tag{4.24}$$

with $H(\varLambda)=-(\varLambda-\tilde{\mu})^2/(2\tilde{\sigma}^2)-q\varLambda$. For large n, the value of this integral tends to $\exp(H(\varLambda^*)\log n)$, where $\varLambda^*$ is the value where H takes a maximum, that is, either at

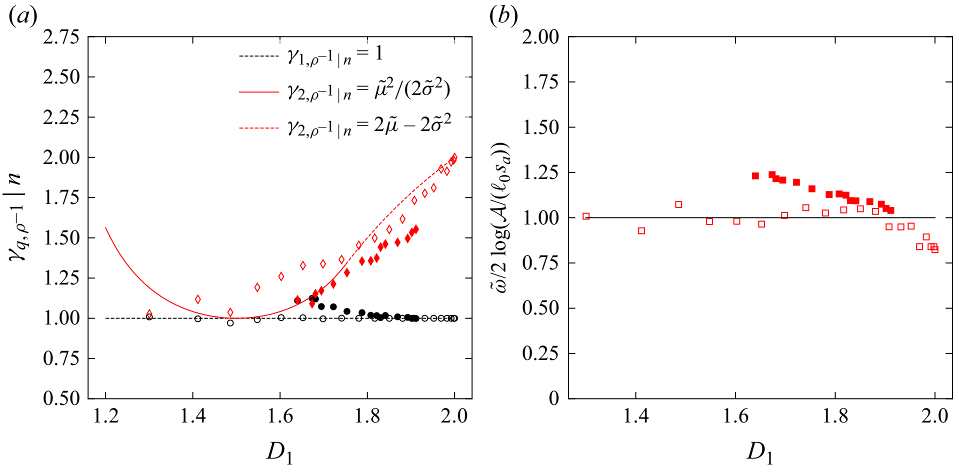

Figure 14. (*a*) Scaling exponents of the q lamellar concentration moments in bundles (4.23). Numerical estimates are plotted with symbols: red diamonds for $q = 2$, and black circles for $q = 1$. Unfilled and filled symbols represent simulations in the baker map and sine flow, respectively. Theoretical predictions (4.27) are represented by lines. (*b*) Intercept $\tilde{\omega}$ of the scaling exponent of the q lamellar concentration moments in bundles (4.23) in the baker map (empty squares) and sine flow (filled squares), and theoretical prediction (line, (4.28)).

$\varLambda^* = \tilde{\mu} - q\tilde{\sigma}^2$ if $\tilde{\mu} - q\tilde{\sigma}^2 > 0$, or at $\varLambda^* = 0$ otherwise. Thus

$$\log \mu_{\rho^{-q}\,|\,n} \approx -(\gamma_{q,\rho^{-1}\,|\,n} \log n + \omega_{q,\rho^{-1}\,|\,n}), \tag{4.25}$$

with

$$\left.\begin{aligned} \gamma_{q,\rho^{-1}\,|\,n} &= q\tilde{\mu} - q^2\tilde{\sigma}^2/2 && \text{if } \tilde{\mu} > q\tilde{\sigma}^2, \\ \gamma_{q,\rho^{-1}\,|\,n} &= \tilde{\mu}^2/(2\tilde{\sigma}^2) && \text{if } \tilde{\mu} \leq q\tilde{\sigma}^2, \end{aligned}\right\} \tag{4.26}$$

and $\omega_{q,\rho^{-1}\,|\,n} = q \log(\mathcal{A}/(s_a \ell_0))$ a constant. In particular, we are interested in the exponent $q = 2$, which is useful to describe fluctuations around the mean. We have

$$\tilde{\gamma} \equiv \gamma_{2,\rho^{-1}\,|\,n} = \begin{cases} 2\tilde{\mu}_\lambda - 2\tilde{\sigma}_\lambda^2 & \text{if } \tilde{\mu} > 2\tilde{\sigma}^2, \\ \tilde{\mu}^2/(2\tilde{\sigma}^2) & \text{if } \tilde{\mu} \leq 2\tilde{\sigma}^2. \end{cases} \tag{4.27}$$

The predicted dependence of $\tilde{\gamma}$ upon D_1 is reproduced in figure 14. Here, $\tilde{\gamma}$ is bounded between 2 (for $D_1 \to 2$) and 1 for $D_1 \approx 1.5$. The prediction agrees reasonably well with numerical simulations of the baker map and sine flow (figure 14). The small discrepancies can be attributed to deviations from log-normally distributed elongation in bundles. We also verify numerically that $\tilde{\omega} \equiv \omega_{2,\rho^{-1}\,|\,n}$ is independent of D_1 (figure 14*b*). In figure 15, we verify that $\tilde{\gamma}$ is independent of the aggregation scale s_a. In contrast,

$$\tilde{\omega} \approx 2 \log(\mathcal{A}/(l_0 s_a)) \tag{4.28}$$

is a function of the aggregation scale only, independent of time and fractal dimension (figure 14*b*). Having determined both the elongation statistics inside aggregates of size n and the spatial distribution of n, we will deduce in the following section the statistics of aggregated scalar levels c.

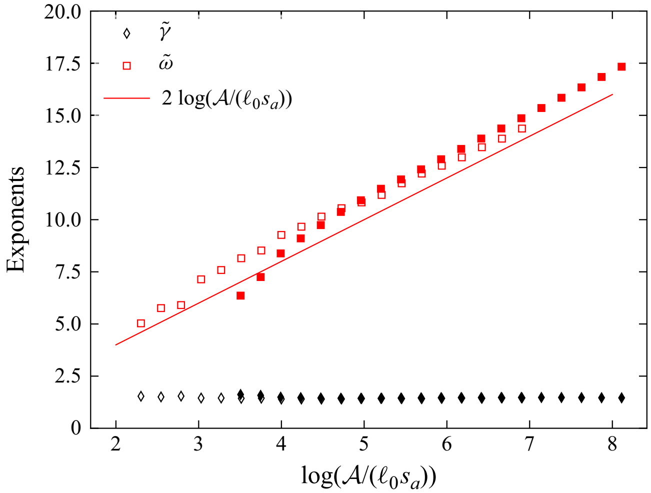

Figure 15. Dependence of $\tilde{\gamma}$ and $\tilde{\omega}$ with the aggregation scale in simulations of the baker map (empty symbols, $a = 0.2$) and the sine flow (filled symbols, $A = 1.2$), and comparison to theoretical prediction (4.28).

5. Aggregated scalar concentrations

In the preceding section, we have shown that the moments of lamellar compression inside a bundle of size n follow

$$\mu_{\rho^{-1}\,|\,n} = \frac{s_a \ell_0}{\mathcal{A}} n^{-1}, \tag{5.1}$$

$$\mu_{\rho^{-2}\,|\,n} = \frac{(s_a \ell_0)^2}{\mathcal{A}^2} n^{-\tilde{\gamma}}, \tag{5.2}$$

with $\tilde{\gamma} \equiv \gamma_{2,\rho^{-1}\,|\,n}$ a flow-dependent exponent varying with the fractal dimension D_1 and taking value between 1 and 2 (figure 14*a*). The variance of lamellar concentrations inside bundles thus follows:

$$\sigma^2_{\rho^{-1}\,|\,n} = \frac{(s_a \ell_0)^2}{\mathcal{A}^2} (n^{-\tilde{\gamma}} - n^{-2}). \tag{5.3}$$

We first assume that the aggregated scalar concentration of a bundle $c\,|\,n$ can be obtained by a sum of independent and identically distributed random variables (2.21), whose statistics have been described previously. In doing so, we make two hypotheses. First, we assume that the elongation statistics of lamellae inside a bundle of size n or among all bundles of size n are comparable, which is true for large n (see Appendix C). Second, we assume that variability in bundle aggregated concentrations arises from independent realizations of the sum. In other words, we assume that the chaotic flow is sufficiently random to shuffle lamellar elongations between each bundles. We will show later that independence is not ensured in the deterministic baker map.

With these hypotheses, the mean aggregated concentration is

$$\mu_{c\,|\,n} = n \times \frac{\sqrt{\pi}\,\theta_0 s_0}{s_a} \mu_{\rho^{-1}\,|\,n} = \frac{\sqrt{\pi}\,\theta_0 \ell_0 s_0}{\mathcal{A}}, \tag{5.4}$$

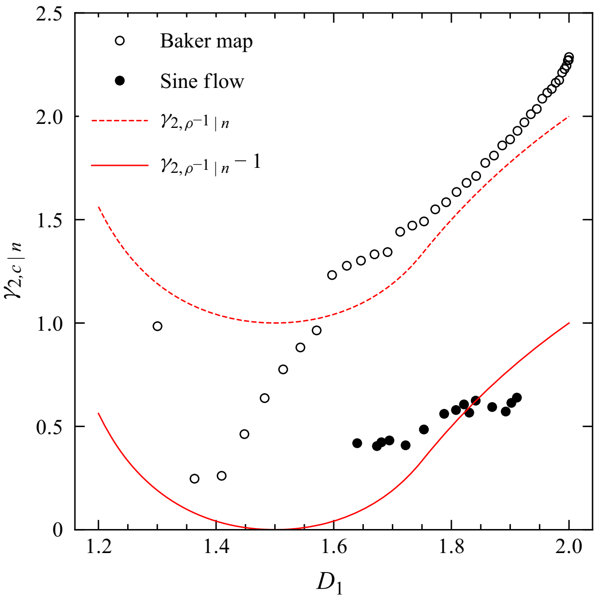

Figure 16. Scaling exponent ξ (5.8) of the variance of bundle concentrations knowing n estimated from simulations (dots) and theoretical predictions with the independent realization hypothesis for the sine flow (solid lines, $\xi = \tilde{\gamma} - 1$, (5.6)) and baker map (dashed lines, $\xi = \tilde{\gamma}$, (5.7)).

which is also the mean concentration μ_c. The variance of aggregated concentrations over bundles of similar size is

$$\sigma^2_{c\,|\,n} \approx n\left(\frac{\sqrt{\pi}\,\theta_0 s_0}{s_a}\right)^2 \sigma^2_{\rho^{-1}\,|\,n} = \frac{(\sqrt{\pi}\,\theta_0 \ell_0 s_0)^2}{\mathcal{A}^2}\,(n^{1-\tilde{\gamma}} - n^{-1}). \tag{5.5}$$

When n is large, this expression further simplifies to

$$\sigma^2_{c\,|\,n} \sim n^{1-\tilde{\gamma}}, \tag{5.6}$$

with $\tilde{\gamma} \in [1, 2]$ given by (4.27).

The scaling obtained with the independent realization hypothesis compares well with numerical simulations of aggregation in random sine flows of various heterogeneity (figure 16). However, it largely underestimates the scaling observed in the deterministic baker map. Indeed, the simplicity and regularity of the deterministic baker map make bundles of similar size not statistically independent. While bundle concentration still results from the addition of variable lamellar concentrations, independent realizations of the sum are not achieved due to the deterministic nature of the map, the same combinations being repeated identically in most bundles, as in the case of a unique realization. In that case, the variance of the sum is the variance of the random variable, and (5.5) transforms into

$$\sigma^2_{c\,|\,n} \sim \sigma^2_{\rho^{-1},n} \sim n^{-\tilde{\gamma}}. \tag{5.7}$$

This scaling indeed fits more accurately the deterministic baker map simulations (figure 16).

To summarize, the addition of lamellar concentration levels in a bundle yields a concentration whose deviation from the mean decays algebraically with the number of

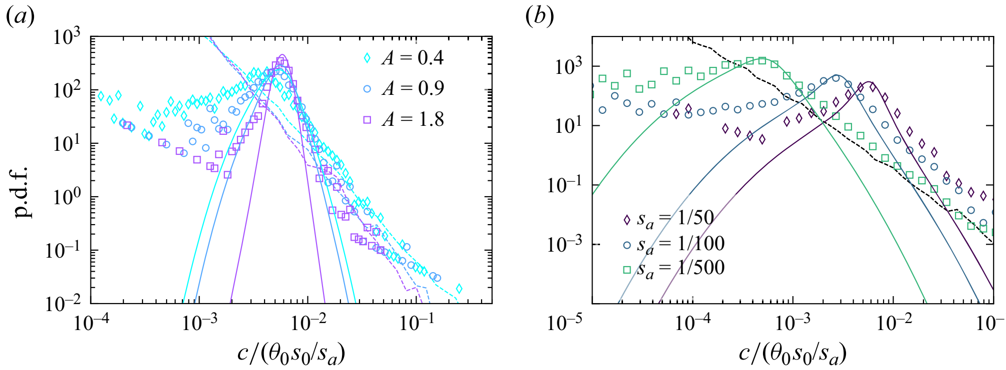

Figure 17. Distributions of aggregated scalar concentrations in the sine flow depending on (*a*) the sine wave amplitude A ($s_a = 1/50$), and (*b*) the aggregation scale s_a ($A = 0.9$). Symbols stand for numerical simulations, solid lines are the aggregation model (5.9), and dashed lines are the isolated strip prediction (2.6). Simulations are all taken at the time when the total filament length reaches $L = 10^7 \ell_0$.

lamellae in the bundle:

$$\sigma^2_{c\,|\,n} \approx \frac{(\theta_0 \ell_0 s_0)^2}{\mathcal{A}^2}\, n^{-\xi}, \tag{5.8}$$

with $\xi = \tilde{\gamma}$ for purely deterministic flows (baker map), and $\xi = \tilde{\gamma} - 1$ for random flows (sine flow). We call ξ the correlation exponent, which can take values between 0 and 2 depending on the flow heterogeneity and randomness.

5.1. *Distribution of c*

We now focus on deriving the expression of the p.d.f. of aggregated scalar concentration P_c over the flow domain. The later can be obtained from the conditional p.d.f. of bundle aggregated concentrations knowing their size $P_{c\,|\,n}$ and the p.d.f. of bundle size P_n by the weighted summation

$$P_c(c) = \int_n \mathrm{d}n\, P_{c\,|\,n}(c)\, P_n(n). \tag{5.9}$$

Here, P_n was found to be well approximated by a Gamma distribution, with moments given by the fractal characteristics of advected material lines (4.14). In turn, we have determined the scaling of the first two moments of bundle aggregated concentration $c\,|\,n$ (5.8), without specifying the precise shape of the p.d.f. A natural choice for $P_{c\,|\,n}$ is the log-normal distribution, since the sum of log-normally distributed random variables is known (Schwartz & Yeh 1982) to be well approximated by log-normal distributions. Adopting the log-normal shape, the parameters of the distribution are

$$\mu_{\log c\,|\,n} = \log \mu_{c\,|\,n} - \log(\sigma^2_{c\,|\,n}/\mu^2_{c\,|\,n} + 1)/2, \tag{5.10}$$

$$\sigma^2_{\log c\,|\,n} = \log(\sigma^2_{c\,|\,n}/\mu^2_{c\,|\,n} + 1). \tag{5.11}$$

In figure 17, we plot the simulated distribution of aggregated concentration levels compared to the prediction (5.9) for the baker map and sine flow. The agreement is fair in the region near μ_c, but deviates for large c. Indeed, this corresponds to lamellae with weak aggregation for which $n \approx 1$. In this region, the solitary strip p.d.f. $P_{\rho^{-1}}$ describes well the

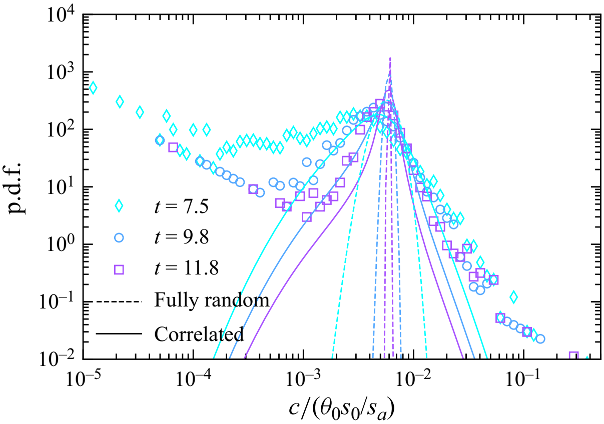

Figure 18. Plots of P_c in sine flows at several times ($A = 0.8$, $s_a = 1/50$) (symbols) compared with the random aggregation (dashed lines) and correlated aggregation (solid lines, (5.9)) models.

tail of P_c because such high concentration excursions are essentially supported by isolated lamellae, while the correlated aggregation model assumes $n \gg 1$. The presence of these weakly aggregated, high concentration levels is particularly evident at high Péclet number (figure 17*b*). The scalar concentration p.d.f. is thus the combination of an aggregated core around the mean following (5.9) and tails following the isolated strip concentration p.d.f.

In figure 18, we compare the correlated aggregation model with the random aggregation model where $\mu_n \sim L(t)$ (2.19). The random aggregation assumption yields gamma p.d.f.s that are narrowing much faster than the simulated p.d.f.s in the sine flow. In contrast, the correlated model better captures the p.d.f.

From the p.d.f. of aggregated scalar concentration, we now derive its moments. They are directly related to the p.d.f. of n, since

$$\mu_c = \int \mathrm{d}c \sum_n c\, P(c\,|\,n)\, P(n) = \sum_n \mu_{c\,|\,n}\, P(n) = \frac{\theta_0 \ell_0 s_0}{\mathcal{A}} \tag{5.12}$$

and

$$\mu_{c^2} = \int_c \mathrm{d}c \int_n \mathrm{d}n\, c^2\, P(c\,|\,n)\, P(n) = \int_n \mathrm{d}n\, \mu_{c^2\,|\,n}\, P(n) = \frac{(\theta_0 \ell_0 s_0)^2}{\mathcal{A}^2}\, (\mu_{n^{-\xi}} + 1). \tag{5.13}$$

Thus the scalar variance is

$$\sigma_c^2 = \frac{(\theta_0 \ell_0 s_0)^2}{\mathcal{A}^2}\, \mu_{n^{-\xi}}. \tag{5.14}$$

Note that because we chose a gamma distribution for n with parameters θ_n and k_n defined in (4.14), $\mu_{n^{-\xi}}$ is not defined for all D_2. Indeed, the gamma distribution admits a power law tail at small n with exponent $k_n - 1$, which affects the value of negative moments: when $k_n(D_2) < \xi(D_2)$, $\mu_n^{-\xi}$ does not exist. However, because fully chaotic flows are space-filling ($D_0 = 2$), we must have $n \geq 1$ at late time. Thus we restrict the p.d.f. of n to $n \geq 1$ and get

$$\mu_{n^{-\xi}} \sim (\theta_n)^{-\min(k_n,\xi)}. \tag{5.15}$$

An intuitive understanding of this equation can be formulated as follows. If the spatial heterogeneity of n is moderate ($\xi < k_n$), then the average of $n^{-\xi}$ is affected by all values

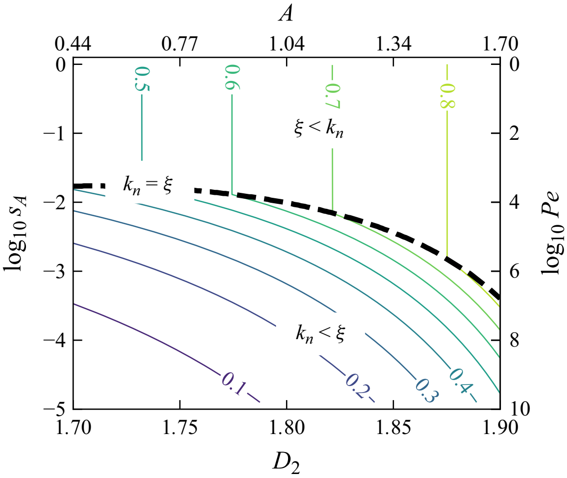

Figure 19. Evolution of the exponent $\min(k_n, \xi)$ (contours) in the sine flow for various Péclet numbers (aggregation scale s_A) and sine amplitude A (fractal dimension D_2). Here, k_n is obtained with (4.13) and $\xi = \tilde{\gamma} - 1$ with (4.27).

of n in the distribution. In contrast, if the heterogeneity is stronger ($\xi > k_n$), then the probability of having low aggregation regions ($n \approx 1$) is high and controls the value of $\mu_{n^{-\xi}}$. In that case, the average no longer scales with ξ, but rather scales with the parameter k_n, explaining the minimum in the exponent. The value of the exponent depending on Péclet number and fractal dimension is plotted in figure 19 for the sine flow. In such random flow, the exponent takes values between 0 and 1. For high Péclet numbers and low fractal dimension, $\mu_{n^{-\xi}}$ is governed by weakly aggregated regions and $\min(k_n, \xi) = k_n$. In contrast, for low Péclet number and large fractal dimension, the whole distribution of n plays a role in the determination of $\mu_{n^{-\xi}}$, and $\min(k_n, \xi) = \xi$.

Combining (5.15) and (4.14) provides the asymptotic scalar variance decay as a function of the domain area $\mathcal{A}$, the elongation of material lines $L(t)$, and the aggregation scale $s_a = \sqrt{2\pi}\, s_B$:

$$\sigma_c^2(t) = \left(\frac{L(t)\, s_a (\sqrt{\mathcal{A}}\, s_a^{D_2 - 2} - 1)}{\mathcal{A}} \right)^{-\min(k_n, \xi)} \sim L(t)^{-\min(k_n, \xi)}. \tag{5.16}$$

Thus in a correlated aggregation scenario, the decay exponent of scalar variance is found to be a fraction of the growth exponent of material lines, respectively $\log 2$ and $\mu_\lambda + \sigma_\lambda^2/2$ in the baker map and sine flow.

Equation (5.16) suggests that the existence of correlations in the aggregation process may be envisioned as a retardation of the purely random aggregation scenario (2.22). From this perspective, the effective number of random aggregation events $L(t)^{\min(k_n, \xi)}$ is smaller than the total number of aggregation $L(t)$, because the correlation exponent $\min(k_n, \xi)$ is generally smaller than 1 (see figure 19).

In contrast, the random aggregation model (2.22) clearly overestimates the variance decay rate in the sine flow. This is explained by the correlated nature of aggregation, which is less efficient at homogenizing concentration levels than a completely random addition. In other words, small concentration levels have a higher probability of coalescing with other small concentrations than with high concentrations, retarding the homogenization of

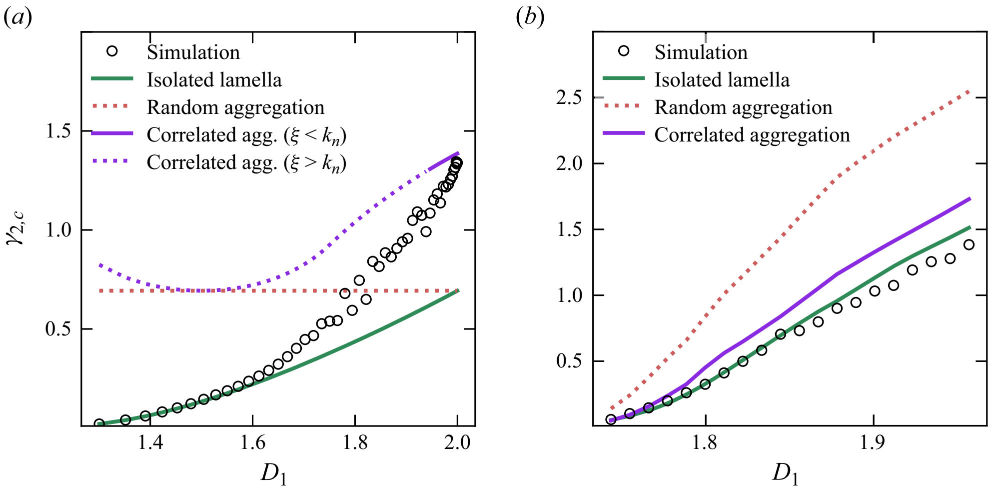

Figure 20. Decay exponent γ_2 of the variance of aggregated scalar levels with time, as a function of fractal dimension D_1 for (*a*) the baker map and (*b*) the random sine flow. Dots stand for numerical simulations, and lines from theoretical predictions for isolated lamellae ((2.17) and exponents of $\mu_{\rho^{-1}}$ in tables 1 and 2), random aggregation ((2.22) and exponents of $1/\mu_\rho$ in tables 1 and 2) and correlated aggregation (5.14).

the mixture. Indeed, the correlated aggregation model proposed herein (5.16) shows decay rates closer to the numerical observations. The solitary lamella model still outperforms the correlated aggregation framework in predicting scalar decay rate, because the latter is dominated by weakly aggregated regions of the flow. These regions are poorly described by statistics obtained in the large-n limit (e.g. (5.5)).

A different picture is observed for scalar decay rates obtained in the baker map (figure 20*b*), because of the deterministic nature of the map. The solitary lamella model is accurate for fractal dimensions $D_1 < 1.6$, where stretching heterogeneity is important and weakly aggregated regions of the flow dominate scalar fluctuations. For $D_1 > 1.6$, the isolated lamella model underpredicts scalar decay due to aggregation effects, as explained in the following. The random aggregation scenario predicts an invariant scalar decay rate ($\gamma_{2,c} = \log 2$) equal to the growth rate of material lines. It thus overpredicts the observed decay rates for $D_1 < 1.8$, and underpredicts them at larger values of D_1. Interestingly, when the flow tends to the uniform case $a \rightarrow 0.5$ ($D_1 \rightarrow 2$), simulations yield scalar decay rate $2 \log 2$, larger than the decay rate of isolated lamellae. This acceleration of mixing by aggregation is a pure consequence of the determinism nature of the map, and is well captured by the correlated aggregation scenario (5.7). However, for all baker maps with $D_1 < 1.9$ ($a < 0.2$), $k_n < \xi$ and scalar fluctuations are mainly governed by the regions where $n \sim 1$ for which the correlated aggregation theory is not expected to be accurate. This explains the limited range of validity of the correlated aggregation theory for describing scalar decay rates in the deterministic baker map.

Finally, note that the simulations do not show the super-exponential decay of scalar fluctuations classically observed for the uniform stretching rate at $a = 0.5$. In fact, the reconstruction of the scalar field by a summation of lamellar maximum concentrations on a fixed grid (3.6) impedes the apparition of the super-exponential mode. As $a \rightarrow 0.5$, all lamellae are subjected to similar stretching rates around $\log 2$, thus yielding a scalar variance decaying as $2 \log 2$.

6. Conclusions

Scalar mixing by incompressible smooth two-dimensional chaotic flows leads to the emergence of two coupled phenomena that are the elongation of interfaces by stretching, and their aggregation and coalescence at the Batchelor scale by compression. In turbulent scalar mixing, aggregation was well captured by random addition rules (Duplat & Villermaux 2008). Here, we show that for smooth chaotic flows in the Batchelor range, aggregation is highly correlated. This correlated aggregation process significantly reduces the flow mixing efficiency compared to a random hypothesis, maintaining it close to the mixing efficiency obtained for solitary lamellae, and explaining the observed monotonic exponential decay of scalar variance before and after coalescence time (Fereday *et al.* 2002).

Using two-dimensional chaotic flows as a reference, we measured the aggregation rate of exponentially stretched material lines across a broad range of chaotic flow regimes. We showed that the most elongated lamellae are also the most aggregated ones, due to the fact that larger compression rates attract larger flow regions. The link between elongation and compression, induced by incompressibility, hence generates a direct correlation between elongation and aggregation. The heterogeneity in stretching rates therefore controls the heterogeneity of the number of lamellae in bundles.

We showed that the statistics of aggregated lamella numbers can be predicted exactly from the fractal dimensions of the elongated material line. We then derived a general theoretical framework that captures the effect of correlated aggregation, where lamellae of similar stretching aggregate preferentially, and predict the p.d.f.s of aggregated scalar levels. We find that correlated aggregation can be uniquely characterized by single correlation exponent $\xi \in [0, 1]$, which provides a measure of the effective number of aggregation events, compared to the total number. In that sense, correlated aggregation delays the route to uniformity compared to a fully random hypothesis, although it does not alter the fundamental nature of the aggregation process (Villermaux & Duplat 2003).

Our results apply for two-dimensional fully chaotic flows in the Batchelor regime, that is, for smooth velocity fields below the integral scale. These flow fields are representative of a large class of flows, including notably porous media flows (Heyman *et al.* 2020; Souzy *et al.* 2020), geophysical flows, and turbulent flows at high Schmidt number. We have also considered small periodic flow domains where no large-scale scalar gradients can appear. When scalar length scales can develop beyond the integral velocity scale in smooth flows, mixing is controlled globally by the slowest dispersing modes (Haynes & Vanneste 2005; Tsang *et al.* 2005). The aggregation rules determined herein are not supposed to change above the velocity scale. Thus it is possible in principle to obtain the statistics of larger-scale scalar fields by the summation of individual lamellae. To do so, one may consider the mean density of aggregation $\mu_n(t)$ (4.4) to vary spatially through macro-dispersion, while microscale fluctuations would remain governed by (4.11). Note that similar ideas were used to predict the evolution of a dispersing scalar plume in two-dimensional porous media (Le Borgne *et al.* 2015).

It should also be possible to extend the correlated aggregation theory to three-dimensional flows that produce two-dimensional sheets (Ngan & Vanneste 2011; Martínez-Ruiz *et al.* 2018; Meunier & Villermaux 2022) that aggregate at late times due to folding. These flows have a unique negative Lyapunov exponent, the mean compression rate normal to the sheets, and thus also simplify to the one-dimensional Lagrangian stretching framework on which aggregation models are framed.

Acknowledgements. J.H., T.L.B. and P.D. acknowledge the European Research Council under the grant 101042466 (CHORUS). T.L.B. and E.V. acknowledge the H2020 MSCA ITN programme under the grant 956457 (COPERMIX). Three anonymous reviewers are thanked for contributions to the manuscript.

Funding. Funded by the European Union (ERC, CHORUS, 101042466).

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

Data availability statement. The data and codes to reproduce the findings of this study are openly available in Github at https://github.com/jorishey1234/aggregation or upon reasonable request to joris.heyman@univ-rennes.fr.

Author ORCIDs.
J. Heyman https://orcid.org/0000-0002-0327-7924;
E. Villermaux https://orcid.org/0000-0001-5130-4862;
P. Davy https://orcid.org/0000-0002-6648-0145;
T. Le Borgne https://orcid.org/0000-0001-9266-9139.

Appendix A. Stretching statistics and averaging

In random chaotic flows, varying stretching rates are experienced by fluid elements (Lester *et al.* 2013). Because of the multiplicative nature of stretching, the log-elongation of material elements $\log\rho$ is well approximated in ergodic chaotic flows by a sum of independent and identically distributed random variables that converges towards the normal distribution with mean $\mu_\lambda t$ and variance $\sigma_\lambda^2 t$ (Meunier & Villermaux 2022).

Thus the elongation ρ of material elements may be expected to follow the log-normal distribution

$$P_{\rho,0}(\rho) \approx \frac{1}{\rho\sqrt{2\pi\sigma_\lambda^2 t}} \exp\left(-\frac{(\log\rho - \mu_\lambda t)^2}{2\sigma_\lambda^2 t}\right). \tag{A1}$$

Non-asymptotic stretching statistics can differ substantially from this limiting behaviour. For instance, the baker map has a binomial distribution of elongations

$$P_{\rho,0}(\rho = (1-a)^k a^{t-k}) = \binom{t}{k} (1-a)^k a^{t-k}, \quad \text{for } k = 0, \ldots, t, \tag{A2}$$

which tends to a log-normal distribution with $\mu_\lambda/t = -a\log(a) - (1-a)\log(1-a)$ and $\sigma_\lambda^2/t = a(1-a)(\log(1-a) - \log(a))^2$.

Note also that because of the multiplicative nature of elongation (Redner 1990), the p.d.f. of ρ (and ρ^{-1}) is highly sensitive to the tails of $P_{\log\rho}$, which may converge only slowly to the Gaussian prediction, depending on the flow heterogeneity. However, for the range of sine flow amplitudes used in this study ($A \in [0.4, 1.8]$), such limitations are relatively weak.

Statistics of lamellar concentrations in the material line can then be obtained by suitable ensemble averaging (denoted by angle brackets) over this distribution. Depending on how sampling is performed through the material line, different moments are obtained. Uniform sampling on the initial filament prior elongation leads to the distribution $P_\rho(\rho)$ in (A1), with main moments summarized in table 1. In contrast, uniform sampling on a the final elongated material line of length L, denoted by $\mu_{\bullet,L}$ leads to the weighted p.d.f. $P_{\rho,L} \sim \rho P_\rho$. Thus $P_{\rho,L}$ is also log-normal, but with different mean, $(\mu_\lambda + \sigma_\lambda^2)t$. Uniform sampling on the final material line gives a stronger weight to highly elongated part of the material

	$\log\rho$	ρ	ρ^{-1}
$\mu_\bullet$	$\mu_\lambda t$	$\exp((\mu_\lambda+\sigma_\lambda^2/2)t)$	$\begin{cases}\exp(-(\mu_\lambda-\sigma_\lambda^2/2)t) & \text{if } \mu_\lambda\geq\sigma_\lambda^2\\ \exp(-\mu_\lambda^2/(2\sigma_\lambda^2)t) & \text{if } \mu_\lambda\leq\sigma_\lambda^2\end{cases}$
$\mu_{\bullet^2}$	$\sigma_\lambda^2 t$	$\exp((2\mu_\lambda+2\sigma_\lambda^2)t)$	$\begin{cases}\exp(-2(\mu_\lambda-\sigma_\lambda^2)t) & \text{if } \mu_\lambda\geq 2\sigma_\lambda^2\\ \exp(-\mu_\lambda^2/(2\sigma_\lambda^2)t) & \text{if } \mu_\lambda\leq 2\sigma_\lambda^2\end{cases}$
$\mu_{\bullet,L}$	$(\mu_\lambda+\sigma_\lambda^2)t$	$\exp((\mu_\lambda+3\sigma_\lambda^2/2)t)$	$\exp((-\mu_\lambda-\sigma_\lambda^2/2)t)$
$\mu_{\bullet^2,L}$	$\sigma_\lambda^2 t$	$\exp((2\mu_\lambda+4\sigma_\lambda^2)t)$	$\begin{cases}\exp(-2\mu_\lambda t) & \text{if } \mu_\lambda\geq\sigma_\lambda^2\\ \exp(-(\mu_\lambda+\sigma_\lambda^2)^2/(2\sigma_\lambda^2)t) & \text{if } \mu_\lambda\leq\sigma_\lambda^2\end{cases}$

Table 1. Moments of log-normally distributed stretching sampled over infinitesimal fluid elements ($\mu_\bullet$), material line ($\mu_{\bullet,L}$).

	$\log\rho$	ρ	ρ^{-1}
$\mu_\bullet$	$(-a\log(a)-(1-a)\log(1-a))t$	2^t	$(1-2a+2a^2)^t$
$\mu_{\bullet^2}$	$a(1-a)(\log(1-a)-\log(a))^2 t$	—	$(1-3a+3a^2)^t$

Table 2. Moments of binomial distributed stretching sampled over infinitesimal fluid elements ($\mu_\bullet$), material line ($\mu_{\bullet,L}$).

line than uniform sampling on the initial filaments. Moments of $\log\rho$, ρ and ρ^{-1} are summarized in table 1 for sine flow and table 2 for the baker map for initial and final sampling. Note that we impose $\rho\geq 1$ since the one-dimensional lamellar framework is valid only when a lamella elongates in the y direction. This lower bound creates a particular scaling of moments of ρ^{-1} when σ_λ^2 is larger than μ_λ, that is, when weak stretching rates dominate the ensemble average.

Appendix B. Fractal dimensions in the baker map

The fractal dimension of order q of the measure p is obtained with (Grassberger 1983)

$$D_q-1=\lim_{\epsilon\to 0}\frac{1}{q-1}\frac{\log I_q(\epsilon)}{\log\epsilon},\quad I_q(\epsilon)\equiv\sum_k^{N=\mathcal{L}/\epsilon}p_k^q, \tag{B1a,b}$$

where the subtraction of 1 on the left-hand side accounts for the counting of one-dimensional structures (lamellae) in a two-dimensional domain. This definition implies the following spatial scaling of the integral of the measure:

$$I_q(\epsilon)\sim\epsilon^{(q-1)(D_q-1)}. \tag{B2}$$

In the baker map, the integral of the measure can then be computed by summing its value on the two replicates created by the map:

$$I_q(\epsilon)=I_{q,a}(\epsilon)+I_{q,1-a}(\epsilon). \tag{B3}$$

We observe that

$$I_{q,a}(a\epsilon) = I_{q,1-a}((1-a)\epsilon) = \sum_{k}^{N=1/\epsilon} \left(\frac{p_k}{2}\right)^q, \tag{B4}$$

where the factor $1/2$ comes from the normalization of the measure due to the doubling of n. Thus

$$I_{q,a}(\epsilon) = I_q(\epsilon/a)\, 2^{-q} \quad \text{and} \quad I_{q,1-a}(\epsilon) = I_q(\epsilon/(1-a))\, 2^{-q}. \tag{B5a,b}$$

Replacing the last expression in (B3) yields

$$I_q(\epsilon) = 2^{-q}\epsilon^{(q-1)D_q}(a^{-(q-1)(D_q-1)} + (1-a)^{-(q-1)(D_q-1)}). \tag{B6}$$

Using the scaling $I_q(\epsilon) \sim \epsilon^{(q-1)D_q}$ thus provides a transcendental equation for D_q independently of ϵ:

$$2^q = a^{-(q-1)(D_q-1)} + (1-a)^{-(q-1)(D_q-1)}, \tag{B7}$$

the solution of which is explicit for $q = 0$ and $q = 1$:

$$D_0 = 2, \quad D_1 = 1 + \frac{2\log 2}{\log(a^{-1} + (1-a)^{-1})}. \tag{B8a,b}$$

Note that the solution for $q = 1$ is obtained with Bernoulli's rule by differentiating (B7) with respect to q, and taking the limit $q \to 1$.

Appendix C. Single and multiple bundle statistics

The variability of a set of random numbers is always greater than the average variability of a subset of these numbers. Thus the stretching variance among bundles of similar sizes – denoted $\sigma^2_{\rho^{-1},n}$ – is always larger than the average stretching variance inside the bundle (5.3). We have

$$\sigma^2_{\rho^{-1},n} = \left(\frac{n}{n-1}\right)\sigma^2_{\rho^{-1}\,|\,n}. \tag{C1}$$

For instance, for bundles made of two lamellae, we expect $\sigma^2_{\rho^{-1},n}$ to be twice as large as $\sigma^2_{\rho^{-1}\,|\,n}$ given by (5.3). The difference between the statistics of the set and its subset tends to reduce at large n, where $\sigma^2_{\rho^{-1}\,|\,n} \to \sigma^2_{\rho^{-1},n}$. For $n = 1$, both $\sigma^2_{\rho^{-1}\,|\,n}$ and $n-1$ cancel out, so the previous equation is undetermined.

Assuming that bundles of similar size have independent stretching histories, the variability of the aggregated scalar concentration is obtained from independent realizations of the random sum (3.6). Thus the variance of $c\,|\,n$ reads

$$\begin{aligned} \sigma^2_{c\,|\,n} &= n\left(\frac{\sqrt{\pi}\,\theta_0 s_0}{s_a}\right)^2 \sigma^2_{\rho^{-1},n} = \left(\frac{\sqrt{\pi}\,\theta_0 s_0}{s_a}\right)^2 \sigma^2_{\rho^{-1}\,|\,n}\left(\frac{n^2}{n-1}\right) \\ &= \frac{(\sqrt{\pi}\,\theta_0 \ell_0 s_0)^2}{\mathcal{A}^2}\left(\frac{n^{2-\tilde{\gamma}}-1}{n-1}\right). \end{aligned} \tag{C2}$$

When n is large, however, we recover

$$\sigma^2_{c\,|\,n} \sim n^{-\tilde{\gamma}+1}. \tag{C3}$$

REFERENCES

AREF, H. 1984 Stirring by chaotic advection. *J. Fluid Mech.* **143**, 1–21.

BALKOVSKY, E. & FOUXON, A. 1999 Universal long-time properties of Lagrangian statistics in the Batchelor regime and their application to the passive scalar problem. *Phys. Rev.* E **60** (4), 4164.

BATCHELOR, G.K. 1959 Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1. General discussion and the case of small conductivity. *J. Fluid Mech.* **5** (1), 113–133.

DUPLAT, J., JOUARY, A. & VILLERMAUX, E. 2010 Entanglement rules for random mixtures. *Phys. Rev. Lett.* **105**, 034504.

DUPLAT, J. & VILLERMAUX, E. 2008 Mixing by random stirring in confined mixtures. *J. Fluid Mech.* **617**, 51–86.

FALKOVICH, G., GAWĘDZKI, K. & VERGASSOLA, M. 2001 Particles and fields in fluid turbulence. *Rev. Mod. Phys.* **73** (4), 913.

FARMER, J.D., OTT, E. & YORKE, J.A. 1983 The dimension of chaotic attractors. *Physica* D **7** (1–3), 153–180.

FEREDAY, D.R., HAYNES, P.H., WONHAS, A. & VASSILICOS, J.C. 2002 Scalar variance decay in chaotic advection and Batchelor-regime turbulence. *Phys. Rev.* E **65** (3), 035301.

FINN, J.M. & OTT, E. 1988 Chaotic flows and fast magnetic dynamos. *Phys. Fluids* **31** (10), 2992–3011.

GARRETT, C. 1983 On the initial streakness of a dispersing tracer in two- and three-dimensional turbulence. *Dyn. Atmos. Oceans* **7** (4), 265–277.

GIONA, M., CERBELLI, S. & ADROVER, A. 2001 Geometry of reaction interfaces in chaotic flows. *Phys. Rev. Lett.* **88** (2), 024501.

GRASSBERGER, P. 1983 Generalized dimensions of strange attractors. *Phys. Lett.* A **97** (6), 227–230.

HAYNES, P.H. & VANNESTE, J. 2005 What controls the decay of passive scalars in smooth flows? *Phys. Fluids* **17** (9), 097103.

HEYMAN, J., LESTER, D.R. & LE BORGNE, T. 2021 Scalar signatures of chaotic mixing in porous media. *Phys. Rev. Lett.* **126**, 034505.

HEYMAN, J., LESTER, D.R., TURUBAN, R., MÉHEUST, Y. & LE BORGNE, T. 2020 Stretching and folding sustain microscale chemical gradients in porous media. *Proc. Natl Acad. Sci. USA* **117** (24), 13359–13365.

LE BORGNE, T., DENTZ, M. & VILLERMAUX, E. 2015 The lamellar description of mixing in porous media. *J. Fluid Mech.* **770**, 458–498.

LE BORGNE, T., HUCK, P.D., DENTZ, M. & VILLERMAUX, E. 2017 Scalar gradients in stirred mixtures and the deconstruction of random fields. *J. Fluid Mech.* **812**, 578–610.

LESTER, D.R., DENTZ, M. & LE BORGNE, T. 2016 Chaotic mixing in three-dimensional porous media. *J. Fluid Mech.* **803**, 144–174.

LESTER, D.R., METCALFE, G. & TREFRY, M.G. 2013 Is chaotic advection inherent to porous media flow? *Phys. Rev. Lett.* **111**, 174101.

MARTÍNEZ-RUIZ, D., MEUNIER, P., FAVIER, B., DUCHEMIN, L. & VILLERMAUX, E. 2018 The diffusive sheet method for scalar mixing. *J. Fluid Mech.* **837**, 230–257.

MEUNIER, P. & VILLERMAUX, E. 2010 The diffusive strip method for scalar mixing in two dimensions. *J. Fluid Mech.* **662**, 134–172.

MEUNIER, P. & VILLERMAUX, E. 2022 The diffuselet concept for scalar mixing. *J. Fluid Mech.* **951**, A33.

NGAN, K. & VANNESTE, J. 2011 Scalar decay in a three-dimensional chaotic flow. *Phys. Rev.* E **83** (5), 056306.

OTT, E. & ANTONSEN, T.M. JR 1989 Fractal measures of passively convected vector fields and scalar gradients in chaotic fluid flows. *Phys. Rev.* A **39** (7), 3660.

OTTINO, J.M. 1990 Mixing, chaotic advection, and turbulence. *Annu. Rev. Fluid Mech.* **22**, 207–253.

PIERREHUMBERT, R.T. 1994 Tracer microstructure in the large-eddy dominated regime. *Chaos, Solitons Fractals* **4** (6), 1091–1110.

RANZ, W.E. 1979 Applications of a stretch model to mixing, diffusion, and reaction in laminar and turbulent flows. *AIChE J.* **25** (1), 41–47.

REDNER, S. 1990 Random multiplicative processes: an elementary tutorial. *Am. J. Phys.* **58** (3), 267–273.

ROTHSTEIN, D., HENRY, E. & GOLLUB, J.P. 1999 Persistent patterns in transient chaotic fluid mixing. *Nature* **401** (6755), 770–772.

SCHWARTZ, S.C. & YEH, Y.-S. 1982 On the distribution function and moments of power sums with log-normal components. *Bell Syst. Tech. J.* **61** (7), 1441–1462.

SINAI, Y.G. & YAKHOT, V. 1989 Limiting probability distributions of a passive scalar in a random velocity field. *Phys. Rev. Lett.* **63** (18), 1962–1964.

SOUZY, M., LHUISSIER, H., MÉHEUST, Y., LE BORGNE, T. & METZGER, B. 2020 Velocity distributions, dispersion and stretching in three-dimensional porous media. *J. Fluid Mech.* **891**, A16.

SUKHATME, J. 2004 Probability density functions of decaying passive scalars in periodic domains: an application of Sinai–Yakhot theory. *Phys. Rev.* E **69** (5), 056302.

SUKHATME, J. & PIERREHUMBERT, R.T. 2002 Decay of passive scalars under the action of single scale smooth velocity fields in bounded two-dimensional domains: from non-self-similar probability distribution functions to self-similar eigenmodes. *Phys. Rev.* E **66** (5), 056302.

TANG, X.Z. & BOOZER, A.H. 1996 Finite time Lyapunov exponent and advection–diffusion equation. *Physica* D **95** (3–4), 283–305.

THIFFEAULT, J.-L. 2004 Stretching and curvature of material lines in chaotic flows. *Physica* D **198** (3), 169–181.

TSANG, Y.-K., ANTONSEN, T.M. JR & OTT, E. 2005 Exponential decay of chaotically advected passive scalars in the zero diffusivity limit. *Phys. Rev.* E **71** (6), 066301.

VILLERMAUX, E. 2012 On dissipation in stirred mixtures. *Adv. Appl. Mech.* **45**, 91–107.

VILLERMAUX, E. 2019 Mixing versus stirring. *Annu. Rev. Fluid Mech.* **51** (1), 245–273.

VILLERMAUX, E. & DUPLAT, J. 2003 Mixing as an aggregation process. *Phys. Rev. Lett.* **91** (18), 184501.

WARHAFT, Z. 2000 Passive scalars in turbulent flows. *Annu. Rev. Fluid Mech.* **32** (1), 203–240.

WONHAS, A. & VASSILICOS, J.C. 2002 Mixing in fully chaotic flows. *Phys. Rev.* E **66** (5), 051205.