UNIVERSITY OF LEEDS

This is a repository copy of *Flow complexity in open systems: interlacing complexity index based on mutual information*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/135022/

Version: Accepted Version

Article:

Pozo, JM, Geers, AJ, Villa-Uriol, M-C et al. (1 more author) (2017) Flow complexity in open systems: interlacing complexity index based on mutual information. Journal of Fluid Mechanics, 825. pp. 704-742. ISSN 0022-1120

https://doi.org/10.1017/jfm.2017.392

(c) 2017, Cambridge University Press. This article has been published in a revised form in the Journal of Fluid Mechanics (https://doi.org/10.1017/jfm.2017.392). This version is free to view and download for private research and study only. Not for re-distribution, re-sale or use in derivative works. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Interlacing Complexity Index for open flow systems based on mutual information.

Jose M. Pozo[1]†, **Arjan J. Geers**[2], **Maria-Cruz Villa-Uriol**[1]
and Alejandro F. Frangi[1]

[1]CISTIB Center for Computational Imaging & Simulation Technologies in Biomedicine,
Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield,
S1 3JD, UK

[2]CISTIB Center for Computational Imaging & Simulation Technologies in Biomedicine,
Department of Information and Communication Technologies, Universitat Pompeu Fabra,
Carrer Tànger, 122-140, 08018 Barcelona, Spain

(Received ?; revised ?; accepted ?. - To be entered by editorial office)

Flow complexity is related to a number of phenomena in science and engineering, which has been approached from the perspective of chaotic dynamical systems, ergodic processes, or mixing of fluids, just to name a few. To the best of our knowledge, all existing methods to quantify flow complexity are only valid for infinite time evolutions, for closed systems or for mixing of two substances. We introduce an index of flow complexity coined *interlacing complexity index* (ICI), valid for a single phase flow in an open system with inlet and outlet regions, involving finite times. ICI is based on Shannon's mutual information (MI), and inspired by an analogy between inlet-outlet open flow systems and communication systems in communication theory. The roles of transmitter, receiver, and communication channel are played, respectively, by the inlet, the outlet, and the flow transport between them. A perfectly laminar flow in a straight tube can be compared to an ideal communication channel where the transmitted and received messages are identical and hence the MI between input and output is maximal. For more complex flows, generated by more intricate conditions or geometries, the ability to discriminate the outlet position by knowing the inlet position is decreased, reducing the corresponding MI. The behaviour of the ICI has been tested with numerical experiments on diverse flows cases. The results indicate that the ICI provides a sensitive complexity measure with intuitive interpretation in a diversity of conditions and in agreement with other observations, such as Dean vortices and subjective visual assessments. As a crucial component of the ICI formulation, we also introduce the *natural distribution of streamlines* and the *natural distribution of world-lines*, with invariance properties with respect to the cross section used to parameterize them, valid for any type of mass-preserving flow.

Key words: Authors should not enter keywords on the manuscript, as these must be chosen by the author during the online submission process and will then be added during the typesetting process (see http://journals.cambridge.org/data/relatedlink/jfm-keywords.pdf for the full list)

---

† Email address for correspondence: j.pozo@sheffield.ac.uk

1. Introduction

The complexity of a flow is often related to the concept of chaos (Ottino 1989, 1990; Wiggins & Ottino 2004; Loskutov 2010; Aref *et al.* 2014). This has been studied from the perspective of dynamical systems and ergodic theory (Mathew & Mezić 2011; Scott *et al.* 2009), introducing measures such as *Lyapunov exponents* and *Kolmogorov–Sinai entropy*, characterizing orbits, and concepts such as *Poincaré sections*. A common feature of these measures is that they characterize flows infinitely propagated in time, in some sense representing the flow properties for large time scales. However, motivated by mixing problems, finite-time Lyapunov exponents have been also defined (Tang & Boozer 1996).

Poincaré sections have been also used (Ling 1994; Funakoshi 2008; Xia *et al.* 2010; Jang & Funakoshi 2010; Lucas & Kerswell 2014) to quantify chaotic mixing on cyclic systems, such as a periodic mixer or a helix-like pipe, after a number of loops. Poincaré sections can be used because the flow is considered cyclic, identifying the cross sections after each cycle with the initial one.

Most methods to quantify fluid flow mixing have focused on measuring the homogenization of a scalar, representing a phase in a two-phase fluid or a physical property like temperature (Lin *et al.* 2011; Thiffeault 2012). Since the seminal work of Danckwerts (1952) introducing the metrics *scale* and *intensity of segregation* (or *variance of concentration*), more complex metrics have been proposed, such as the *Sobolev norm* and the *mix-norm* (Mathew *et al.* 2005), or the *Shannon entropy* (Camesasca *et al.* 2006) and *Rényi entropy* (Camesasca & Manas-Zloczower 2009) of the scalar spatial distribution. In accordance with their purpose, all these metrics are dependent on the initial distribution of the studied scalar.

Brandani *et al.* (2013) extended the idea of Camesasca *et al.* (2006) by using conditional entropy, but still applicable only to two-phase fluid flow mixing. It was applied to the design and evaluation of mixer devices (Domingues *et al.* 2008). In contrast, Guida *et al.* (2010) used conditional entropy to introduce a mixing measure of the flow itself, independent of any scalar distribution. They independently developed the same idea previously presented by Alemaskin *et al.* (2004).

In most approaches, the fluid domain is discretized in a number of *cells*. This set of cells plays two different roles: defining the observation scale and acting as a convenient structure for computations. This double role can introduce some confusion. An explicit description of the observation scale and its relevance is found in Tucker & Peters (2003), which introduced two mixing measures as functions of scale: *sample standard deviation* and *sample max error*, for two-phase fluids. Particles inside each cell are distributed with different proportions over several cells after some time. The *distribution matrix* (Kruijt *et al.* 2001) encodes this information and has been used as an ingredient in many works. In contrast, the multi-scale aggregate *mix-norm* proposed by Mathew *et al.* (2005) assignes to each point centred balls of different radius. Each point contribution is then continuously integrated (no discretization), so that both roles are clearly separated.

A scalar-independent measure of chaotic mixing can be interpreted as a quantification of the complexity of the flow. Following this notion, we introduce in this paper the *Interlacing Complexity Index* (ICI). It is independent of any scalar distribution, and valid for open systems with inlet and outlet flow regions. None of the existing flow complexity measures can be applied for these systems. The large-time-scale measures from dynamical systems theory are not applicable since the flow transport from the inlet to the outlet involves a finite time. The ICI quantifies the loss of spatial coherence owing to this flow transport. The inlet and outlet are surfaces on the control volume under study. Similarly to Poincaré sections, the observation locus is given by the crossing of the streamlines

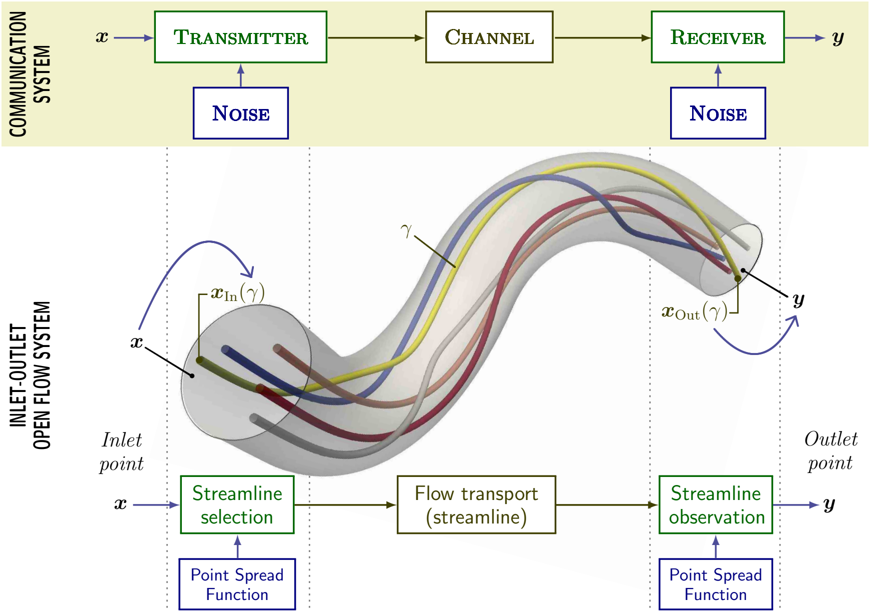

FIGURE 1. Inlet-outlet open flow system analogy to a communication system. Each inlet point, $\boldsymbol{x} \in \mathcal{S}_{\mathrm{In}}$ (*input alphabet*), is *transmitted* as a streamline, $\gamma \in \Gamma_{\mathcal{S}}$, propagated by the flow (*ideal channel*) up to the outlet. The streamline is *received* as a position $\boldsymbol{y} \in \mathcal{S}_{\mathrm{Out}}$ (*output alphabet*) selected around the streamline crossing $\boldsymbol{x}_{\mathrm{Out}}(\gamma)$. The *noise* is modelled as a Gaussian point spread function, representing imprecision in the identification of points, with a standard deviation depending on the observation scale.

through these surfaces. However, they cannot be treated as Poincaré sections, since no identification between the inlet and the outlet is possible for general non-cyclic flows and control volumes.

The ICI is inspired by communication theory (Shannon & Weaver 1949), which deals with information flow as opposed to fluid flow. The basic scheme of a communication system involves a transmitter, a receiver, and a communication channel. We can recognize parallel roles in those of, respectively, the inlet, the outlet, and the flow transport between them. This analogy is illustrated in figure 1. The transmitted message is a position at the inlet, which is received as a corresponding position at the outlet, connected by a flow streamline and observed at a particular scale. In general, noise sources are also present in a communication system, producing some loss of information in the communication flow. Shanon's mutual information (MI) measures the amount of information effectively communicated. Thus, for a perfectly laminar and parallel flow in a straight tube, the MI between input and output will be maximal. For more complex flows, generated by more intricate conditions or geometries, the ability to discriminate the outlet position by knowing the inlet position is decreased, reducing the corresponding MI.

Hence, MI is a natural candidate to quantify fluid flow complexity in open flow systems as it is conceptually related to fluid transport from inlet to outlet. The ICI is defined as the normalization of this MI, so that $ICI = 0$ for the simplest parallel flow, and $ICI \to 1$ for a very complex flow producing practically unpredictable positions at the outlet from the positions at the inlet.

The ICI is not a single index but a function of the observation scale (ICI_s). Thus, it can distinguish features appearing at different scales. The observation scale, s, can be introduced by different protocols, for instance by a discretization or by a ball of specific radius as in Mathew *et al.* (2005). Our implementation will follow this second approach but using Gaussian kernels instead of balls. To better understand the dependence of ICI_s with the scale, we derive its limit for infinitesimal scales in appendix B.

At first glance, the global mixing measure proposed by Alemaskin *et al.* (2004) and Guida *et al.* (2010) is the most similar to the ICI proposed here. It is independent of any scalar distribution and is defined as the average of Shannon's conditional entropy. However, there are important differences. First, the definition of MI and conditional entropy are related but not equivalent and, hence, have different properties. The most notable difference is that conditional entropy, as entropy, is not well defined for continuous variables. The entropy of continuous variables, or differential entropy, does not coincide with the limit of the discrete entropy and is not invariant under parameter changes (Shannon 1948; Kolmogorov 1956; Jaynes 1968, 2003). In contrast, the MI of continuous variables is the limit of the discrete case and is invariant to reparameterizations. Second, their mixing measure is based on the fluid particle propagation as a function of the time elapsed from the initial state and is observed in the full fluid domain. Thus, it is not applicable to open systems, i.e. those with specified inlets and outlets.

In order to properly define the ICI, the probability distribution of streamlines must be established. Herewith, we introduce the *natural probability distribution of streamlines*, which satisfies invariance with respect to the cross section used to generate the streamlines. This probability density distribution is a relevant contribution in itself, which could be applied for visualization purposes or for any other statistics based on streamlines. The natural distribution of streamlines is valid for any steady flow and for any incompressible flow (steady or unsteady). For unsteady flows, a more natural alternative to streamlines are pathlines. Taking into account the temporal dimension, pathlines correspond to world-lines in the 4-dimensional space-time. Thus, we also introduce the *natural distribution of world-lines*, with invariant properties for any flow satisfying the continuity equation. This involves the definition of the inlet and outlet as spatio-temporal hypersurfaces. For unsteady incompressible flows, two versions of ICI are possible, considering either streamlines or world-lines. Although in that case world-lines reflects the actual flow transport, for some applications, considering the streamlines could be interesting. In particular, this allows a neat discrimination of flow features at different time instants.

The paper is organized as follows. Section 2 introduces the natural distribution of streamlines, demonstrating its invariant properties, together with a rigorous definition of the inlet and outlet. Section 3 defines the ICI, providing an intuitive interpretation. To illustrate and explore the properties of the proposed ICI, we apply it to several flow fields obtained either analytically or from numerical simulations. Section 4 describes the geometries and flows used for the numerical experiments in the subsequent sections, and the algorithm used for estimating the ICI using a finite sample of streamlines. In § 5, we evaluate the accuracy and precision of the estimated ICI for different algorithm settings, and the invariance with respect to the cross section generating the streamlines. For the accuracy evaluation, we consider the analytical computation of the MI from a steady flow in a cylindrical straight tube, presented in appendix A. In § 6, we analyse the ICI of steady flows in two types of idealised geometries (non-coplanar double-bend tubes and rotating concentric cylinders) under different conditions. In § 7, the ICI of unsteady incompressible flows in more complex geometries (anatomically realistic vasculatures with aneurysms) is compared with subjective classifications of flow complexity and flow stability. In § 8, the world-lines version of ICI is obtained for unsteady compressible flows (2-dimensional (2D) periodic cellular flows), and its behaviour for infinitesimal scales, derived in appendix B, is tested. Finally, § 9 presents the general conclusions.

2. Natural distribution of streamlines

2.1. *Streamlines through a control volume*

Let us consider any flow field $\boldsymbol{v}(\boldsymbol{x},t)$ defined in a spatio-temporal domain $\mathcal{D}\times\mathcal{T}$, representing the local fluid speed at any point $\boldsymbol{x}\in\mathcal{D}$ at any time $t\in\mathcal{T}$, and a fluid density $\rho(\boldsymbol{x},t)$. A *streamline* at a time t is a line γ whose tangent vector is parallel to the flow field at that time at all points along the line. Parameterized with a pseudo-time τ,

$$\boldsymbol{\gamma}: I \to \mathcal{D}; \quad \tau \mapsto \boldsymbol{\gamma}(\tau), \quad \text{such that} \quad \frac{\mathrm{d}}{\mathrm{d}\tau}\boldsymbol{\gamma}(\tau) = \boldsymbol{v}(\boldsymbol{\gamma}(\tau),t),$$

where we consider the interval, $I\subset\mathbb{R}$, that maximally extends the streamline in $\mathcal{D}$. We will denote by γ (no boldface) the streamline trajectory regardless of its parameterization. The *congruence of streamlines*, $\Gamma(t)$, is the set of all streamlines in $\mathcal{D}$ at any given instant t. For simplicity, let us fix a time t, and assume it in the notation, as for $\boldsymbol{v}(\boldsymbol{x}) := \boldsymbol{v}(\boldsymbol{x},t)$ and $\Gamma := \Gamma(t)$.

Let us consider a control volume $\mathcal{D}_\mathcal{S}\subset\mathcal{D}$ enclosed by a control surface $\mathcal{S}$, with a strictly positive scalar field, $\alpha(\boldsymbol{x})>0\ \forall x\in\mathcal{D}$, defining a current $\boldsymbol{j}(\boldsymbol{x}) = \alpha(\boldsymbol{x})\,\boldsymbol{v}(\boldsymbol{x})$.

Definition 1. *At any point on the control surface, $\boldsymbol{x}\in\mathcal{S}$, the $\boldsymbol{j}$-*flux density *(flow rate per unit area) through the surface $\mathcal{S}$ will be*

$$\varphi_{\boldsymbol{j}}^{\mathcal{S}}(\boldsymbol{x}) := \boldsymbol{j}(\boldsymbol{x})\cdot\hat{\boldsymbol{n}}_\mathcal{S}(\boldsymbol{x}) \tag{2.1}$$

where $\hat{\boldsymbol{n}}_\mathcal{S}(\boldsymbol{x})$ is the unit normal vector to $\mathcal{S}$ at the point $\boldsymbol{x}$, pointing outwards. For any subregion $\mathcal{A}\subset\mathcal{S}$, the corresponding flux *will be the surface integral on the subregion,*

$$\Phi_{\boldsymbol{j}}(\mathcal{A}) := \int_\mathcal{A}\varphi_{\boldsymbol{j}}^{\mathcal{S}}(\boldsymbol{x})\ \mathrm{d}S = \int_\mathcal{A}\boldsymbol{j}(\boldsymbol{x})\cdot\mathbf{d}\boldsymbol{S}, \tag{2.2}$$

where $\mathrm{d}S$ *and* $\mathbf{d}\boldsymbol{S}$ *denote the scalar and vector surface element, respectively.*

In particular, we will consider the volume flow rate, $\varphi_{\boldsymbol{v}}^{\mathcal{S}}(\boldsymbol{x})$, corresponding to $\alpha(\boldsymbol{x})=1$, and the mass flow rate generated by $\boldsymbol{j}(\boldsymbol{x}) = \rho(\boldsymbol{x})\,\boldsymbol{v}(\boldsymbol{x})$.

The flux density will be negative at entry points and positive at exit points. Thus, the control surface $\mathcal{S}$ can be accordingly split into two regions:

$$\bar{\mathcal{S}}_\mathrm{In} := \{\boldsymbol{x}\in\mathcal{S}\,|\,\varphi_{\boldsymbol{v}}^{\mathcal{S}}(\boldsymbol{x})<0\} \qquad \text{and} \qquad \bar{\mathcal{S}}_\mathrm{Out} := \{\boldsymbol{x}\in\mathcal{S}\,|\,\varphi_{\boldsymbol{v}}^{\mathcal{S}}(\boldsymbol{x})>0\}.$$

There is indeed a third region, $\bar{\mathcal{S}}_0 := \{\boldsymbol{x}\in\mathcal{S}\,|\,\varphi_{\boldsymbol{v}}^{\mathcal{S}}(\boldsymbol{x})=0\}$, where the streamlines are tangent to the surface. In particular, if $\mathcal{S}$ includes rigid walls delimiting the fluid, they will be included in $\bar{\mathcal{S}}_0$.

A streamline $\gamma\in\Gamma$ can miss $\mathcal{D}_\mathcal{S}$, enter once, or repeatedly enter and exit $\mathcal{D}_\mathcal{S}$. Let us denote the number of times that the streamline enters and exits $\mathcal{D}_\mathcal{S}$ by, respectively,

$$r_\mathcal{S}^\mathrm{In}(\gamma) := \mathrm{Card}(\gamma\cap\bar{\mathcal{S}}_\mathrm{In}) \qquad \text{and} \qquad r_\mathcal{S}^\mathrm{Out}(\gamma) := \mathrm{Card}(\gamma\cap\bar{\mathcal{S}}_\mathrm{Out}).$$

The $r_\mathcal{S}^\mathrm{In}(\gamma)$ points in the intersection $\gamma\cap\bar{\mathcal{S}}_\mathrm{In}$ can be sorted according to the crossing order: $\boldsymbol{x}_{\mathrm{In},i}(\gamma)$, for $i=1,2,\ldots,r_\mathcal{S}^\mathrm{In}(\gamma)$. Analogously for the $r_\mathcal{S}^\mathrm{Out}(\gamma)$ points in $\gamma\cap\bar{\mathcal{S}}_\mathrm{Out}$.

Assuming that there is no sink or source inside the control volume, each streamline will in general exit as many times as it enters. Strictly speaking, some exceptional cases can appear, involving equilibrium points or cycles. However, under reasonable assumptions, they would represent a zero-measure set of streamlines, and will be discarded in the following definition.

Definition 2. *The* congruence of $\mathcal{S}$-crossing streamlines *will be the set*

$$\Gamma_\mathcal{S} := \left\{\,\gamma\in\Gamma\;\middle|\;r_\mathcal{S}^\mathrm{In}(\gamma) = r_\mathcal{S}^\mathrm{Out}(\gamma)>0\,\right\}.$$

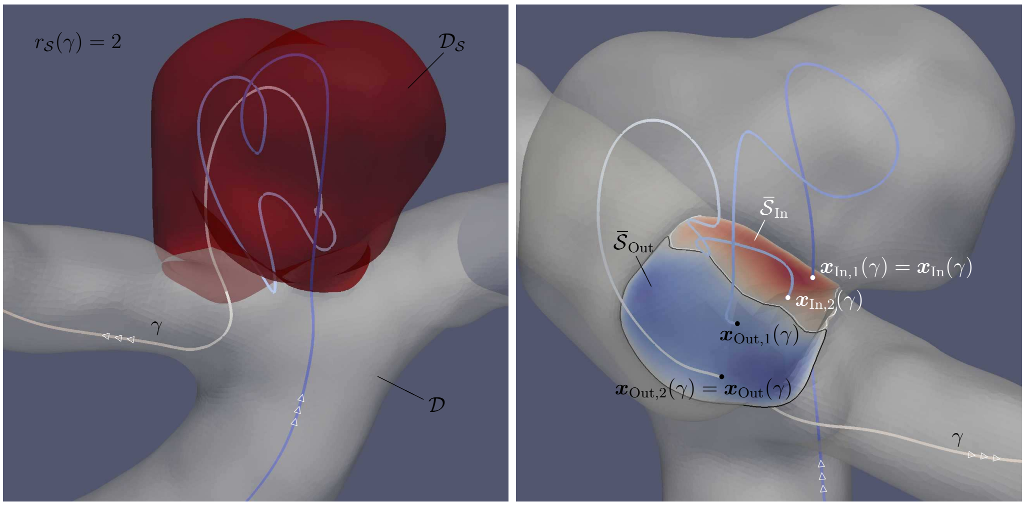

FIGURE 2. Example of recirculation at the entrance of a control volume $\mathcal{D}_{\mathcal{S}}$, represented in red (left). One streamline is shown with recirculation number $r_{\mathcal{S}}(\gamma) = 2$. The two inward points are represented in white, and the two outward points in black (right). The entrance is divided in two surfaces, $\bar{\mathcal{S}}_{\mathrm{In}}$ and $\bar{\mathcal{S}}_{\mathrm{Out}}$, according to the flux sign.

The number of times that a streamline $\gamma \in \Gamma_{\mathcal{S}}$ enters the control volume $\mathcal{D}_{\mathcal{S}}$ will be called its recirculation number, *and will be denoted by $r_{\mathcal{S}}(\gamma)$.*

Figure 2 illustrates these concepts.

DEFINITION 3. *For each streamline $\gamma \in \Gamma_{\mathcal{S}}$, the first inwards crossing point will be called its* inlet point, *$\boldsymbol{x}_{\mathrm{In}}(\gamma) := \boldsymbol{x}_{\mathrm{In},1}(\gamma)$, and the last outwards crossing point its* outlet point, *$\boldsymbol{x}_{\mathrm{Out}}(\gamma) := \boldsymbol{x}_{\mathrm{Out},r_{\mathcal{S}}(\gamma)}(\gamma)$. The sets containing all inlet and outlet points,*

$$\mathcal{S}_{\mathrm{In}} := \{\, \boldsymbol{x}_{\mathrm{In}}(\gamma) \mid \gamma \in \Gamma_{\mathcal{S}} \,\} \qquad \text{and} \qquad \mathcal{S}_{\mathrm{Out}} := \{\, \boldsymbol{x}_{\mathrm{Out}}(\gamma) \mid \gamma \in \Gamma_{\mathcal{S}} \,\}\,, \tag{2.3}$$

will be called inlet *and* outlet, *respectively.*

This definition guarantees that there exist bijective mappings between inlet, outlet, and the congruence of streamlines $\Gamma_{\mathcal{S}}$. Notice that, although the union of the streamlines fills the 3-dimensional domain, the congruence of streamlines is a 2-dimensional set. Since $\boldsymbol{x}_{\mathrm{In}} : \Gamma_{\mathcal{S}} \to \mathcal{S}_{\mathrm{In}}$ is a bijection, $\boldsymbol{x}_{\mathrm{In}}(\gamma)$ can be used as a parameterization of γ. Analogously, $\boldsymbol{x}_{\mathrm{Out}}(\gamma)$ gives an alternative parameterization.

2.2. *Probability distribution of streamlines*

The computation of any statistical quantity from a streamline distribution requires the definition of a probability density of streamlines. A natural selection is to make the probability density proportional to the local flux, thereby considering denser presence of streamlines where the flow rate is larger.

DEFINITION 4. *Parameterizing streamlines $\gamma \in \Gamma_{\mathcal{S}}$ by their crossing at the inlet $\boldsymbol{x} = \boldsymbol{x}_{\mathrm{In}}(\gamma)$, the* ($\boldsymbol{j}$-)natural distribution of streamlines *will be given by the surface probability density*

$$p_{\mathrm{In}}(\boldsymbol{x}) = \frac{\varphi_{\boldsymbol{j}}^{\mathcal{S}}(\boldsymbol{x})}{\Phi_{\boldsymbol{j}}(\mathcal{S}_{\mathrm{In}})} \qquad \forall \boldsymbol{x} \in \mathcal{S}_{\mathrm{In}}. \tag{2.4}$$

Any bundle of streamlines, $\Upsilon \subset \Gamma_{\mathcal{S}}$, will define a subregion of the inlet by its crossing,

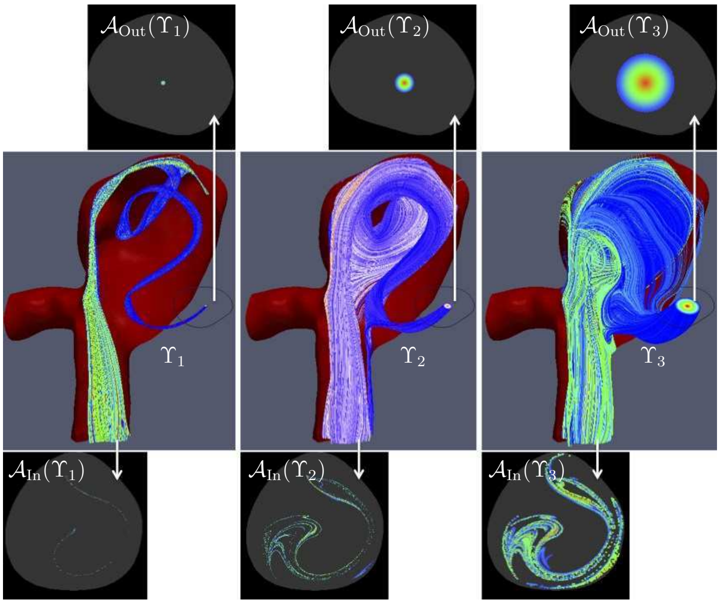

FIGURE 3. Three examples of streamlines bundles generated by their crossing at the outlet (top) defined as a circle around a point with three different radii. The crossing at the inlet (bottom) is very elongated and the complexity of the obtained pattern depends on the scale considered.

$\mathcal{A}_{\mathrm{In}}(\Upsilon) = \Upsilon \cap \mathcal{S}_{\mathrm{In}}$, and a subregion of the outlet, $\mathcal{A}_{\mathrm{Out}}(\Upsilon) = \Upsilon \cap \mathcal{S}_{\mathrm{Out}}$ (see figure 3). Accordingly, the probability of a streamline bundle $\Upsilon \subset \Gamma_{\mathcal{S}}$ will be

$$P_{\mathrm{In}}(\Upsilon) = \int_{\mathcal{A}_{\mathrm{In}}(\Upsilon)} p_{\mathrm{In}}(\boldsymbol{x})\,\mathrm{d}S = \frac{\Phi_{\boldsymbol{j}}\left(\mathcal{A}_{\mathrm{In}}(\Upsilon)\right)}{\Phi_{\boldsymbol{j}}(\mathcal{S}_{\mathrm{In}})}. \tag{2.5}$$

Analogously, $p_{\mathrm{Out}}(\boldsymbol{x})$ and $P_{\mathrm{Out}}(\Upsilon)$ are defined by their crossing at the outlet. Indeed, for any surface intersecting once every streamline in $\Gamma_{\mathcal{S}}$, we could define the corresponding probability density.

Strictly speaking, we are considering here *volumetric probabilities* instead of *probability densities*, in the sense discussed in Tarantola (2005) and Tarantola & Mosegaard (2007), since they are defined with respect to the surface element, $\mathrm{d}S$, and independently of any coordinate system.

LEMMA 1. *For any divergence-free current,* $\boldsymbol{\nabla}\cdot\boldsymbol{j}(\boldsymbol{x}) = 0$, *the flux corresponding to any bundle of streamlines* $\Upsilon \subset \Gamma_{\mathcal{S}}$ *is constant along the streamlines (invariant with respect to the cross section considered):*

$$\Phi_{\boldsymbol{j}}\left(\mathcal{A}_{\mathrm{In}}(\Upsilon)\right) = \Phi_{\boldsymbol{j}}\left(\mathcal{A}_{\mathrm{Out}}(\Upsilon)\right),$$

and its natural probability distribution of streamlines is invariant with respect to the cross section used to parameterize the streamlines:

$$P_{\mathrm{In}}(\Upsilon) = P_{\mathrm{Out}}(\Upsilon).$$

Thus, we can simply denote it as $P(\Upsilon)$.

Proof. It follows from the divergence theorem applied to small streamtubes. □

PROPOSITION 1. (Incompressible flows). *For an incompressible flow, the* $\boldsymbol{v}$*-natural* *probability distribution is invariant with respect to the cross section used to parameterize the streamlines.*

Proof. An incompressible flow is characterized by $\boldsymbol{\nabla} \cdot \boldsymbol{v} = 0$. ☐

PROPOSITION 2. (Steady flows). *For a steady flow, the* $\boldsymbol{j}$*-natural probability distribution, for the mass current* $\boldsymbol{j} = \rho\, \boldsymbol{v}$*, is invariant with respect to the cross section used to parameterize the streamlines.*

Proof. For a steady flow, the density at each point is temporally constant, $\partial\rho/\partial t = 0$. Then, the continuity equation implies

$$\frac{\partial}{\partial t}\rho + \boldsymbol{\nabla} \cdot (\rho\, \boldsymbol{v}) = \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0.$$

☐

Observe that for a homogeneous incompressible flow, the probability distribution generated by the mass flux coincides with the probability distribution generated by the volume flux.

PROPOSITION 3. (General flow with density propagated by (pseudo) advection). *The propagation by advection along any instantaneous flow* $\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x}, t_0)$*, of any density* $\alpha(x)$ *defined at the inlet or outlet, provides a conserved current,* $\boldsymbol{j} = \alpha\boldsymbol{v}$*, with invariant* $\boldsymbol{j}$*-natural probability density of streamlines.*

Proof. Let $\tau(\boldsymbol{x})$ denote the function providing the pseudo-time parameter, τ, in the corresponding streamline γ, so that $\boldsymbol{x} = \boldsymbol{\gamma}(\tau)$. Let us consider a density $c(\boldsymbol{x}, \tau)$ generated by advection along the flow $\boldsymbol{v}(\boldsymbol{x})$:

$$\frac{\partial c(\boldsymbol{x}, \tau)}{\partial \tau} + \boldsymbol{\nabla} \cdot \big(c(\boldsymbol{x}, \tau)\, \boldsymbol{v}(\boldsymbol{x})\big) = 0, \qquad \text{with} \quad c(\boldsymbol{x}_{\mathrm{In}}, \tau_{\mathrm{In}}) = \alpha(\boldsymbol{x}_{\mathrm{In}}) \quad \forall \gamma \in \Gamma_{\mathcal{S}}. \tag{2.6}$$

Then, the density $\alpha(\boldsymbol{x}) := c(\boldsymbol{x}, \tau(\boldsymbol{x}))$ provides a conserved current, $\boldsymbol{j}(\boldsymbol{x}) = \alpha(\boldsymbol{x})\, \boldsymbol{v}(\boldsymbol{x})$:

$$\boldsymbol{\nabla} \cdot \boldsymbol{j}(\boldsymbol{x}) = \boldsymbol{\nabla} \cdot \big(c(\boldsymbol{x}, \tau)\, \boldsymbol{v}(\boldsymbol{x})\big)\Big|_{\tau = \tau(\boldsymbol{x})} + \frac{\partial c(\boldsymbol{x}, \tau)}{\partial \tau}\bigg|_{\tau = \tau(\boldsymbol{x})} \boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{\nabla}\tau(\boldsymbol{x}) = 0,$$

where we have used (2.6) and the property $\boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{\nabla}\tau(\boldsymbol{x}) = 1$. ☐

In these cases, the natural probability density is a property of the streamlines, independently of the cross section considered. We will symbolically denote it as $p(\gamma)$. Note that the naive homogenous distribution would be cross-section dependent. In general, it would be inhomogeneous for any other cross section.

2.3. *Source of non-invariance of the natural distribution of streamlines*

Whenever the current $\boldsymbol{j} = \alpha\boldsymbol{v}$ is not divergence-free, $\boldsymbol{\nabla} \cdot \boldsymbol{j} \neq 0$, the corresponding flux cannot be guaranteed to be conserved, and the probability distribution of streamlines will not be invariant. In our experiments, we will consider homogeneous incompressible flows, which should satisfy $\boldsymbol{\nabla} \cdot \boldsymbol{j} = 0$. However, when a flow is obtained by computer flow simulation, numerical inaccuracies might arise, which justifies studying their influence in the probability distribution invariance.

LEMMA 2. *Given a streamline bundle,* $\mathrm{d}\Upsilon$*, of infinitesimal cross section (streamtube) around a streamline* $\gamma \in \Gamma_{\mathcal{S}}$*, propagating from inlet to outlet:*

$$\boldsymbol{\gamma}(\tau_{\mathrm{In}}) = \boldsymbol{x}_{\mathrm{In}} \in \mathcal{A}_{\mathrm{In}}(\mathrm{d}\Upsilon) \subset \mathcal{S}_{\mathrm{In}}, \qquad \textit{and} \qquad \boldsymbol{\gamma}(\tau_{\mathrm{Out}}) = \boldsymbol{x}_{\mathrm{Out}} \in \mathcal{A}_{\mathrm{Out}}(\mathrm{d}\Upsilon) \subset \mathcal{S}_{\mathrm{Out}},$$

the fluxes of the current $\boldsymbol{j} = \alpha \boldsymbol{v}$ *at the inlet and outlet are related by*

$$\Phi_{\boldsymbol{j}}(\mathcal{A}_{\mathrm{Out}}(\mathrm{d}\Upsilon)) = \mathrm{e}^{\xi(\gamma)}\, \Phi_{\boldsymbol{j}}(\mathcal{A}_{\mathrm{In}}(\mathrm{d}\Upsilon)),$$

where the factor exponent is given by the temporal integral of the flow divergence along the streamline,

$$\xi(\gamma) := \int_{\tau_{\mathrm{In}}}^{\tau_{\mathrm{Out}}} \frac{1}{\alpha(\boldsymbol{\gamma}(\tau))} \boldsymbol{\nabla} \cdot \boldsymbol{j}\, (\boldsymbol{\gamma}(\tau))\, \mathrm{d}\tau.$$

For divergence-free currents, $\xi(\gamma) = 0$, so that the flux is conserved. Thus, the factor exponent $\xi(\gamma)$ is the source of non-invariance of the corresponding natural distribution of streamlines. For incompressible flows ($\boldsymbol{j} = \boldsymbol{v}$), the integral simplifies to

$$\xi(\gamma) = \int_{\tau_{\mathrm{In}}}^{\tau_{\mathrm{Out}}} \boldsymbol{\nabla} \cdot \boldsymbol{v}\, (\boldsymbol{\gamma}(\tau))\, \mathrm{d}\tau. \tag{2.7}$$

Proof. Let us consider any density $\alpha'(\boldsymbol{x})$ generated by advection along $\boldsymbol{v}(\boldsymbol{x})$, as defined in proposition 3. Then, the current $\boldsymbol{j}' = \alpha' \boldsymbol{v}$ provides a conserved flux. Since the new current is proportional to the original one, $\boldsymbol{j}' = \beta \boldsymbol{j}$, where $\beta = \alpha'/\alpha$, their fluxes are also proportional. Thus, the conservation of $\boldsymbol{j}'$ provides the relationship

$$\Phi_{\boldsymbol{j}}(\mathcal{A}_{\mathrm{Out}}(\mathrm{d}\Upsilon)) = \frac{\beta(\boldsymbol{x}_{\mathrm{In}})}{\beta(\boldsymbol{x}_{\mathrm{Out}})} \Phi_{\boldsymbol{j}}(\mathcal{A}_{\mathrm{In}}(\mathrm{d}\Upsilon)). \tag{2.8}$$

Besides,

$$\boldsymbol{\nabla} \cdot \boldsymbol{j}' = 0 \quad \Rightarrow \quad \boldsymbol{j} \cdot \boldsymbol{\nabla}\beta + \beta\, \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0 \quad \Rightarrow \quad \boldsymbol{j} \cdot \boldsymbol{\nabla} \log(\beta) = -\boldsymbol{\nabla} \cdot \boldsymbol{j}. \tag{2.9}$$

For any point in the streamline, $\boldsymbol{x} = \boldsymbol{\gamma}(\tau)$, the directional derivative along $\boldsymbol{j} = \alpha \boldsymbol{v}$ can be expressed in terms of the time derivative:

$$\boldsymbol{j} \cdot \boldsymbol{\nabla} \log(\beta(\boldsymbol{x})) = \alpha(\boldsymbol{x}) \frac{\mathrm{d}}{\mathrm{d}\tau} \log(\beta(\boldsymbol{\gamma}(\tau))). \tag{2.10}$$

Substituting (2.10) into (2.9) and integrating along the streamline from inlet to outlet:

$$\log \frac{\beta(\boldsymbol{x}_{\mathrm{In}})}{\beta(\boldsymbol{x}_{\mathrm{Out}})} = \int_{\tau_{\mathrm{In}}}^{\tau_{\mathrm{Out}}} \frac{1}{\alpha(\boldsymbol{\gamma}(\tau))}\, \boldsymbol{\nabla} \cdot \boldsymbol{j}(\boldsymbol{\gamma}(\tau))\, \mathrm{d}\tau = \xi(\gamma).$$

□

2.4. *Natural probability distribution of world-lines*

For unsteady flows, particle trajectories (pathlines) are generally different from streamlines. Pathlines are normally understood as the trajectories in the space, considering the temporal dimension only as the pathline parameter,

$$\boldsymbol{\gamma} : I \to \mathcal{D}; \quad t \mapsto \boldsymbol{\gamma}(t), \quad \text{such that} \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\gamma}(t) = \boldsymbol{v}(\boldsymbol{\gamma}(t), t).$$

However, particle trajectories can be understood as streamlines in the 4-dimensional (4D) Galilean space-time, called *world-lines.* Let us consider the 4D Cartesian coordinates and the 4D velocity field

$$x^{\mu} = (x^0, x^1, x^2, x^3) = (t, \boldsymbol{x}) \qquad \text{and} \qquad v^{\mu}(\{x^{\nu}\}) = (1, \boldsymbol{v}(\boldsymbol{x}, t)),$$

where Greek indices range from 0 to 3. This notation is partially borrowed from special relativity physics, but it is compatible with Galilean relativity and Newtonian mechanics.

World-lines represent the particle trajectories in the space-time and are computed as the streamlines of this 4D flow:

$$\gamma^\mu : I \to \mathcal{T} \times \mathcal{D}; \quad \tau \mapsto \gamma^\mu(\tau), \quad \text{such that} \quad \frac{\mathrm{d}}{\mathrm{d}\tau}\,\gamma^\mu(\tau) = v^\mu(\{\gamma^\mu(\tau)\}),$$

which, together with the initial condition $\gamma^0(\tau_0) = \tau_0$, is equivalent to the pathline equation

$$\gamma^0(\tau) = \tau \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}\tau}\boldsymbol{\gamma}(\tau) = \boldsymbol{v}(\boldsymbol{\gamma}(\tau), \tau).$$

The 4D nabla operator is denoted

$$\nabla_\mu = \left(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\right) = \left(\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right).$$

The mass flow current is then $j^\mu := \rho v^\mu$, for which the continuity equation is simply written in terms of the 4D divergence:

$$\nabla_\mu j^\mu = \frac{\partial}{\partial t}\,\rho + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) = 0. \tag{2.11}$$

Hence, all definitions and properties described above for streamlines of incompressible flows are valid for world-lines of any (mass preserving) flow. Thus the *natural distribution of world-lines* appears as follows. Considering a 4D control hyper-volume, its boundary would be a 3-dimensional hyper-surface, from which the corresponding inlet and outlet hyper-surfaces can be defined according to the flux density derived from j^μ. This flux defines then a probability density invariant to the cross-sectional hyper-surface.

3. Interlacing complexity index based on mutual information

If a small area, $\mathcal{A}_{\mathrm{In}}$, is singled out at the inlet, the corresponding bundle of streamlines, Γ, will determine the corresponding outlet area, $\mathcal{A}_{\mathrm{Out}}$. Depending on the complexity of the flow from inlet to outlet, a simple and compact input area can correspond to a complex and elongated output (see figure 3). We propose to measure the flow complexity by an index derived from Shanon's mutual information (MI) between corresponding inlet position, $\boldsymbol{x}_{\mathrm{In}}$, and outlet position, $\boldsymbol{x}_{\mathrm{Out}}$, of the streamlines, $\gamma \in \Gamma_{\mathcal{S}}$.

The MI of two variables of interest, $\boldsymbol{X}_{\mathrm{In}}$ and $\boldsymbol{X}_{\mathrm{Out}}$, is computed from the corresponding joint probability density, $p(\boldsymbol{x}_{\mathrm{In}}, \boldsymbol{x}_{\mathrm{Out}})$:

$$I(\boldsymbol{X}_{\mathrm{In}}, \boldsymbol{X}_{\mathrm{Out}}) = \int_{\mathcal{S}_{\mathrm{In}}} \int_{\mathcal{S}_{\mathrm{Out}}} p(\boldsymbol{x}_{\mathrm{In}}, \boldsymbol{x}_{\mathrm{Out}}) \log\left(\frac{p(\boldsymbol{x}_{\mathrm{In}}, \boldsymbol{x}_{\mathrm{Out}})}{p(\boldsymbol{x}_{\mathrm{In}})\, p(\boldsymbol{x}_{\mathrm{Out}})}\right) \mathrm{d}S_{\mathrm{In}}\, \mathrm{d}S_{\mathrm{Out}} \tag{3.1}$$

and it is a measure of how much is known of one variable by knowing the other. Here $p(\boldsymbol{x}_{\mathrm{In}})$ and $p(\boldsymbol{x}_{\mathrm{Out}})$ are the corresponding marginal probability densities, and we use natural logarithms. MI is well defined for continuous stochastic variables, being the correct limit of the discrete MI and invariant to any change of parameters or coordinates.

As described in the introduction and illustrated in figure 1, there is a clear analogy of the transport from inlet to outlet by the flow streamlines with communication channels in information theory. The emitted message is encoded in an inlet position $\boldsymbol{x}_{\mathrm{In}} \in \mathcal{S}_{\mathrm{In}}$, and the received message is encoded by the position at the outlet, $\boldsymbol{x}_{\mathrm{Out}} \in \mathcal{S}_{\mathrm{Out}}$. The channel is the flow, which transmits the message following the streamlines.

Given a flow, the corresponding streamlines are exact lines, where no stochastic process is considered. Thus, the streamline position at the inlet, $\boldsymbol{x}_{\mathrm{In}}$, is completely determined from its position at the outlet, $\boldsymbol{x}_{\mathrm{Out}}$, and vice versa, defining a bijective function. This is analogous to a noise-free channel, providing exactly predictable continuous variables, for which MI will be infinite. However, this description assumes infinite-precision observation of the positions. To measure the flow complexity, we need to consider a spatial scale with respect to which the mixing is evaluated. This will be analogous to considering two noise sources in this channel, one at the selection of the streamline from the inlet point, and another at the determination of the outlet point from the streamline, affecting their precision (figure 1). This noise is not, however, a characteristic of the flow system, but imposed by the observer. Thus, the proposed complexity index will be relative to this scale, providing actually a function of the scale, instead of a single index. Note also that the introduction of the observation scale does not make the variables discrete; they are still continuous but with local uncertainty.

3.1. *Scale-dependent probability model*

A typical method to estimate a probability density from a sample is based on binning the variables of interest and extracting the frequencies in each bin (histogram). If the bin size is small compared with the spatial frequency of change of the considered probability density, this discretization provides a valid approximation. However, for the joint probability density, $p(\boldsymbol{x}_{\mathrm{In}}, \boldsymbol{x}_{\mathrm{Out}})$, defined with infinite precision, any binning would be, by definition, large. Thus, it would impose a coarse-grained scale of observation. This observation scale is, indeed, what we need to introduce in the probability model. Unfortunately, this binning method has sometimes led to confusion between two different concepts: observation scale and numerical estimation. Thus, for clarity, we will explicitly decouple both.

We will introduce the scale by considering that each streamline does not correspond to an exact point at the inlet or outlet, but to a distribution around that point, which we will call *point spread function* (PSF), noticing its analogy with imaging systems. The PSF is given by independent conditional probability densities at the inlet, $p_{\mathrm{In},s}(\boldsymbol{x} \,|\, \gamma)$, and outlet, $p_{\mathrm{Out},s}(\boldsymbol{y} \,|\, \gamma)$, providing the joint probability density

$$p_s(\boldsymbol{x}, \boldsymbol{y}) = \int_{\Gamma_{\mathcal{S}}} p_{\mathrm{In},s}(\boldsymbol{x} \,|\, \gamma)\, p_{\mathrm{Out},s}(\boldsymbol{y} \,|\, \gamma)\, p(\gamma)\, \mathrm{d}\gamma \tag{3.2}$$

where $\boldsymbol{x} \in \mathcal{S}_{\mathrm{In}}$ and $\boldsymbol{y} \in \mathcal{S}_{\mathrm{Out}}$. The index s identifies the scale considered. As introduced in § 2.2, $p(\gamma)$ denotes the natural probability density of streamlines. The integral can be computed either at the inlet or the outlet, parameterizing the streamlines by the corresponding crossing points.

From this joint probability density, we obtain the MI relative to the PSF at scale s:

$$I_s(\boldsymbol{X}, \boldsymbol{Y}) = \int_{\mathcal{S}_{\mathrm{In}}} \int_{\mathcal{S}_{\mathrm{Out}}} p_s(\boldsymbol{x}, \boldsymbol{y}) \log\left(\frac{p_s(\boldsymbol{x}, \boldsymbol{y})}{p_s(\boldsymbol{x})\, p_s(\boldsymbol{y})}\right) \mathrm{d}S_{\mathrm{In}}\, \mathrm{d}S_{\mathrm{Out}}. \tag{3.3}$$

In contrast with the binning approach, here the scale is explicitly stated as part of the exact definition of the MI, and decoupled from the discretization needed for its numerical estimation.

It is possible to define PSFs equivalent to the binning method, using a discretization of the inlet and outlet. However, in that case, the definition of the joint probability density and its MI will be dependent, not only on the scale, but also on the particular partition considered. Besides, the definition of a discretization is not straightforward for inlets or outlets of general irregular shape.

We will consider as PSF, Gaussian kernels with isotropic standard deviation σ:

$$p_{\mathrm{In},s}(\boldsymbol{x}\,|\,\gamma) = \frac{1}{C}\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_{\mathrm{In}}(\gamma)\|^2}{2\sigma^2}\right), \qquad C = \int_{\mathcal{S}_{\mathrm{In}}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_{\mathrm{In}}(\gamma)\|^2}{2\sigma^2}\right) \mathrm{d}S_{\mathrm{In}}, \tag{3.4}$$

where the normalization constant depends on the inlet geometry. The standard deviation, $\sigma = sR$, is defined as proportional to the scale, s, and the inlet *area-equivalent* radius, $R = \sqrt{\mathrm{Area}(\mathcal{S}_{\mathrm{In}})/\pi}$. For the outlet, $p_{\mathrm{Out},s}(\boldsymbol{y}\,|\,\gamma)$ is analogously defined. This PSF can be applied independently of the geometry of the control volume, is homogeneous and isotropic by construction, and produces continuous and smooth probability densities.

3.2. *Interlacing complexity index (ICI)*

Several versions of normalized MI for discrete variables exist (McDaid *et al.* 2011; Cahill 2010; Vinh *et al.* 2010). They run along the finite range between 0 (if and only if the two variables are independent) and 1 (if and only if there exists a bijection between them). The normalization facilitates interpretation and comparison across different conditions, where the (unnormalized) MI might have different ranges. Two of the normalized MIs are similarity metrics (Chen *et al.* 2009),

$$NMI_{\mathrm{max}} = \frac{I(\boldsymbol{X},\boldsymbol{Y})}{\max\{H(\boldsymbol{X}),H(\boldsymbol{Y})\}} \qquad \text{and} \qquad NMI_{\mathrm{joint}} = \frac{I(\boldsymbol{X},\boldsymbol{Y})}{H(\boldsymbol{X},\boldsymbol{Y})}$$

and their complement,

$$d_{\mathrm{max}} = 1 - NMI_{\mathrm{max}} \qquad \text{and} \qquad d_{\mathrm{joint}} = 1 - NMI_{\mathrm{joint}},$$

are normalized distance metrics (Vinh *et al.* 2010) (satisfying positive definiteness, symmetry and triangle inequality). Among them, NMI_{max} seems to provide the most linear and intuitive behaviour (McDaid *et al.* 2011).

Unfortunately, none of the normalized MI versions is directly applicable to continuous variables, since the normalization factor involves either the entropy of each variable or the joint entropy, which are inconsistent for continuous variables. For discrete variables the entropy can be expressed as self-information, $H(\boldsymbol{X}) = I(\boldsymbol{X},\boldsymbol{X})$, but it diverges for continuous variables. However, analogous to (3.2), we can introduce a self-probability density dependent on the observation scale:

$$p_s(\boldsymbol{x}_1,\boldsymbol{x}_2) = \int_{\Gamma_S} p_{\mathrm{In},s}(\boldsymbol{x}_1\,|\,\gamma)\, p_{\mathrm{In},s}(\boldsymbol{x}_2\,|\,\gamma)\, p(\gamma)\,\mathrm{d}\gamma\ , \tag{3.5}$$

and the corresponding inlet self-information, $I_s(\boldsymbol{X},\boldsymbol{X})$, analogously to (3.3), but integrating both variables in $\mathcal{S}_{\mathrm{In}}$. In case the scale was introduced by a discretization, it can be proved that $I_s(\boldsymbol{X},\boldsymbol{X})$ coincides with the entropy of the corresponding discrete variable. Thus, this allows us to introduce a generalization of the normalization involved in NMI_{max} and d_{max}.

DEFINITION 5. *The* Interlacing Complexity Index (ICI) *at scale s, is defined from the the mutual information* (3.3) *of the joint distribution given by* (3.2), *as*

$$ICI_s = 1 - \frac{I_s(\boldsymbol{X},\boldsymbol{Y})}{\max\{I_s(\boldsymbol{X},\boldsymbol{X}),I_s(\boldsymbol{Y},\boldsymbol{Y})\}}\ . \tag{3.6}$$

Note that the ICI is not directly the normalized distance d_{max}, but a generalization of it. Thus, we have not proved that it satisfies all its properties. However, the results of our experiments are compatible with the triangle inequality and the expected range,

$ICI_s \in [0,1]$. In particular, $ICI_s = 0$ for the case of a steady flow in a straight cylindrical tube, where the streamlines are straight and parallel. ICI increases with the complexity of the flow, approaching $ICI_s = 1$ when the positions at the outlet are highly unpredictable from the positions at the inlet.

Figures 7, 9, 11, and 15 show the ICI obtained for different types of geometries and flows, which are described in the following sections. Each of them is presented together with a corresponding illustration of the complexity introduced by the interlacing of the streamlines when propagating from the inlet to the outlet. This provides an intuitive understanding of the meaning of the proposed ICI as a flow complexity quantification.

Observe that all the elements involved in the definition of the ICI are valid for an inlet and outlet of any geometry and topology. In particular, inlet and outlet may consist of multiple, separate regions. which would be analogous to a multiple-inputs multiple-outputs (MIMO) communication system.

4. Numerical experiments design

4.1. *Flow cases*

In §§ 5, 6, 7, and 8, we present several numerical experiments involving 5 different flow cases. Their characteristics are summarised in table 1. They illustrate a diversity of geometries and flow conditions, including steady and unsteady flows, and different inlet and outlet geometries and topologies. In all cases, except for case (v), rigid impermeable walls and incompressible Newtonian flow is assumed.

Depending on the case, the flow fields and streamlines have been obtained analytically or by numerical simulation and integration. For the simplest case (i), flow field, streamlines, and MI have been computed both analytically and numerically. This have been used in § 5 to estimate the accuracy of the algorithms and to test the correctness of their implementation. Cases (ii) and (iii) represent idealized geometries and steady flows, and their corresponding ICI under different conditions is investigated in § 6. In contrast, case (iv) includes anatomically realistic vasculatures and unsteady flows, which are presented and investigated in § 7. A subset of them has been also used in § 5 to evaluate the algorithm accuracy in more complex cases. Case (v) is analytically defined and represents a family of unsteady compressible 2D periodic cellular flows. In § 8, we analyse the obtained ICI_s for different parameter values and test its limit for infinitesimal scales.

4.2. *Flow simulations*

From the surface triangular meshes representing the different geometries, unstructured volumetric meshes have been created using an octree approach with ICEM CFD 13.0 (ANSYS, Canonsburg, PA, USA). The considered geometries have a typical radius of 2 mm, and the resulting meshes are composed of tetrahedral elements with a side length of 0.24 mm and, at the wall, three layers of prism elements with a total height of 0.08 mm and a side length of 0.12 mm. This mesh configuration was selected according to the mesh convergency study performed in the project @neurist (Aneurist Consortium 2010; Villa-Uriol *et al.* 2011).

The flow fields have been simulated with the vertex-centred finite volume solver CFX 13.0 (ANSYS) for Navier–Stokes equations, using a second order advection scheme, a second order backward Euler transient scheme for unsteady simulations, and CFX' automatic time scale control for steady-state simulations. An incompressible Newtonian fluid with viscosity $\mu = 3.5\,\mathrm{mPa\ s}$ and density $\rho = 1066\,\mathrm{kg\ m^{-3}}$ (typical values for blood) was considered.

Rigid walls, volumetric inflow rate, and parabolic velocity profile in the inlet, with

	Geometry	Steadiness	Variants	Flow & streamlines computation	MI & ICI computation
(i)	Straight cylindrical tube	Steady	-	Analytical & Numerical	Analytical & Numerical
(ii)	Rotating concentric cylinders	Steady	3 different ratios internal/external radii	Analytical	Numerical
(iii)	Non-coplanar double-bend tube	Steady	5 inter-bend angles × 2 Reynolds numbers	Numerical	Numerical
(iv)	Cerebral vasculatures with aneurysm	Unsteady pulsatile	49 subjects	Numerical	Numerical
(v)	2D Periodic cellular flow	Unsteady	4 unsteadiness × 3 compressibilities	Analytical & Numerical	Numerical

Table 1. Summary of the characteristics of the 4 flow cases considered for the numerical experiments.

moderate Reynolds numbers ($Re \leqslant 500$), have been imposed as boundary conditions. Steady flow have been simulated for the straight (i) and the double-bend tubes (iii), with maximum residual of 10^{-5}.

Unsteady cyclic flow has been simulated for the arteries with aneurysm (iv). The specific boundary conditions at the inlet (flow rate) and outlets (pressure) were computed for a cardiac cycle using a one-dimensional model of the human vascular tree (Reymond *et al.* 2009). The cardiac cycle, with a period of 0.8 s, has been discretized in 160 uniformly distributed time steps (5 ms) and, to reduce the effect of initial transients, the third of three simulated cardiac cycles has been analysed. A maximum residual of 5×10^{-3} was set as convergence criterion. This corresponds to the state-of-the art protocol for haemodynamic simulation studies defined in the @neurist project. Although not strictly converged in some cases, it represents a balance between accuracy and computational time, being sufficient to illustrate how the ICI could be used in practical applications. The protocol and convergence criterion is stricter for the idealized flow simulations.

For cases (ii) and (v) the flow fields are obtained analytically.

4.3. *Streamlines computation*

From the flow velocity fields, the streamlines have been obtained by numerical integration using 4th order Runge–Kutta algorithm, implemented in the Visualization ToolKit library (VTK community 2014; Schroeder *et al.* 2003). The integration is controlled by the length-step, which we have set to of 1/20 of the cell length. Inside each cell the velocity is linearly interpolated from the values at the cell nodes. The integration is terminated if either the number of steps reach 10^5 or the speed is below 10^{-100} m s^{-1}. Observe that this termination speed will be virtually never reached, since those stagnant streamlines have null provability according to the natural distribution of streamlines.

The seed points are selected according to the natural distribution (2.4) in either inlet or outlet. The streamlines are propagated in both directions, forwards and backwards. In the inlet case, recirculation is controlled by discarding any seed point for which the backwards propagated streamline crosses again the inlet. In the outlet case, the seed is discarded if the forwards propagated streamline crosses again the outlet. This algorithm implements definition 3 of inlet and outlet.

The streamlines sample is generated on the fly from the flow field while computing the MI. In order to reduce the memory consumption, they are generated in groups of 2k. Then, the inlet and outlet crossings, $\boldsymbol{x}_{\mathrm{In}}(\gamma)$ and $\boldsymbol{x}_{\mathrm{Out}}(\gamma)$, are identified and the streamlines are removed to free memory. For the different experiments, MI has been computed from sets of streamlines going from 5k to 100k.

4.4. *ICI estimation*

The joint probability density (3.2) can be estimated from a sample of N streamlines, $\{\gamma_\alpha\}$, generated from any flow cross section (in particular, either inlet or outlet) with the natural probability density (2.4):

$$p_s(\boldsymbol{x},\boldsymbol{y}) \simeq \frac{1}{N}\sum_{\alpha=1}^{N} p_{\mathrm{In},s}(\boldsymbol{x}\,|\,\gamma_\alpha)\,p_{\mathrm{Out},s}(\boldsymbol{y}\,|\,\gamma_\alpha) \tag{4.1}$$

The factor $p(\gamma)$ disappears because it is taken into account when generating the sample of streamlines. Then, Monte Carlo integration (Robert & Casella 2013) is applied for the computation of the MI in (3.3), using uniform samplings of M points in both the inlet and the outlet. The same procedure is applied for the estimation of $I_s(\boldsymbol{X},\boldsymbol{Y})$, $I_s(\boldsymbol{X},\boldsymbol{X})$ and $I_s(\boldsymbol{Y},\boldsymbol{Y})$. From them, ICI_s is computed as defined in (3.6). For computation and memory efficiency, the PSF is truncated for distances $> 3.717\sigma$, corresponding to the 99.9% confidence region, and sparse matrices are used in the algorithm.

This algorithm is linear in the number of streamlines, N, and quadratic in the number of points, M. The complete algorithm, including the streamlines computation, has been implemented in C++. All the experiments have been run in a MacBook Pro, 2.8GHz dual-core Intel Core i7, 16GB SDRAM. The mean computational time for the full process, from flow field to ICI_s, for the scale sequence $s=\frac{1}{3},\frac{1}{4},\ldots,\frac{1}{20}$, with $N=20\mathrm{k}$ and $M=5\mathrm{k}$, was 5 minutes. But this time depends on the geometry extent and the flow complexity.

5. Accuracy, precision and invariance

5.1. *Algorithm accuracy and precision evaluation*

To check the correctness and accuracy of the algorithm described above for the estimation of the ICI, we have considered the steady flow in a cylindrical straight tube (i), and the pulsatile flow in a subset of the dataset of anatomically realistic vasculatures with aneurysm (iv).

For case (i), the MI can be semi-analytically computed. This computation and the resulting values for $I_s(\boldsymbol{X},\boldsymbol{Y})$ are presented in appendix A. Alternatively, the corresponding MI has been also numerically computed from the flow simulated for a tube of radius $R=2\,\mathrm{mm}$ and length $L=0.3\,\mathrm{m}$, and a flow rate $Q=5\,\mathrm{ml\,s^{-1}}$, corresponding to $Re=485$. $I_s(\boldsymbol{X},\boldsymbol{Y})$ has been estimated for a series of streamline samples of sizes $N=5\mathrm{k},10\mathrm{k},20\mathrm{k},50\mathrm{k}$, and of point samples of sizes $M=2\mathrm{k},5\mathrm{k},10\mathrm{k},20\mathrm{k}$. Each combination has been instantiated 8 times to evaluate also their variability. The resulting values have been compared with the semi-analytical computation, considered as the ground truth (GT). Figure 4 presents the distribution of signed relative errors, $(I_s - I_s^{(\mathrm{GT})})/I_s^{(\mathrm{GT})}$, obtained with the different sample sizes, for scales $s=\frac{1}{3},\frac{1}{10},\frac{1}{20}$.

To estimate the algorithm precision and accuracy in more complex geometries and flows, we have selected a random subset of 10 cases from the dataset of cerebral vasculatures with aneurysms (iv), considering their ICI_s computed from the streamlines at one cardiac phase (peak systole). Since no analytical ground truth is available in this case, each ICI_s has been computed using $N=100\mathrm{k}$ streamlines and $M=50\mathrm{k}$ points, which is in the limit of the feasible computation, needing for each case 20 hours

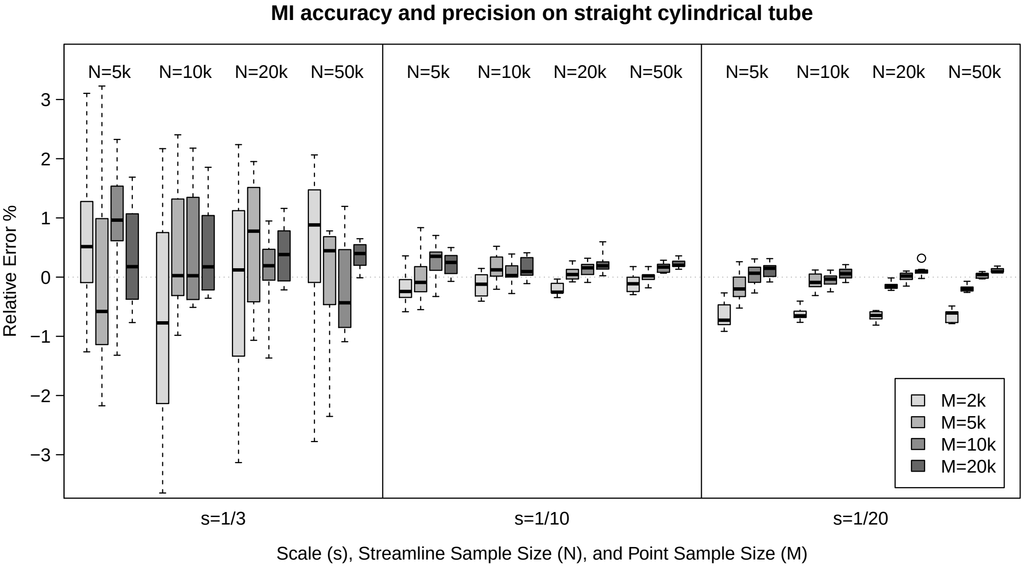

FIGURE 4. Evaluation of the accuracy and precision of the algorithm implementation for the computation of the MI. For a stationary flow in a cylindrical straight tube, the numerical estimation is compared with the semi-analytical computation. The signed error is evaluated for 3 scales (s), 4 streamline sample sizes (N) and 4 point sample sizes (M).

	$N = 5\text{k}, M = 2\text{k}$	$N = 20\text{k}, M = 5\text{k}$	$N = 50\text{k}, M = 20\text{k}$
$s = 1/3$	0.001 ± 0.032	0.002 ± 0.012	0.001 ± 0.008
$s = 1/10$	-0.005 ± 0.010	0.000 ± 0.006	0.000 ± 0.003
$s = 1/20$	-0.018 ± 0.022	-0.003 ± 0.007	-0.001 ± 0.003

TABLE 2. Accuracy ($\overline{\varepsilon}$) and precision ($2\sigma_\varepsilon$) of ICI_s for 3 scales and 3 settings of N and M, estimated from the flow in 10 vasculatures with aneurysm. The values are expressed as $\overline{\varepsilon} \pm 2\sigma_\varepsilon$ (approximate 95% confidence interval).

of computation with peaks of 15 GB of memory usage. The obtained values have been taken as ground truth. The ICI has been also computed for $N = 5\text{k}, 10\text{k}, 20\text{k}, 50\text{k}$ and $M = 2\text{k}, 5\text{k}, 10\text{k}, 20\text{k}$, instantiated twice for each combination. For each scale, the signed error is given by the difference with respect to the ground truth, $\varepsilon_s = ICI_s - ICI_s^{(\mathrm{GT})}$. Since the ICI is normalized between 0 and 1, this absolute error is directly interpretable as relative to the index range and comparable between scales. Figure 5 presents the obtained distribution of errors. Table 2 presents the estimated accuracy and precision for 3 different settings for N and M, where the accuracy is quantified by the mean signed error, $\overline{\varepsilon}$, and the precision by twice its standard deviation, $2\sigma_\varepsilon$, representing approximately the 95% confidence interval ($\overline{\varepsilon} \pm 2\sigma_\varepsilon$).

As expected, both accuracy and precision increase with the number of streamlines (N) and with the number of points (M). For the straight tube, the convergence to the semi-analytical values evidences the correctness of the algorithm implementation.

The accuracy is degraded from coarser scales (large s) to finer scales (small s), and from larger samples sizes, N and M, to smaller ones. This bias for small samples is a known effect for experimental MI estimators (Paninski 2003). It is interesting that the

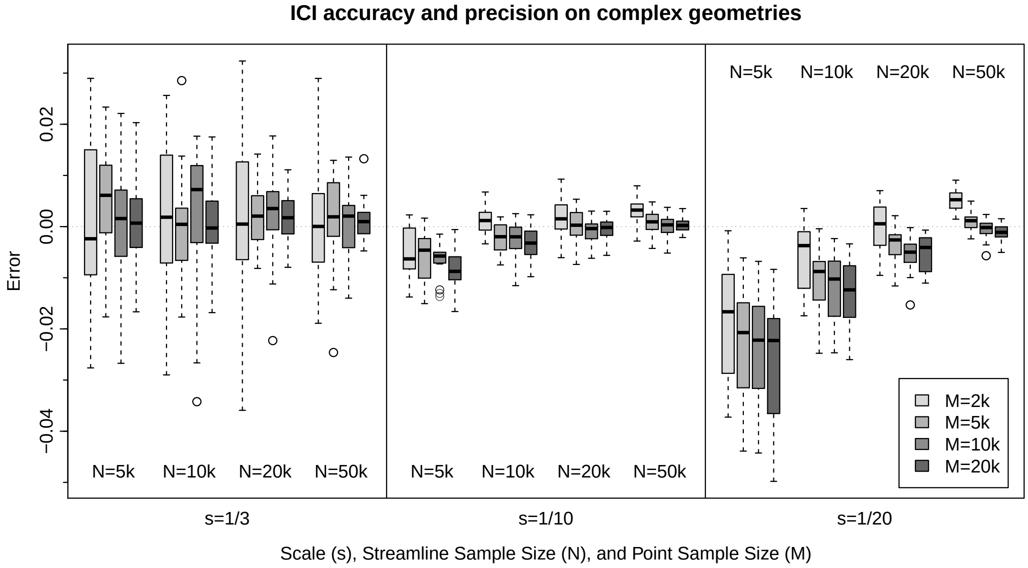

FIGURE 5. Evaluation of the accuracy and precision of ICI_s computation on the flow at peak systole in 10 vasculatures with aneurysm. For each scale (s), the signed error against the ground truth is evaluated for several streamline sample sizes (N) and point sample sizes (M).

bias introduced by small N and M acts in opposite directions. Reducing the number of streamlines (N) should presumably reduce the apparent complexity, which explains the observed decrease in the ICI (figure 5), specially for $s = 1/20$. Accordingly, this bias dependence with N is not observed for the simple flow in the straight tube (figure 4). For smaller s, the area covered by the PSF is smaller, requiring denser samples. This explains the bias increase when s decreases, for the same N and M.

The estimation of ICI is most precise for the intermediate scale, reducing the precision towards both extremes. Two competing effects appear here. The variability of the MI estimator increases for small sample sizes compared with the observation scale. Thus, for the same N and M, the MI variability decrease when s increase. However, the ICI normalization factor also decreases, magnifying the ICI errors for the same MI errors.

From the obtained values, a reasonable selection seems to be $N = 20$k and $M = 5$k, involving an acceptable mean computational time of 5 minutes per case, representing a balance between accuracy and computational time and memory. Hence, these algorithm settings have been used in the subsequent experiments.

5.2. *Streamline distribution invariance*

Although the natural probability distribution of streamlines is analytically invariant to the cross section generating them, this invariance can be broken by the numerical approximations involved in the flow simulations. As shown in § 2.3, the source of variation of the streamline distribution is the integral of the flow divergence, $\xi(\gamma)$, along each streamline, γ, as defined in (2.7).

For testing this invariance and the influence of the numerical errors, we have considered the unsteady flow cases (iv) in cerebral vasculatures. For each case, 5 different streamline samples have been generated either forward from the inlet or backward from the outlet, and with either the natural or homogeneous distribution:

(FNa) Forward, natural distribution (Sample A).

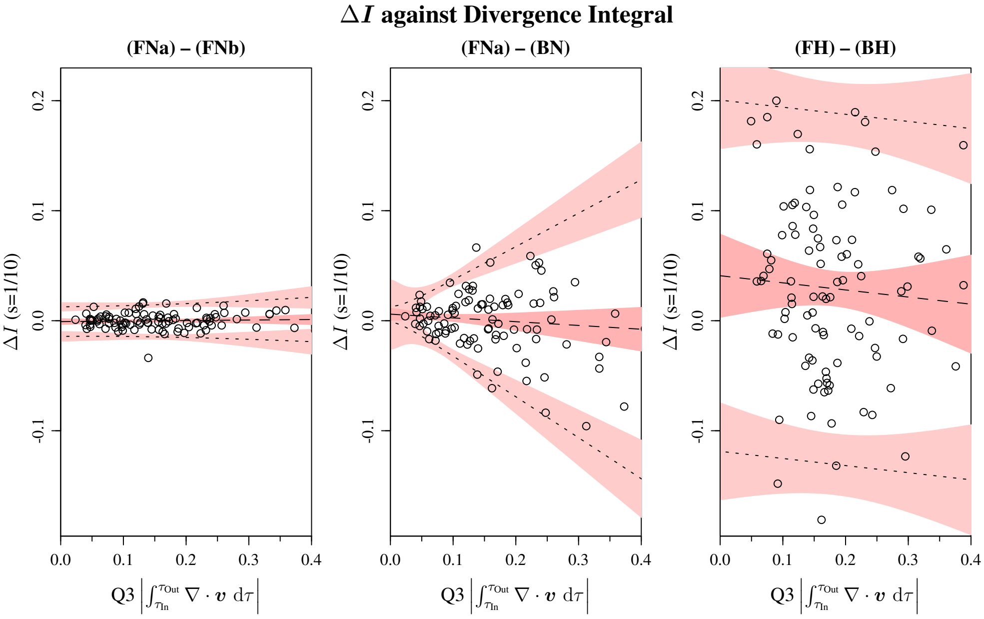

FIGURE 6. Dependance of the inter-sample variability of the MI on the divergence integral. The differences are plotted for two streamline samples generated at the inlet (FNa−FNb), sample at the inlet minus sample at the outlet generated by the natural distribution of streamlines (FNa−BN) and generated by homogeneous distribution (FH−BH).

(FNb) Forward, natural distribution (Sample B).
(BN) Backward, natural distribution.
(FH) Forward, homogeneous distribution.
(BH) Backward, homogeneous distribution.

Here, *homogeneous* denotes the distribution of streamlines obtained from a homogeneous sample of seed points at the inlet (forward) or outlet (backwards). It is used as a naive alternative against which to compare the natural distribution. The corresponding MI has been computed from each of these samples. The flow divergence has been also computed and integrated along each streamline, providing $\xi(\gamma)$.

Figure 6 presents the MI difference, ΔI, between two streamlines samples (inter-sample variability) for 3 different pairs of samples: (FNa)-(FNb), (FNa)-(BN), and (FH)-(BH). The MI difference is plotted against the third quartile of the absolute value of the divergence integrals, $\tilde{\xi} = \mathrm{Q}_3(|\xi(\gamma)|)$. We have considered the linear model $\Delta I = a + b\tilde{\xi} + \varepsilon + \beta\tilde{\xi}$, with two independent normally distributed errors: a constant noise, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon)$, and an effect linearly increasing with the divergence integral, $\beta\tilde{\xi} \sim \mathcal{N}(0, \sigma_\beta\tilde{\xi})$. The fitted line and 95% confidence intervals for the differences, are given by the minimum likelihood estimates of the linear model parameter, a, b, σ_ε, σ_β. The statistical significance of the dependency of the error with the divergence integral ($\sigma_\beta > 0$) has been further confirmed by the studentized Breusch–Pagan test (Koenker 1981).

As expected, the difference, ΔI, for two samples equally generated from the inlet, (FNa)-(FNb), shows no bias and the 95% confidence interval is independent of the divergence integral.

For (FNa)-(BN), ΔI shows a bias and 95% confidence interval significantly dependent on the divergence integral ($p = 6 \times 10^{-6}$). But the limit for null divergence coincides with the differences obtained from (FNa)-(FNb). This is in agreement with the expected invariance of the natural distribution of streamlines with respect to the generating cross section.

In contrast, for (FH)-(BH), the 95% confidence interval is independent of $\tilde{\xi}$ and significantly larger than the confidence interval for (FNa)-(FNb). This reflects the importance of using the natural distribution of streamlines.

6. ICI along steady flows in simple geometries

6.1. *Rotating concentric cylinders with axial flow*

The flow resulting in the annulus between two concentric cylinders, of inner radius R_1 and outer radius R_2, rotating with different angular velocities, ω_1 and ω_2 respectively, is a well studied case with a long history (Childs 2010). In general, the flow can be unstable involving Taylor vortices. For low angular velocity or high viscosity, with Reynolds and Taylor numbers below a critical value, the steady flow is cylindrically symmetric and can be expressed analytically. For the case with static outer cylinder, $\omega_2 = 0$, and a constant pressure gradient in the axial direction, $\partial p/\partial z = -p'$, a steady incompressible Newtonian flow under no-slip condition can be expressed in cylindrical coordinates as

$$v^z = \frac{p' {R_2}^2}{4\mu}\left(1 - \hat{r}^2 - \frac{1-k^2}{\log k}\log\hat{r}\right), \qquad v^\theta = \frac{\omega_1 k^2 R_2}{1-k^2}\frac{1-\hat{r}^2}{\hat{r}}, \qquad v^r = 0,$$

where $k = R_1/R_2$ and $\hat{r} = r/R_2$. This velocity field can be integrated to obtain the streamlines:

$$\theta(t) = \theta_0 + \frac{v^\theta}{r}t, \qquad z(t) = z_0 + v^z t, \qquad r(t) = r_0.$$

Thus, for inlet and outlet defined by the cross sections at $z = z_{\mathrm{In}}$ and $z = z_{\mathrm{In}} + L$, a streamline crossing the inlet at $(r_{\mathrm{In}}, \theta_{\mathrm{In}}, z_{\mathrm{In}})$, will cross the outlet at $(r_{\mathrm{Out}}, \theta_{\mathrm{Out}}, z_{\mathrm{Out}})$, where

$$r_{\mathrm{Out}} = r_{\mathrm{In}}, \qquad \theta_{\mathrm{Out}} = \theta_{\mathrm{In}} + L\frac{v^\theta}{r v^z}, \qquad z_{\mathrm{Out}} = z_{\mathrm{In}} + L. \tag{6.1}$$

The local flux at the inlet or outlet is given by $\varphi_{\boldsymbol{v}}^{\mathcal{S}} = v^z$, which gives the total volumetric flow rate

$$Q = \frac{p' {R_2}^2}{4\mu} A\, C, \qquad \text{where} \quad C = \frac{1}{2}\left(1 + k^2 + \frac{1-k^2}{\log k}\right),$$

and $A = \pi({R_2}^2 - {R_1}^2)$ is the cross-sectional area. After normalization we obtain the natural probability density of streamlines

$$p(\gamma) = \frac{1}{A\, C}\left(1 - \hat{r}^2 - \frac{1-k^2}{\log k}\log\hat{r}\right). \tag{6.2}$$

Finally, expanding v^θ and v^z in (6.1), we get

$$\theta_{\mathrm{Out}} = \theta_{\mathrm{In}} + \lambda\,\frac{\pi C}{\hat{r}^2}\left(1 - \frac{(1-k^2)\log\hat{r}}{(1-\hat{r}^2)\log k}\right)^{-1}, \quad \text{with} \quad \lambda = \frac{\omega_1 R_1^2 L}{Q}. \tag{6.3}$$

Observe that λ is a dimensionless quantity, which can be interpreted as the product of two ratios:

$$\lambda = \frac{Ta}{Re}\,\frac{R_1 L}{A} = \frac{v_1^\theta}{\overline{v}^z}\,\frac{R_1 L}{A},$$

where the Taylor number (or rotational Reynolds number), Ta, and the (axial) Reynolds number, Re, are defined as in Lueptow *et al.* (1992) and Wereley & Lueptow (1999), so that their ratio is the quotient of the velocity of the interior cylinder, $v_1^\theta = \omega_1 R_1$, by the bulk axial velocity, $\overline{v}^z = Q/A$.

We have run experiments for three values of the ratio between internal and external radius, $k = 0.1$, 0.5, and 0.8, and an exponential sequence of values for λ, ranging along several orders of magnitude from 0.01 to 100. For each value of k and λ, a sample of crossing points at the inlet and outlet from 20k streamlines have been generated following (6.2) and (6.3). The corresponding ICI has then been computed from each of these samples.

Figure 7 shows the evolution of ICI when λ increases, for two scales, $s = 1/3$ and $1/20$, and for the three radius ratios. Different combinations of fluid properties or boundary conditions can result in the same value of λ. Two simple interpretations of the curves are that they represent either the ICI evolution when increasing the distance, L, from inlet to outlet, or the ICI change when increasing the rotational velocity, ω_1.

Figure 7 also illustrates the degree of mixing by rendering how the azimuthally coloured inlet is propagated by the streamlines, producing deformed images by their intersection at the outlet. This gives an intuition of the mixing corresponding to each value of the ICI.

The ICI increases faster and further for coarser scales. In particular, $ICI_{1/3}$ converges very close to 1 for large λ, while $ICI_{1/20}$ is always below 0.9. At first glance, we could expect that ICI_s converges to 1 for any scale. However, this laminar rotating flow produces no mix in the radial direction. Thus, there is always some amount of information preserved. The difference between both scales is larger for the smallest radius ratio, $k = 0.1$, while for $k = 0.8$ the ICI of both scales approach each other. The value of ICI_s also increases with k, since then the contribution of the radial direction is reduced with respect to the angular one.

6.2. *Double-bend tubes*

One of the simplest 3-dimensional geometries is a tube including the sequence of two non-coplanar 90° bends. Lee *et al.* (2008) investigated the characteristics of the steady flow obtained under 5 different inter-bend angles ($\alpha = 0°, 45°, 90°, 135°, 180°$) and 2 Reynolds numbers ($Re = 125, 500$), with tube radius $R = 2\,\mathrm{mm}$ and bend radius $R_{\mathrm{bend}} = 8\,\mathrm{mm}$. We have reproduced the same geometries (figure 8) and flow conditions considered there, investigating the ICI obtained in each case.

For each angle, a series of cross sections have been generated every 1 mm along the tube centreline, numbered according to their position relative to the inter-bend centre (see figure 8). The ICI has been computed from the inlet cross section at −30 mm, to each of the outlets in the cross-section series, from position −20 mm to 40 mm. Figure 9 shows the ICI obtained along this sequence of outlets, for the 5 angles and 2 Reynolds numbers, and for 2 scales: $s = 1/3$ and $1/20$.

The inlet has been coloured radially from red in the centre to blue in the perimeter, adding 6 white radius. This image was propagated by the flow streamlines, producing deformed images by their intersection with each cross section. These images illustrates the degree of mixing and complexity of the flow from the inlet to each of the series of outlets. They are presented in figure 9 at positions -5, 0, 12 and 40 mm, for each angle and Reynolds, together with the corresponding ICI.

When comparing the 5 angles, both the coloured cross sections and the ICI coincide for outlets along the first bend and the initial portion of the second bend, where the dynamics is essentially characterized by the Dean vortices developed by the first bend. The divergence between angles starts earlier for $Re = 500$, as seen in both scales. This

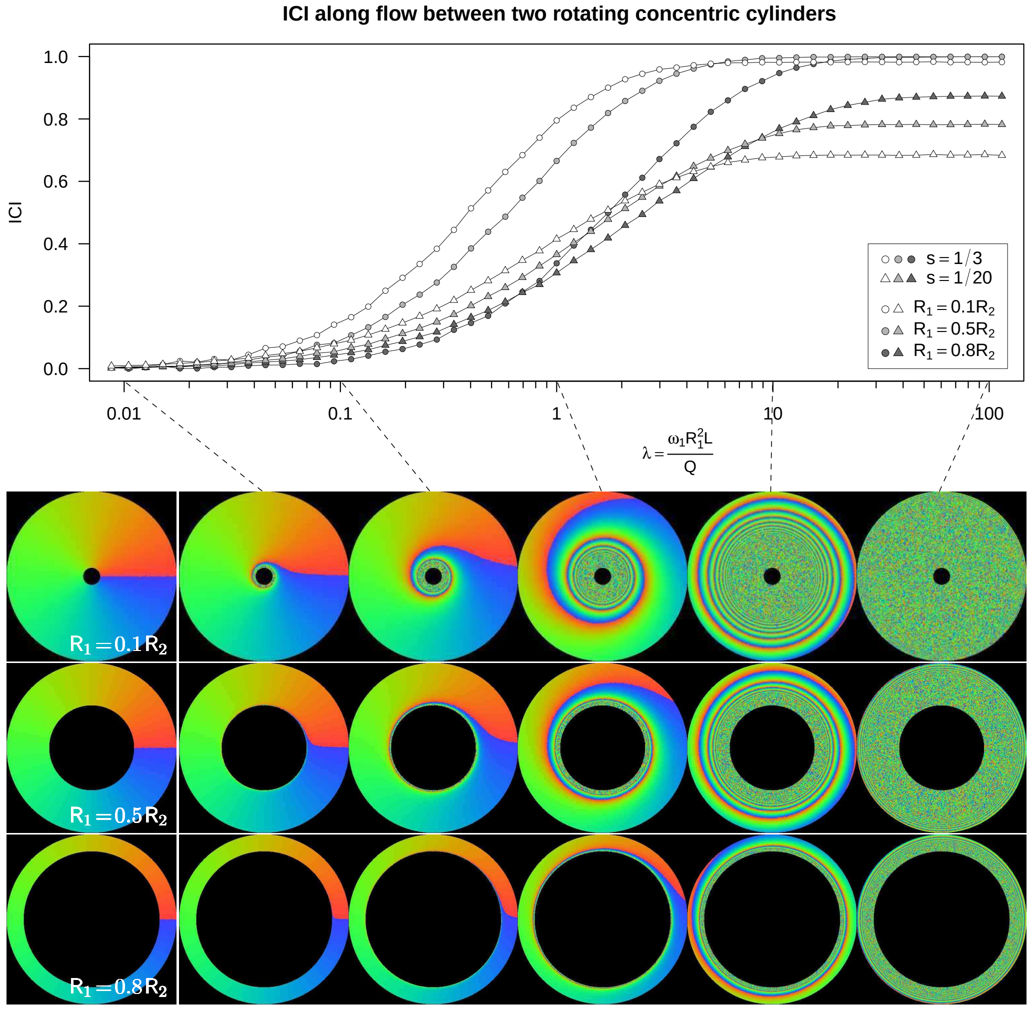

FIGURE 7. ICI and illustration of the mixing produced by a steady flow in annulus with rotating inner cylinder and axial pressure gradient, for 3 radius ratios. The ICI at two scales (s) is plotted as function of the dimensionless parameter λ. For each radius ratio, the azimuthally coloured inlet (left column) is propagated by the streamlines, producing deformed images by their intersection at the outlet for $\lambda = 0.01$, 0.1, 1, 10, and 100.

can be explained since the Dean vortices generated by the second bend are stronger and faster developed for higher Reynolds number.

The ICI for the planar bend with $\alpha = 180°$ is almost the same for both Reynolds numbers at the end of the second bend (12 mm). This can be understood since the higher energy in the Dean vortices for $Re = 500$ is compensated by the higher axial velocity, so that the streamlines are similar. Lee *et al.* (2008) mention that at $Re = 500$ the strength of the primary Dean vortices is sufficient to lead to secondary and even tertiary vortices. They do not explicitly state this not happening at $Re = 125$, but any possible differences between both does not result in a clear difference in ICI. However, the ICI still grows significantly after the second bend for $Re = 500$, indicating that the vortices are more persistent. This effect is common also for all the angles, so that the ICI

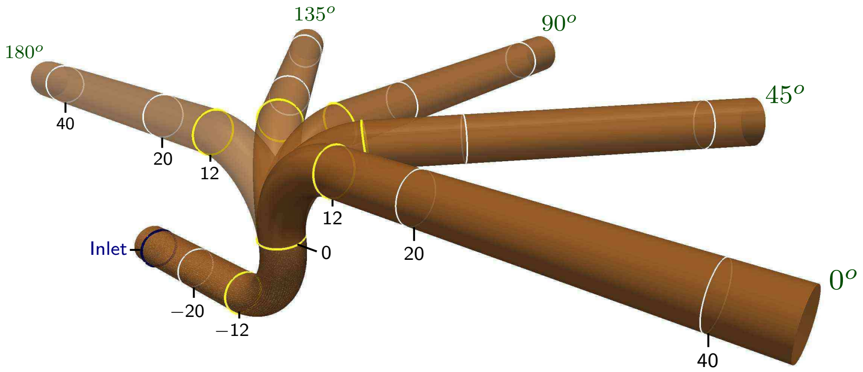

FIGURE 8. Double-bend tubes with inter-bend angles 0°, 45°, 90°, 135°, and 180°. For each angle, a series of cross sections have been generated along the tube centreline, each separated 1 mm from the previous one. They are numbered according to their position relative to the inter-bend centre. The inter-bend centre (0), the first ring at the beginning of the first bend (−12) and the last ring at the end of the second bend (12) are shown in yellow.

at 40 mm is always larger for $Re = 500$, and not even completely developed, in contrast with $Re = 125$.

As expected, the ICI at the outlet is higher for non-planar than for planar models. Both $\alpha = 45°$ and $\alpha = 90°$ present very similar coloured cross sections and ICI. They share the highest ICI in both scales and both Reynolds numbers. This is in agreement with the observation in Lee *et al.* (2008) that there appears to be an optimal inter-bend angle between 45° and 90° to achieve asymmetry in the Dean vortex pattern.

In contrast with $\alpha = 180°$, for $\alpha = 0°$ the ICI is substantially higher at $Re = 500$ than at $Re = 125$. Apparently the reverse Dean vortices developed in the second bend, partially undo the mixing for low Reynolds number. This even produces a reduction of ICI for $Re = 125$ at the end of the second bend. This effect can be seen in the coloured cross sections, with the reduction of the deformation in the centre at 12 mm.

For $Re = 125$, we can also observe a small reduction of the ICI at the end of the first bend, only for scale $s = 1/3$, but shared by all the angles. We have not found a satisfactory explanation of this effect.

7. Unsteady incompressible flow on complex geometries

7.1. *Cerebral arteries with aneurysms*

As a case study on non-idealized geometries, we have investigated the ICI obtained for unsteady flows in anatomically realistic cerebral arteries with aneurysms. Aneurysms are pathologic dilations of the vessel wall, and their flow characteristics have been extensively investigated as mechanically and biologically related with the initiation, growth, and eventual rupture of aneurysms (Cebral *et al.* 2011; Geers *et al.* 2011; Schnell *et al.* 2014; Gopalakrishnan *et al.* 2014; Arzani *et al.* 2016).

A set of 49 geometries obtained from segmentation of 3D Rotational Angiography (3DRA) images including an aneurysm in the middle cerebral artery have been selected from the @neurIST database (Aneurist Consortium 2010; Villa-Uriol *et al.* 2011). Surface models of patient-specific vascular geometries were obtained with the geodesic active region segmentation method (Bogunović *et al.* 2011) and manually corrected for topological errors using the software suite @neuFuse (Villa-Uriol *et al.* 2011). The vasculature of interest includes at least 12 vessel diameters upstream and 4 vessel diameters downstream

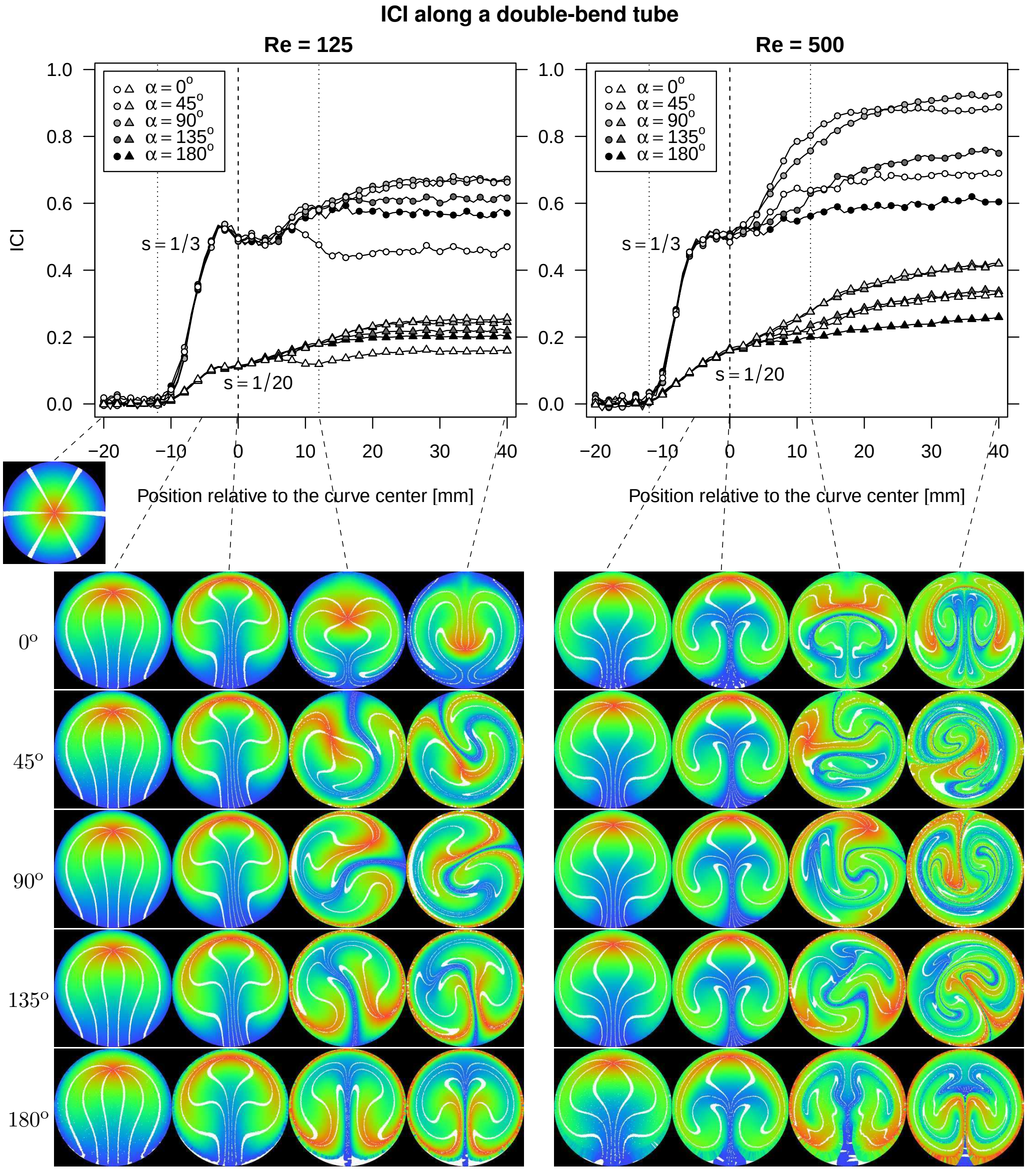

FIGURE 9. ICI and illustration of the mixing along double-bend tubes with 5 different inter-bend angles, and for two Reynolds numbers. The corresponding ICI_s is computed for two scales, s=1/3 (circles) and s=1/20 (triangles), from the inlet to each of the outlets in the cross-section series (see figure 8). To visualise the degree of mixing, the radially coloured inlet is propagated by the streamlines, producing deformed images by their intersection with each cross section. They are presented at positions −5, 0, 12 and 40 mm, for each angle and Reynolds number.

from the aneurysm. Inlets and outlets have been clipped with planes perpendicular to the vessel centerline. Unsteady flow simulations have been performed imposing flow rate (inlet) and pressure waveforms (outlets) extracted from a one-dimensional model of all the large arteries in the human body (Reymond *et al.* 2009), for a cardiac cycle of period 0.8 s.

For the resulting unsteady incompressible flow, both versions of ICI could be applied:

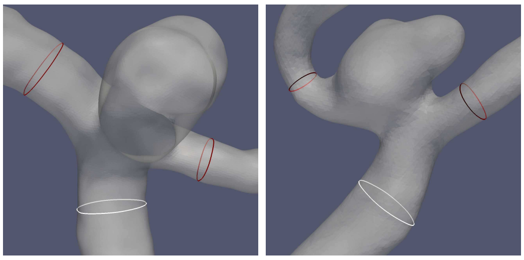

FIGURE 10. Example of the cross sections automatically selected for the inlet vessel and each of the two outlet vessels, at one diameter from the aneurysm neck at the bifurcation.

considering either streamlines or world-lines. We have considered here the streamlines for two physiologically relevant cardiac phases: peak systole (PS) and end diastole (ED). This allows us to directly compare the results to some existing subjective classifications of aneurysm flow complexity and stability, defined also for these two cardiac phases.

The considered control volume, $\mathcal{D}_{\mathcal{S}}$, comprises the aneurysm and a portion of each adjacent vessel. We have selected inlet and outlet at one vessel diameter from the aneurysm neck, as illustrated in figure 10. This is the same criterion introduced by Millan *et al.* (2007) for aneurysm morphological characterization. The limiting cross sections were automatically computed based on the aneurysm removal methodology proposed by Ford *et al.* (2009) and implemented in the Vascular Modeling Toolkit (Piccinelli *et al.* 2009). Using vessel centerlines and Voronoi diagrams, the aneurysm neck and vessel bifurcations were detected and used as a reference to reliably identify the cross sections.

For this control volume, the outlet consists of multiple, separate regions. But this does not affect the definition of ICI.

7.2. *ICI as a function of scale*

For each case, ICI_s was obtained for the sequence of scales, $s = \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{20}$. This provides ICI as a function of the scale. Figure 11, displays this function for several representative examples.

In general, ICI_s increases with the scale s. The obtained continuous and smooth behaviour makes evident a high correlation between the values at different scales. Thus, a dimensionality reduction can be obtained, for instance, by Principal Component Analysis (PCA). Effectively, the first principal component (PC) explains 99.1% of the variance, and the 2 first PCs explain 99.9% of the variance. These two PCs are also plotted in figure 11.

An illustration of the mixing corresponding to each of the ICI values is also presented by coloured cross sections, similar to the ones in previous cases, but adapted to the multiplicity of the outlet. Each of the outlets have been coloured by two different patterns: radially from red to blue, and by homogenous colour identifying each outlet vessel. Two coloured cross section are respectively obtained at the inlet, by their propagation backwards along the streamlines.

7.3. *Comparison with subjective assessment of flow complexity*

For each aneurysm, the flow field has been qualitatively assessed by an expert, through visual inspection of the streamlines, according to four qualitative variables (Cebral *et al.* 2011; Geers *et al.* 2011). Three variables are measured at PS: flow complexity (*simple* or *complex*), inflow concentration (*concentrated* or *diffuse*), and impingement size (*small* or *large*). The fourth variable, flow stability (*stable* or *unstable*), is measured by comparing the flow patterns at PS and ED.

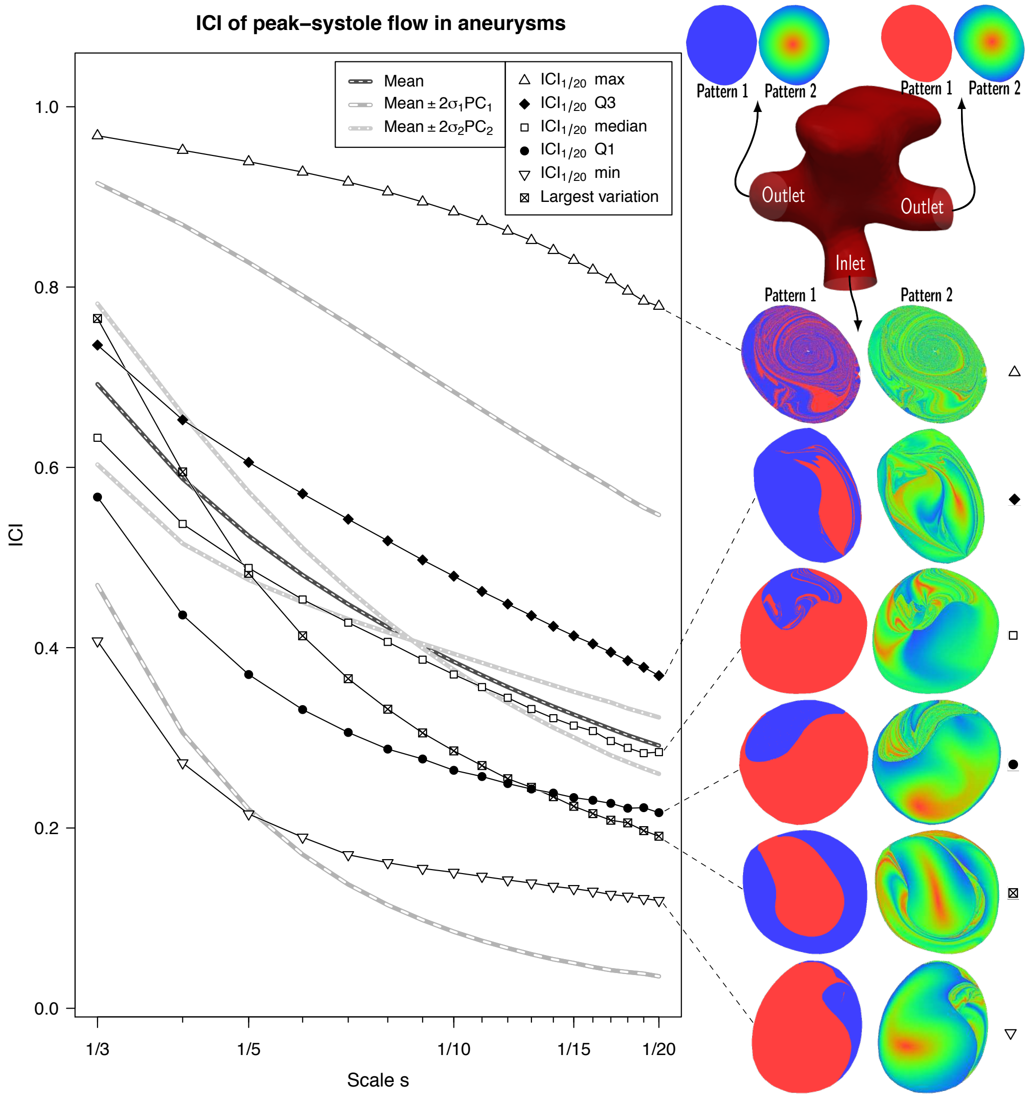

FIGURE 11. ICI as function of scale and illustration of mixing in typical cases from the aneurysm dataset. The line-connected symbols are the ICI at PS of examples representing the minimum, maximum, median, and 1st and 3rd quartiles, for $s = 1/20$. The case with largest change in the order from $s = 1/3$ to $s = 1/20$ pass from the 1st to the 4th quartile. The ICI mean, and the variation explained by the first 2 principal components are also plotted. Each of the outlets have been coloured by two different patterns: by homogenous colour identifying each outlet vessel as blue or red, and radially with a rainbow colour map. Their propagation backwards by the streamlines, produce respectively two coloured cross sections at the inlet.

We have investigated the correlation of the proposed quantitative ICI at PS with the subjective qualitative classification. For each qualitative variable, the corresponding two categories have been plotted as a box-plot of $ICI_{1/10}$ (figure 12). Non-parametric Mann–Whitney U test (Mann & Whitney 1947; Hollander *et al.* 2013) has been applied to each pair of categories. Three of them gave statistically significant differences (flow complexity: $p = 9 \times 10^{-3}$, flow stability: $p = 2 \times 10^{-4}$, impingement size: $p = 4 \times 10^{-4}$). Only inflow concentration did not show significant differences. Surprisingly, flow stability showed

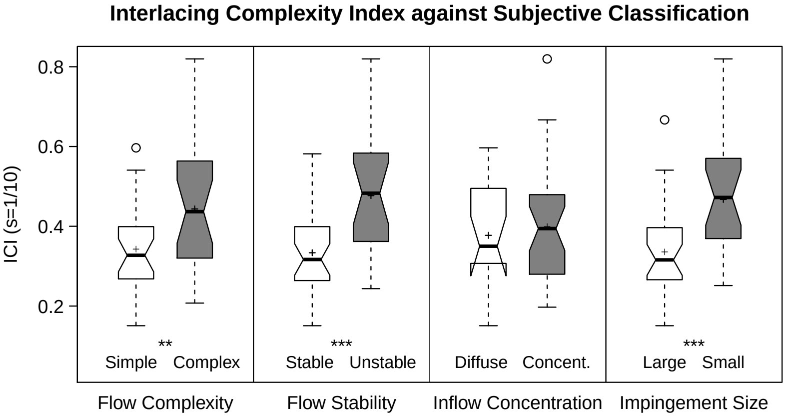

FIGURE 12. Correlation of ICI with each of the 4 subjective flow classifications. The significant difference between each pair of categories is assessed by non-parametric Mann–Whitney U test. $**$ $(0.01 \geqslant p > 0.001)$, $***$ $(0.001 \geqslant p > 0.0001)$.

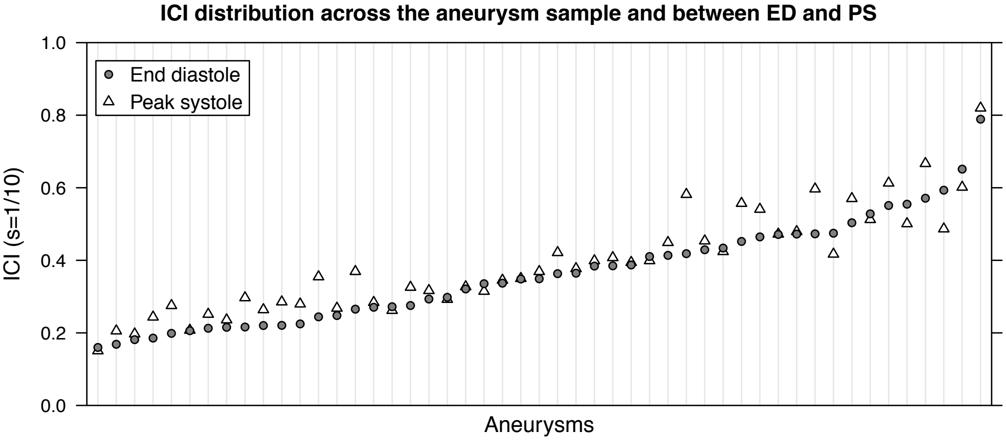

FIGURE 13. Comparison between ICI of the flow at peak systole (PS) and end diastole (ED) for each aneurysm, at scale $s = 1/10$. The ICI is generally larger at PS than at ED, but the opposite behaviour is also observed. The aneurysms have been ordered according to their ICI at ED.

significant differences for the ICI exclusively measured at PS, although it is supposed to account for the variation between ED and PS. Probably, this reflects the fact that more complex flows are also more unstable.

7.4. *Comparison between peak systole and end diastole*

Figure 14 presents a dot-plot with $ICI_{1/10}$ obtained for each aneurysm at PS and ED. The values spread quite homogeneously across the ICI range. In general, the ICI obtained for PS is larger than the one obtained for ED, but the opposite behaviour is also observed.

The ICI variation from ED to PS is presented in figure 14, separated according to the qualitative classification into stable and unstable. A clear correlation appears, which is confirmed by one-tailed Mann–Whitney U test, giving a high significant difference ($p = 5 \times 10^{-4}$). For completeness, Mann–Whitney test was also applied to the ICI variation

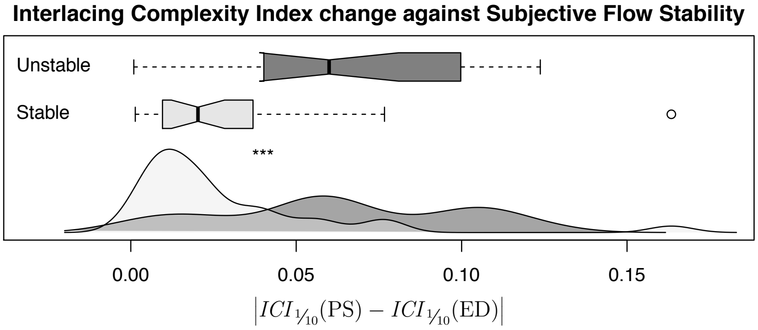

FIGURE 14. Subjective flow stability classes against absolute value of the difference of ICI between peak systole and end diastole. A strong significant difference ($p = 0.0005$, one-tailed Mann–Whitney U-test) is found between subjectively stable and unstable flows.

for the classes obtained from the qualitative flow complexity, inflow concentration, and impingement size. None of them provided significant differences.

8. Unsteady compressible 2D periodic cellular flows

8.1. *2D periodic cellular flows*

Periodic cellular flows are simply defined models that can, however, show very complex evolutions. We have considered a modification of the 2D unsteady model used by Zu *et al.* (2015):

$$\boldsymbol{v} = \frac{1}{\rho}\left[\big(\cos(2\pi y), \cos(2\pi x)\big) + \theta\cos(t)\big(\sin(2\pi y), \sin(2\pi x)\big)\right]$$

defined in Cartesian coordinates $\boldsymbol{x} = (x, y)$ for the square domain $\Omega = [0,1] \times [0,1]$ with periodic boundary conditions (torus). The *base* flow ($\rho = 1, \theta = 0$) produce two counter-rotating cells centred at $(\frac{1}{4}, \frac{3}{4})$ and $(\frac{3}{4}, \frac{1}{4})$, and two saddles at $(\frac{1}{4}, \frac{1}{4})$ and $(\frac{3}{4}, \frac{3}{4})$. The parameter θ introduce a time-dependent perturbation, making it an unsteady flow. The original model did not include the density ρ. We have considered a density

$$\rho = 1 + \kappa\sin(2\pi x)\sin(2\pi y)$$

controlled by the parameter κ, which introduces an inhomogeneity, making the flow compressible. This density is maximum at the saddles and minimum at the rotating centres. The resulting unsteady compressible flow satisfies the continuity equation (2.11). Thus, the natural distribution of world-lines is invariant, and we can consider the corresponding version of ICI.

For this flow, we consider the spatio-temporal control volume $\mathcal{D}_{\mathcal{S}} = \Omega \times \mathcal{T}$ given by the full spatial cyclic domain Ω (torus) along a lapse of time $\mathcal{T} = [0, T]$. Thus, the inlet and outlet are not spatial surfaces, but the full space, Ω, at instances $t = 0$ and $t = T$, respectively.

Figure 15 illustrates the resulting flow for different values of θ and κ, and for different time lapses. The ICI for each of these parameter values and time lapses is displayed for a series of scales from $s = 0.01$ to $s = 0.5$. The flow propagation of a vertical colour pattern is also shown for each of the parameter values.

The dominant factor is the increase of ICI with the time lapse. The increase of ICI with the parameter θ is pronounced for all scales and time lapses, reflecting that the higher the unsteadiness, the more complex the flow. The ICI for compressible flows ($\kappa = 0.9$) is

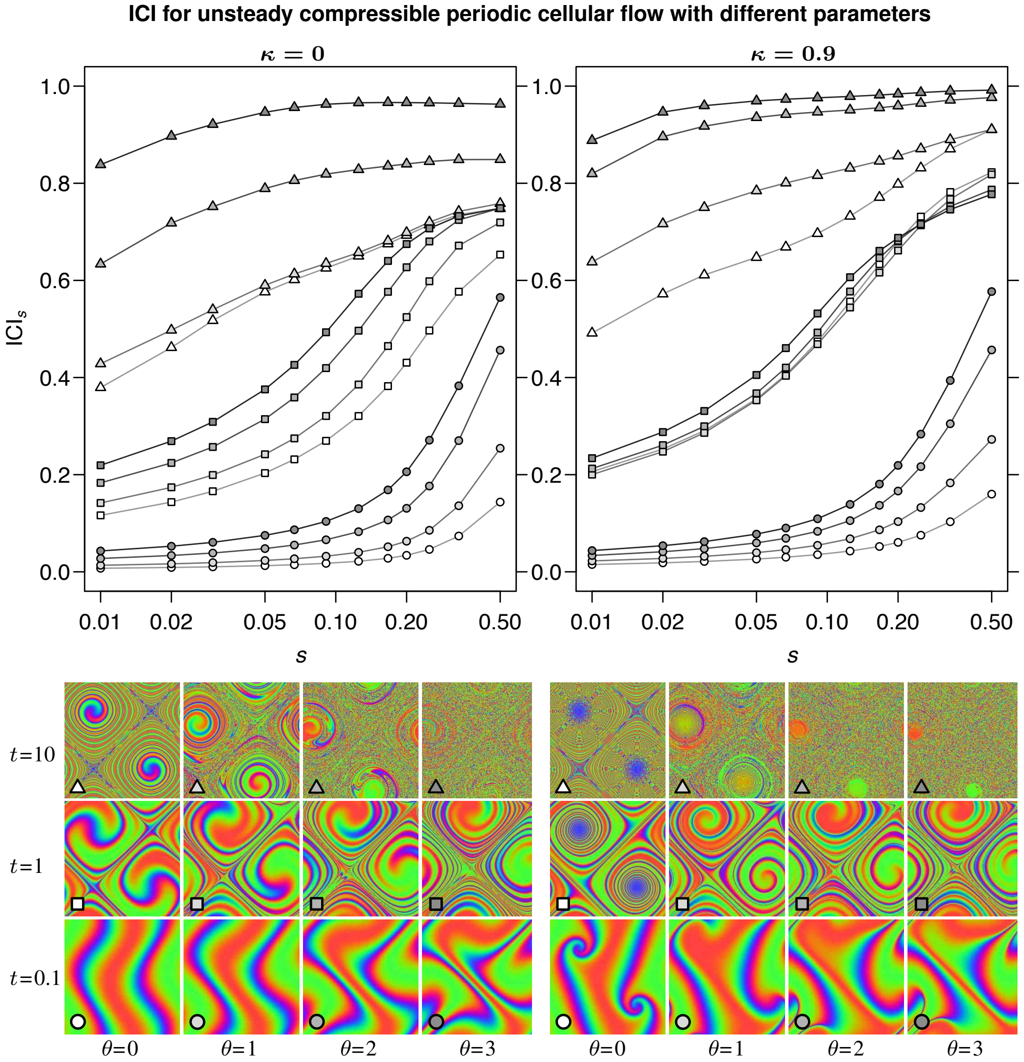

Figure 15. ICI as a function of scale for the 2D periodic cellular flow with $\kappa = 0$ (incompressible) and $\kappa = 0.9$ (compressible), with 4 different values of θ (unsteadiness), and for 3 time lapses. The legend is presented at the bottom, together with the illustration of the propagation by each flow of a vertical colour pattern.

also higher than for incompressible flows ($\kappa = 0$), although the difference is less marked. It is more evident for the steady flow ($\theta = 0$) and intermediate time ($t = 1$), for which the difference between steady and unsteady flow is reduced. This is in agreement with the intuitive impression from the propagated colour pattern.

The curves of ICI_s increase with the scale. But each of the cases presents a clear maximum slope at different scales, where the complexity increases faster. The plots evidence that the complexity is progressively propagated to smaller scales with time, in agreement with the colour patterns. We can also see that the higher the unsteadiness (θ), the smaller the scale showing fast ICI increase.

8.2. *Test of the limit of ICI for infinitesimal scales*

In appendix B, we demonstrate (B 1) that $\lim_{s\to 0} ICI_s = 0$, and that the dominant term for infinitesimal scales is proportional to $-1/\log(s)$.

For the proposed cellular flow, the probability densities at inlet at outlet coincide and are given be the density,

$$p_{\mathrm{In}}(\boldsymbol{x}) = \rho(\boldsymbol{x}) \qquad \text{and} \qquad p_{\mathrm{In}}(\boldsymbol{y}) = \rho(\boldsymbol{y}).$$

Thus, $\Delta = 0$, and (B 1) becomes

$$ICI_s = \frac{-1}{\log(s)}\Psi + \mathcal{O}\left(\left(\frac{1}{\log(s)}\right)^2\right), \quad \text{with} \quad \Psi = \frac{1}{4}\int_\Omega \log\left(\tfrac{1}{4}|\boldsymbol{Q_x}|\right)\rho(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{8.1}$$

Observe that the term $\log\left(\frac{1}{4}|\boldsymbol{Q_x}|\right)$ shares some elements with the finite-time Lyapunov exponents (Tang & Boozer 1996), since it can be expressed in terms of the eigenvalues of the Cauchy–Green deformation tensor $\boldsymbol{J}_{\boldsymbol{x}}^{\mathsf{T}}\boldsymbol{J}_{\boldsymbol{x}}$. Actually, the higher the Lyapunov exponents, the higher will be Ψ. However, Lyapunov exponents are *intensive* quantities (average flow characteristics per unit of time), whereas ICI is *extensive* (total flow characteristics in a finite spatio-temporal domain). Thus, in contrast with Lyapunov exponents, Ψ is not normalized by the time lapse, so that it is expected to unboundedly grow with time.

We have numerically computed the factor Ψ for flows with different values of $\theta = 0, 1, 2, 3$ and $\kappa = 0, 0.5, 0.9$, and for different time-lapses, $t = 0.1, 1, 10$. To compute the integral, the factor $\log\left(\frac{1}{4}|\boldsymbol{Q_x}|\right)$ has been computed in a regular grid of 100×100 points. For this, the Jacobian matrix, $\boldsymbol{J_x}$, has been computed from the flow equations applying the method described in (Sandri 1996), using 4th-order Runge–Kutta algorithm with time-step 10^{-4}.

Figure 16 compares the behaviour of ICI_s for small scales to the deduced behaviour dependent on Ψ. The left plot displays ICI_s against $-1/\log(s)$ together with the expected limiting straight line of slope Ψ. The right plot complements it by displaying $-\log(s) \times ICI_s$, whose limit for $s \to 0$ is directly Ψ, being more easily observable. In all cases, the expected limit seems to be satisfied. This is more evident for $t = 0.1$ and $t = 1$, for which a linear behaviour is reached for the shown scales, $s \geqslant 0.01$. For $t = 10$, this is not still reached, but the limit, Ψ, seems compatible with a smooth propagation of $-\log(s) \times ICI_s$. Observe that they show ($t = 1$) or suggest ($t = 0.1$ and $t = 10$) a maximum for some scale. The larger t and θ, the smaller this scale.

9. Conclusions

In this paper, we have introduced the interlacing complexity index (ICI) as a measure of the flow complexity for open systems with inlet and outlet regions of arbitrary shape and topology. No previous flow complexity measure is applicable to these systems. ICI is defined using Shannon's mutual information, inspired by an analogy between open flow systems and information systems from information theory, where inlet, outlet, and flow transport between them correspond to emitter, receiver, and communication channel, respectively. The more complex the flow, the larger the information loss in the propagation from inlet to outlet.

The behaviour of ICI has been tested with numerical experiment on several flows cases, involving geometries and conditions of different complexity, steady and unsteady flows, and inlets and outlets of different topologies. The results indicate that the ICI provides a sensitive complexity measure with intuitive interpretation in a diversity of conditions

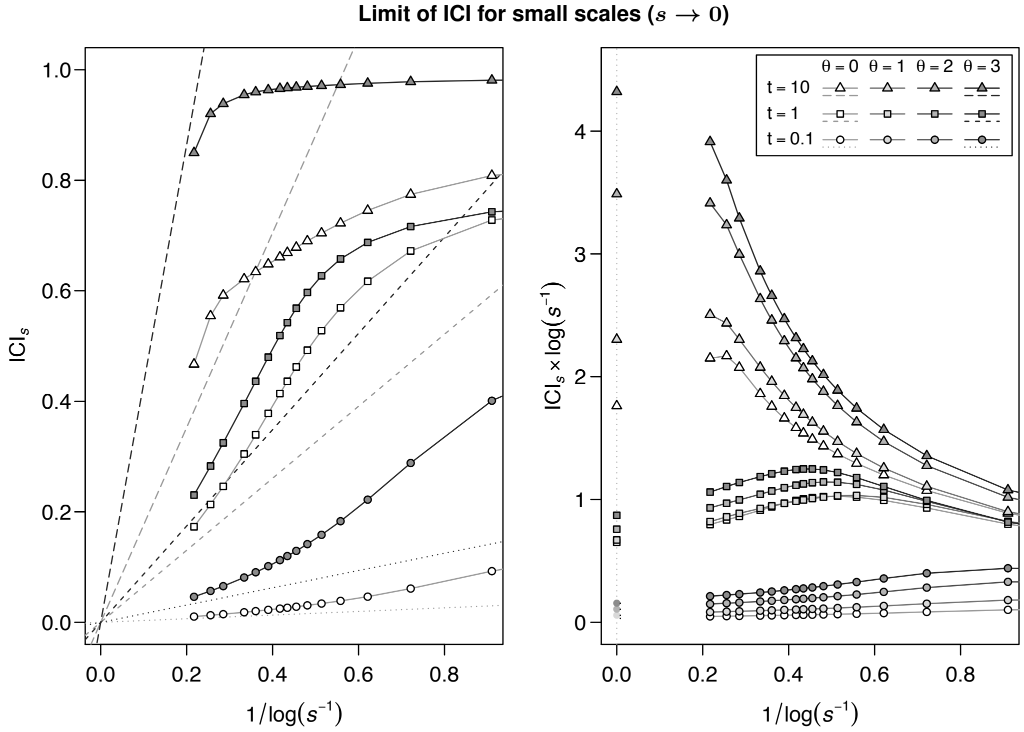

FIGURE 16. ICI_s for small scales in comparison to the expected leading-order term $\Psi/\log(s^{-1})$. They are shown for the periodic cellular flow with $\kappa = 0.5$, for 4 values of θ and 3 time lapses t. Left: ICI_s against $1/\log(s^{-1})$ (connected symbols) and limit slope Ψ (dashed lines). Right: $ICI_s \times \log(s^{-1})$ against $1/\log(s^{-1})$ (connected symbols) and limit $\lim_{s\to 0} ICI_s \times \log(s^{-1}) = \Psi$ (symbols at $s = 0$).

and in agreement with the observation of flow features and subjective classifications. The experiments have also shown a good accuracy and precision of the algorithm proposed for the ICI estimation from finite samples of streamlines.

For the flow cases in rotating concentric cylinders, the ICI demonstrates a good sensitivity on a broad range of parameters spanning over 4 orders of magnitude, showing also differential behaviour depending on the scale of observation. The ICI values obtained for the flow in double-bend tubes with different angles and Reynolds numbers are consistent with previous observations in the literature on the presence and strength of Dean vortices. For the unsteady pulsatile flow cases in vasculatures with aneurysms, we have compared the ICI with qualitative subjective visual assessments of flow complexity and stability, which have been previously used as haemodynamic indicators of aneurysm rupture risk. The high correlations obtained indicate that ICI captures a great part of these subjective classifications, but providing an objective continuous index. In the case of unsteady compressible periodic cellular flows, we have explored the complexity introduced by the increase in the unsteadiness and compressibility. We have also tested in this case the analytically derived limit of ICI for infinitesimal scales.

The experiments have been run on flows obtained analytically or by numerical simulations. We have only dealt with limited Reynolds numbers, involving non-turbulent flows, although including Dean vortices and recirculations. Although ICI is defined for any flow, for highly turbulent flows the numerical simulation and the streamline integration could be more challenging, representing a possible limitation for its applicability.

The focus of this paper was on showing the relevance, intuitive behaviour, and capacities of the proposed ICI, not on our capacity to simulate turbulent flows. For experimental flow measurements, the resolution of the observation will also limit the scale at which the flow and the ICI can be computed.

We have presented a careful definition of all the elements involved in the ICI formulation. A crucial component is the natural distribution of streamlines for steady flows) or of world-lines (for unsteady flows). Given any steady or mass-preserving flow, this probability distribution is invariant with respect to the cross section used to generate or parameterize the streamlines. For unsteady incompressible flows, the natural distributions of streamlines and of world-lines are different, but both are invariant. Thus, it is possible to consider either of them according to the intended application. We believe that this is a relevant contribution in itself, which should be considered for any other statistics based on streamlines to describe the flow. It could also be applied for flow visualization by representative sets of streamlines.

Acknowledgements

The dataset of vasculatures with aneurysms used for this work has been provided by the European integrated project @neurIST (IST-027703) (see http://www.aneurist.org).

Appendix A. Analytical computation for a cylindrical straight tube

A steady flow in a cylindrical straight tube under no-slip condition is described analytically by straight streamlines parallel to the cylinder axis and with a parabolic velocity profile. In this case, the probability densities can be computed analytically and the corresponding MI can be obtained by numerical integration. We present this computation, which is used to assess the correctness and accuracy of the numerical algorithm employed in the experiments.

The parabolic profile gives the natural distribution of streamlines

$$p(\gamma) = \frac{2}{\pi R^4}(R^2 - \|\boldsymbol{x}(\gamma)\|^2),$$

where R is the vessel radius and $\boldsymbol{r} := \boldsymbol{x}(\gamma)$ is the position of the streamline γ at any cross section relative to the cross section centre. For scale s, the corresponding truncated Gaussian kernel of standard deviation $\sigma = sR$ is

$$p_{\mathrm{In},s}(\boldsymbol{x} \mid \gamma) = \begin{cases} \frac{1}{C} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{r}\|^2}{2s^2R^2}\right), & \|\boldsymbol{x}\| < R \\ 0, & \|\boldsymbol{x}\| \geqslant R \end{cases}$$

and analogous for $p_{\mathrm{Out},s}(\boldsymbol{y} \mid \gamma)$. Due to the truncation at the circular boundary, the normalization C depends on $\|\boldsymbol{r}\|$, and is given by the cumulative non-central χ^2 distribution function with 2 degrees of freedom, equivalent to the Marcum Q-function (Nuttall 1975):

$$C = \int_{\|\boldsymbol{x}\|<R} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{r}\|^2}{2s^2R^2}\right) \mathrm{d}S_{\mathrm{In}} = 2\pi s^2 R^2 \left(1 - \mathrm{Q}_1\left(\frac{\|\boldsymbol{r}\|}{sR}, \frac{1}{s}\right)\right).$$

The scale dependent joint probability density between the positions at any two sections (inlet and outlet) is then

$$p_s(\boldsymbol{x}, \boldsymbol{y}) = \int_{\|\boldsymbol{r}\|<R} \frac{1}{C^2} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{r}\|^2 + \|\boldsymbol{y}-\boldsymbol{r}\|^2}{2s^2R^2}\right) \frac{2}{\pi R^4}(R^2 - \|\boldsymbol{r}\|^2)\,\mathrm{d}S,$$

Normalizing the positions by the radius, $\hat{\boldsymbol{x}} = \boldsymbol{x}/R$, $\hat{\boldsymbol{y}} = \boldsymbol{y}/R$, and $\hat{\boldsymbol{r}} = \boldsymbol{r}/R$, the joint probability density can be reparameterized as

$$p_s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \frac{1}{2\pi^3 s^4} \int_{\|\hat{\boldsymbol{r}}\|<1} \frac{1 - \|\hat{\boldsymbol{r}}\|^2}{\left(1 - \mathrm{Q}_1(s^{-1}\|\hat{\boldsymbol{r}}\|, s^{-1})\right)^2} \exp\left(-\frac{\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{r}}\|^2 + \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{r}}\|^2}{2s^2}\right) \mathrm{d}\hat{S}.$$

Since MI is invariant to any independent reparameterization of the two variables, we can use this normalized version, which implies that $I_s(\boldsymbol{X}, \boldsymbol{Y})$ will be independent of the vessel radius R. The corresponding marginal probability densities are

$$p_s(\hat{\boldsymbol{x}}) = \frac{1}{\pi^2 s^2} \int_{\|\hat{\boldsymbol{r}}\|<1} \frac{1 - \|\hat{\boldsymbol{r}}\|^2}{1 - \mathrm{Q}_1(s^{-1}\|\hat{\boldsymbol{r}}\|, s^{-1})} \exp\left(-\frac{\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{r}}\|^2}{2s^2}\right) \mathrm{d}\hat{S},$$

and analogously, $p_s(\hat{\boldsymbol{y}})$ is given by the same function. Re-expressing the integral in polar coordinates we obtain

$$p_s(\hat{\boldsymbol{x}}) = \frac{2}{\pi s^2} \int_0^1 \frac{\rho(1 - \rho^2)}{1 - \mathrm{Q}_1(s^{-1}\rho, s^{-1})} \exp\left(-\frac{\|\hat{\boldsymbol{x}}\|^2 + \rho^2}{2s^2}\right) \mathrm{I}_0\left(\frac{\rho\|\hat{\boldsymbol{x}}\|}{s^2}\right) \mathrm{d}\rho = \mathrm{F}_{s,1}(\|\hat{\boldsymbol{x}}\|), \tag{A 1}$$

where I_0 denotes the modified Bessel function of the first kind. The joint probability density can also be re-expressed in polar coordinates, with some refactorization of the exponential arguments, giving

$$p_s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \frac{1}{4\pi s^2} \exp\left(-\frac{\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\|^2}{4s^2}\right) \mathrm{F}_{s,2}\left(\frac{\|\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}\|}{2}\right)$$

with

$$\mathrm{F}_{s,k}(a) = \frac{2k}{\pi s^2} \int_0^1 \rho(1 - \rho^2) \left(\frac{\exp\left(-\frac{1}{2}s^{-2}(a^2 + \rho^2)\right)}{1 - \mathrm{Q}_1(s^{-1}\rho, s^{-1})}\right)^k \mathrm{I}_0\left(\frac{k\rho a}{s^2}\right) \mathrm{d}\rho \tag{A 2}$$

A closed formula in terms of known functions have not been found for the integral over ρ. However, $F_{s,k}$ has been computed for $s^{-1} = 1, \ldots, 20$ and $k = 1, 2$, using the statistical software R (R Core Team 2013), by numerical integration based on the package QUADPACK (Piessens 1983), giving a smooth, monotonic decreasing and log-concave function on $a \in [0, 1] \subset \mathbb{R}$.

The MI can be computed from the joint and marginal differential entropies as

$$I_s(\boldsymbol{X}, \boldsymbol{Y}) = h[p_s(\hat{\boldsymbol{x}})] + h[p_s(\hat{\boldsymbol{y}})] - h[p_s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})] = 2h[p_s(\hat{\boldsymbol{x}})] - h[p_s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})], \tag{A 3}$$

where we have used that both marginal probabilities follow the same distribution. The marginal differential entropy can be simplified using polar coordinates:

$$h[p_s(\hat{\boldsymbol{x}})] = -\int_{\|\hat{\boldsymbol{x}}\|<1} p_s(\hat{\boldsymbol{x}}) \log\left(p_s(\hat{\boldsymbol{x}})\right) \mathrm{d}\hat{S}_{\mathrm{In}} = -2\pi \int_0^1 a\, \mathrm{F}_{s,1}(a) \log\left(\mathrm{F}_{s,1}(a)\right) \mathrm{d}a \tag{A 4}$$

For computing the joint differential entropy, we first use the change of variables $\boldsymbol{a} = (\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}})/2$ and $\boldsymbol{b} = (\hat{\boldsymbol{y}} - \hat{\boldsymbol{x}})/2$, and express them in polar coordinates, taking into account the conditions $\|\hat{\boldsymbol{x}}\| < 1$ and $\|\hat{\boldsymbol{y}}\| < 1$ for the integrals. Thus, the angular variables can be analytically integrated, resulting in the double integral

$$\begin{aligned} h[p_s(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})] = -\frac{8}{s^2} \int_0^1 \mathrm{d}a \int_0^{\sqrt{1-a^2}} \mathrm{d}b\;\, ab \left[\tfrac{\pi}{2} - \arccos\Big(\min\big(\tfrac{1-a^2-b^2}{2ab}, 1\big)\Big)\right] \\ \times \exp\big(\tfrac{-b^2}{s^2}\big)\, \mathrm{F}_{s,2}(a) \log\Big(\tfrac{1}{4\pi s^2} \exp\big(\tfrac{-b^2}{s^2}\big)\, \mathrm{F}_{s,2}(a)\Big). \end{aligned} \tag{A 5}$$

s^{-1}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$I_s(\boldsymbol{X},\boldsymbol{Y})$	0.0013	0.0917	0.3741	0.7004	1.0070	1.2845	1.5344	1.7602	1.9655	2.1532	2.3260	2.4858	2.6343	2.7730	2.9031	3.0255	3.1411	3.2506	3.3544	3.4535

Table 3. Values of the MI for a steady flow in a cylindrical straight tube, for scales $s^{-1} = 1, \ldots, 20$. They have been computed analytically and by numerical integration, and are used as ground truth to evaluate the correctness and accuracy of the proposed algorithm.

The integrals (A 4) and (A 5) have been computed by numerical integration and, applying (A 3), $I_s(\boldsymbol{X},\boldsymbol{Y})$ have been obtained for $s^{-1} = 1, \ldots, 20$ (see table 3).

Appendix B. Limit of ICI for infinitesimal scale

We can write the joint probability density (3.2) parameterizing the natural distribution of streamlines by its crossing at the inlet (2.4):

$$p_s(\boldsymbol{x},\boldsymbol{y}) = \int_{\mathcal{S}_{\mathrm{In}}} \frac{1}{C_{\mathrm{In}}C_{\mathrm{Out}}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2s^2R_{\mathrm{In}}^2}\right) \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{x}')\|^2}{2s^2R_{\mathrm{Out}}^2}\right) p_{\mathrm{In}}(\boldsymbol{x}')\,\mathrm{d}S'_{\mathrm{In}}$$

where the corresponding crossing at the outlet defines the map $\boldsymbol{f} : \mathcal{S}_{\mathrm{In}} \to \mathcal{S}_{\mathrm{Out}};\ \boldsymbol{x}' \mapsto \boldsymbol{f}(\boldsymbol{x}')$. For inifinitesimal scales, $s \to 0$, the Gaussian kernels will select points separated by distances of the order of s. Thus, we can denote

$$\boldsymbol{x} - \boldsymbol{x}' = sR_{\mathrm{In}}\boldsymbol{\alpha} \qquad \text{and} \qquad \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}) = sR_{\mathrm{Out}}\boldsymbol{\beta}$$

which imply $p_{\mathrm{In}}(\boldsymbol{x}') = p_{\mathrm{In}}(\boldsymbol{x}) + \mathcal{O}(s)$ and

$$\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}') = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x} - sR_{\mathrm{In}}\boldsymbol{\alpha}) = s\big(R_{\mathrm{Out}}\boldsymbol{\beta} + R_{\mathrm{In}}\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}\boldsymbol{f}(\boldsymbol{x})\big) + \mathcal{O}(s^2).$$

Here, the linearization of $\boldsymbol{f}(\boldsymbol{x}')$ around $\boldsymbol{x}$ is analogous to the one assumed for the definition of Lyapunov exponents, sharing the same exceptions at bifurcation points.

Considering that for $s \to 0$ the effects of the inlet boundary and curvature are negligible, we can rewrite the integral over $\mathcal{S}_{\mathrm{In}}$ as the unbounded integral over the tangent space parameterized by $\boldsymbol{\alpha}$:

$$p_s(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{(2\pi)^2 s^2 R_{\mathrm{Out}}^2} \int_{\mathbb{R}^2} \exp\left(-\frac{\|\boldsymbol{\alpha}\|^2 + \|\boldsymbol{\beta} - \frac{R_{\mathrm{In}}}{R_{\mathrm{Out}}}\mathsf{J}_{\boldsymbol{x}}\,\boldsymbol{\alpha}\|^2}{2}\right)\big(p_{\mathrm{In}}(\boldsymbol{x}) + \mathcal{O}(s)\big)\,\mathrm{d}\boldsymbol{\alpha}$$

where matrix notation is used for the Jacobian, $\mathsf{J}_{\boldsymbol{x}} = \big(\partial f^i(\boldsymbol{x})/\partial x^j\big)$. Reorganizing the exponent, the integral can be exactly solved, resulting in

$$p_s(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2\pi s^2 R_{\mathrm{Out}}^2\sqrt{|\mathsf{Q}_{\boldsymbol{x}}|}} \exp\left(-\tfrac{1}{2}\,\boldsymbol{\beta}^{\mathsf{T}}\mathsf{Q}_{\boldsymbol{x}}^{-1}\,\boldsymbol{\beta}\right)\big(p_{\mathrm{In}}(\boldsymbol{x}) + \mathcal{O}(s)\big),$$

where

$$\mathsf{Q}_{\boldsymbol{x}} = \mathsf{I} + \tfrac{R_{\mathrm{In}}^2}{R_{\mathrm{Out}}^2}\mathsf{J}_{\boldsymbol{x}}\mathsf{J}_{\boldsymbol{x}}^{\mathsf{T}}.$$

The limit of the marginal probabilities can be analogously deduced to be

$$p_s(\boldsymbol{x}) = p_{\mathrm{In}}(\boldsymbol{x}) + \mathcal{O}(s) \qquad \text{and} \qquad p_s(\boldsymbol{y}) = p_{\mathrm{Out}}(\boldsymbol{y}) + \mathcal{O}(s).$$

The MI (3.3) can be reexpressed by separating the factors under the logarithm in two terms,

$$I_s(\boldsymbol{X},\boldsymbol{Y}) = \int_{\mathcal{S}_{\mathrm{In}}}\int_{\mathcal{S}_{\mathrm{Out}}} p_s(\boldsymbol{x},\boldsymbol{y}) \log\left(\frac{p_s(\boldsymbol{x},\boldsymbol{y})}{p_s(\boldsymbol{x})}\right) \mathrm{d}S_{\mathrm{Out}}\,\mathrm{d}S_{\mathrm{In}} - \int_{\mathcal{S}_{\mathrm{Out}}} p_s(\boldsymbol{y})\log(p_s(\boldsymbol{y}))\,\mathrm{d}S_{\mathrm{Out}}.$$

The first term can be integrated by changing the integrating variable $\boldsymbol{y} \to \hat{\boldsymbol{\beta}} = {Q_{\boldsymbol{x}}}^{-1/2}\boldsymbol{\beta}$

$$\begin{aligned}\int_{\mathcal{S}_{\mathrm{In}}} \left(p_{\mathrm{In}}(\boldsymbol{x}) + \mathcal{O}(s)\right) \int_{\mathbb{R}^2} \frac{1}{2\pi}\exp\left(-\tfrac{1}{2}\|\hat{\boldsymbol{\beta}}\|^2\right)\left[-\tfrac{1}{2}\|\hat{\boldsymbol{\beta}}\|^2 - \log\left(2\pi s^2 R_{\mathrm{Out}}^2\sqrt{|Q_{\boldsymbol{x}}|}\right)\right]\mathrm{d}\hat{\boldsymbol{\beta}}\,\mathrm{d}S_{\mathrm{In}} \\ = -2\log(s) - \frac{1}{2} - \log\left(2\pi R_{\mathrm{Out}}^2\right) - \frac{1}{2}\int_{\mathcal{S}_{\mathrm{In}}} \log\left(|Q_{\boldsymbol{x}}|\right) p_{\mathrm{In}}(\boldsymbol{x})\,\mathrm{d}S_{\mathrm{In}} + \mathcal{O}(s).\end{aligned}$$

The second term gives the differential entropy of $p_{\mathrm{Out}}(\boldsymbol{y})$. Thus

$$I_s(\boldsymbol{X},\boldsymbol{Y}) = -2\log(s) - \frac{1}{2} - \log\left(2\pi R_{\mathrm{Out}}^2\right) - \frac{1}{2}\int_{\mathcal{S}_{\mathrm{In}}} \log\left(|Q_{\boldsymbol{x}}|\right) p_{\mathrm{In}}(\boldsymbol{x})\,\mathrm{d}S_{\mathrm{In}} + h[p_{\mathrm{Out}}] + \mathcal{O}(s).$$

We can specialize this expression to the inlet and outlet self-informations,

$$\begin{aligned} I_s(\boldsymbol{X},\boldsymbol{X}) &= -2\log(s) - \frac{1}{2} - \log(4\pi R_{\mathrm{In}}^2) + h[p_{\mathrm{In}}] + \mathcal{O}(s) \\ I_s(\boldsymbol{Y},\boldsymbol{Y}) &= -2\log(s) - \frac{1}{2} - \log(4\pi R_{\mathrm{Out}}^2) + h[p_{\mathrm{Out}}] + \mathcal{O}(s). \end{aligned}$$

Combining them, we obtain the behaviour of ICI (3.6) in the limit for $s \to 0$:

$$ICI_s = \frac{-1}{\log(s)}\left[\frac{1}{4}\int_{\mathcal{S}_{\mathrm{In}}} \log\left(\tfrac{1}{4}|Q_{\boldsymbol{x}}|\right) p_{\mathrm{In}}(\boldsymbol{x})\,\mathrm{d}S_{\mathrm{In}} + \Delta\right] + \mathcal{O}\left(\left(\tfrac{1}{\log(s)}\right)^2\right), \qquad (\mathrm{B}\,1)$$

where

$$\Delta = \max\left(0, \tfrac{1}{2}(h[p_{\mathrm{In}}] - h[p_{\mathrm{Out}}]) + \log\left(R_{\mathrm{Out}}/R_{\mathrm{In}}\right)\right).$$

Thus, $\lim_{s\to 0} ICI_s = 0$, being proportional to the slowly convergent factor $-1/\log(s)$.

REFERENCES

Alemaskin, K., Camesasca, M., Manas-Zloczower, I. & Kaufman, M. 2004 Entropic measures of mixing tailored for various applications. In *Materials Processing and Design: Modeling, Simulation and Applications — NUMIFORM 2004*, *AIP Conference Proceedings*, vol. 712, pp. 169–173. AIP Publishing.

Aneurist Consortium 2010 The @neurIST Integrated Project. `http://www.aneurist.org`.

Aref, H., Blake, J. R., Budišić, M., Cartwright, J. H. E., Clercx, H. J. H., Feudel, U., Golestanian, R., Gouillart, E., Guer, Y. L., van Heijst, G. F. *et al.* 2014 Frontiers of chaotic advection. *arXiv preprint arXiv:1403.2953* .

Arzani, A., Gambaruto, A. M., Chen, G. & Shadden, S. C. 2016 Lagrangian wall shear stress structures and near-wall transport in high-schmidt-number aneurysmal flows. *Journal of Fluid Mechanics* **790**, 158–172.

Bogunović, H., Pozo, J. M., Villa-Uriol, M. C., Majoie, C. B. L. M., van den Berg, R., Gratama van Andel, H. A. F., Macho, J. M., Blasco, J., San Román, L. & Frangi, A. F. 2011 Automated segmentation of cerebral vasculature with aneurysms in 3DRA and TOF-MRA using geodesic active regions: An evaluation study. *Medical Physics* **38**, 210.

Brandani, G. B., Schor, M., MacPhee, C. E., Grubmüller, H., Zachariae, U. & Marenduzzo, D. 2013 Quantifying disorder through conditional entropy: An application to fluid mixing. *PLoS ONE* **8** (6), e65617.

Cahill, N. D. 2010 Normalized measures of mutual information with general definitions of entropy for multimodal image registration. In *Biomedical Image Registration*, *Lecture Notes in Computer Science*, vol. 6204, pp. 258–268. Springer.

Camesasca, M., Kaufman, M. & Manas-Zloczower, I. 2006 Quantifying fluid mixing with the Shannon entropy. *Macromolecular Theory and Simulations* **15** (8), 595–607.

Camesasca, M. & Manas-Zloczower, I. 2009 Danckwerts revisited — the use of entropy to define scale and intensity of segregation. *Macromolecular Theory and Simulations* **18** (2), 87–96.

Cebral, J. R., Mut, F., Weir, J. & Putman, C. M. 2011 Association of hemodynamic characteristics and cerebral aneurysm rupture. *American Journal of Neuroradiology* **32** (2), 264–270.

Chen, S., Ma, B. & Zhang, K. 2009 On the similarity metric and the distance metric. *Theoretical Computer Science* **410** (24), 2365–2376.

Childs, P. R. N. 2010 *Rotating flow*. Elsevier.

Danckwerts, P. V. 1952 The definition and measurement of some characteristics of mixtures. *Applied Scientific Research, Section A* **3** (4), 279–296.

Domingues, N., Gaspar-Cunha, A. & Covas, J. A. 2008 Global mixing indices for single screw extrusion. *International Journal of Material Forming* **1** (1), 723–726.

Ford, M. D., Hoi, Y., Piccinelli, M., Antiga, L. & Steinman, D. A. 2009 An objective approach to digital removal of saccular aneurysms: technique and applications. *The British Journal of Radiology* **82** (special_issue_1), S55–S61.

Funakoshi, M. 2008 Chaotic mixing and mixing efficiency in a short time. *Fluid Dynamics Research* **40** (1), 1–33.

Geers, A. J., Larrabide, I., Radaelli, A. G., Bogunovic, H., Kim, M., van Andel, H. A. F. G., Majoie, C. B., VanBavel, E. & Frangi, A. F. 2011 Patient-specific computational hemodynamics of intracranial aneurysms from 3D rotational angiography and CT angiography: an in vivo reproducibility study. *American Journal of Neuroradiology* **32** (3), 581–586.

Gopalakrishnan, S. S., Pier, B. & Biesheuvel, A. 2014 Dynamics of pulsatile flow through model abdominal aortic aneurysms. *Journal of Fluid Mechanics* **758**, 150–179.

Guida, A., Nienow, A. W. & Barigou, M. 2010 Shannon entropy for local and global description of mixing by lagrangian particle tracking. *Chemical Engineering Science* **65** (10), 2865–2883.

Hollander, M., Wolfe, D. A. & Chicken, E. 2013 *Nonparametric statistical methods*. John Wiley & Sons.

Jang, B. & Funakoshi, M. 2010 Chaotic mixing in a helix-like pipe with periodic variations in curvature and torsion. *Fluid Dynamics Research* **42** (3), 035506.

Jaynes, E. T. 1968 Prior probabilities. *Systems Science and Cybernetics, IEEE Transactions on* **4** (3), 227–241.

Jaynes, E. T. 2003 *Probability theory: the logic of science*. Cambridge university press.

Koenker, R. 1981 A note on studentizing a test for heteroscedasticity. *Journal of Econometrics* **17** (1), 107–112.

Kolmogorov, A. 1956 On the Shannon theory of information transmission in the case of continuous signals. *Information Theory, IRE Transactions on* **2** (4), 102–108.

Kruijt, P. G. M., Galaktionov, O. S., Anderson, P. D., Peters, G. W. M. & Meijer, H. E. H. 2001 Analyzing mixing in periodic flows by distribution matrices: mapping method. *AIChE Journal* **47** (5), 1005–1015.

Lee, K., Parker, K., Caro, C. & Sherwin, S. 2008 The spectral/hp element modelling of steady flow in non-planar double bends. *International Journal for Numerical Methods in Fluids* **57** (5), 519–529.

Lin, Z., Thiffeault, J.-L. & Doering, C. R. 2011 Optimal stirring strategies for passive scalar mixing. *Journal of Fluid Mechanics* **675**, 465–476.

Ling, F. H. 1994 Interpolated Poincaré map and application to mixing problems. *Chaos, Solitons & Fractals* **4** (5), 681–692.

Loskutov, A. 2010 Fascination of chaos. *Physics-Uspekhi* **53** (12), 1257.

Lucas, D. & Kerswell, R. 2014 Spatiotemporal dynamics in two-dimensional kolmogorov flow over large domains. *Journal of Fluid Mechanics* **750**, 518–554.

Lueptow, R. M., Docter, A. & Min, K. 1992 Stability of axial flow in an annulus with a rotating inner cylinder. *Physics of Fluids A: Fluid Dynamics* **4** (11), 2446–2455.

Mann, H. B. & Whitney, D. R. 1947 On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* **18** (1), 50–60.

Mathew, G. & Mezić, I. 2011 Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Physica D: Nonlinear Phenomena* **240** (4), 432–442.

Mathew, G., Mezić, I. & Petzold, L. 2005 A multiscale measure for mixing. *Physica D: Nonlinear Phenomena* **211** (1–2), 23–46.

McDaid, A. F., Greene, D. & Hurley, N. 2011 Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv preprint arXiv:1110.2515* .

Millan, R. D., Dempere-Marco, L., Pozo, J. M., Cebral, J. R. & Frangi, A. F. 2007 Morphological characterization of intracranial aneurysms using 3-D moment invariants. *Medical Imaging, IEEE Transactions on* **26** (9), 1270–1282.

Nuttall, A. H. 1975 Some integrals involving the Q_M function (corresp.). *Information Theory, IEEE Transactions on* **21** (1), 95–96.

Ottino, J. M. 1989 *The Kinematics of Mixing: Stretching, Chaos, and Transport*, *Cambridge Texts in Applied Mathematics*, vol. 3. Cambridge University Press.

Ottino, J. M. 1990 Mixing, chaotic advection, and turbulence. *Annual Review of Fluid Mechanics* **22** (1), 207–254.

Paninski, L. 2003 Estimation of entropy and mutual information. *Neural Computation* **15** (6), 1191–1253.

Piccinelli, M., Veneziani, A., Steinman, D., Remuzzi, A. & Antiga, L. 2009 A framework for geometric analysis of vascular structures: application to cerebral aneurysms. *Medical Imaging, IEEE Transactions on* **28** (8), 1141–1155.

Piessens, R. 1983 *Quadpack: a Subroutine Package for Automatic Integration*, *Springer Series in Computational Mathematics*, vol. 1. Springer-Verlag.

R Core Team 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reymond, P., Merenda, F., Perren, F., Rüfenacht, D. & Stergiopulos, N. 2009 Validation of a one-dimensional model of the systemic arterial tree. *American Journal of Physiology-Heart and Circulatory Physiology* **297** (1), H208–H222.

Robert, C. & Casella, G. 2013 *Monte Carlo Statistical Methods*. Springer Science & Business Media.

Sandri, M. 1996 Numerical calculation of Lyapunov exponents. *Mathematica Journal* **6** (3), 78–84.

Schnell, S., Ansari, S. A., Vakil, P., Wasielewski, M., Carr, M. L., Hurley, M. C., Bendok, B. R., Batjer, H., Carroll, T. J., Carr, J. & Markl, M. 2014 Three-dimensional hemodynamics in intracranial aneurysms: influence of size and morphology. *Journal of Magnetic Resonance Imaging* **39** (1), 120–131.

Schroeder, W., Martin, K. & Lorensen, B. 2003 *The Visualization Toolkit: An Object Oriented Approach to 3D Graphics 3rd Edition*. Kitware Inc.

Scott, S. E., Redd, T. C., Kuznetsov, L., Mezić, I. & Jones, C. K. R. T. 2009 Capturing deviation from ergodicity at different scales. *Physica D: Nonlinear Phenomena* **238** (16), 1668–1679.

Shannon, C. E. 1948 A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423.

Shannon, C. E. & Weaver, W. 1949 *The Mathematical Theory of Information*. Urbana, Illinois: University of Illinois Press.

Tang, X. Z. & Boozer, A. H. 1996 Finite time Lyapunov exponent and advection-diffusion equation. *Physica D: Nonlinear Phenomena* **95** (3), 283–305.

Tarantola, A. 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Tarantola, A. & Mosegaard, K. 2007 Mapping of probabilities theory for the interpretation of uncertain physical measurements. `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.177.9042`.

Thiffeault, J.-L. 2012 Using multiscale norms to quantify mixing and transport. *Nonlinearity* **25** (2), R1.

Tucker, C. L. & Peters, G. W. M. 2003 Global measures of distributive mixing and their behavior in chaotic flows. *Korea-Australia Rheology Journal* **15** (4), 197–208.

Villa-Uriol, M. C., Berti, G., Hose, D. R., Marzo, A., Chiarini, A., Penrose, J., Pozo, J. M., Schmidt, J. G., Singh, P., Lycett, R., Larrabide, I. & Frangi, A. F. 2011 @neurist complex information processing toolchain for the integrated management of cerebral aneurysms. *Interface Focus* **1** (3), 308–319.

Vinh, N. X., Epps, J. & Bailey, J. 2010 Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research* **11**, 2837–2854.

VTK community 2014 VTK visualization toolkit. `http://www.vtk.org`.

Wereley, S. T. & Lueptow, R. M. 1999 Velocity field for Taylor–Couette flow with an axial flow. *Physics of Fluids* **11** (12), 3637–3649.

Wiggins, S. & Ottino, J. M. 2004 Foundations of chaotic mixing. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **362** (1818), 937–970.

Xia, H. M., Shu, C., Chew, Y. T. & Wang, Z. P. 2010 Approximate mapping method for prediction of chaotic mixing in spatial-periodic microchannel. *Chemical Engineering Research and Design* **88** (10), 1419–1426.

Zu, P., Chen, L. & Xin, J. 2015 A computational study of residual kpp front speeds in time-periodic cellular flows in the small diffusion limit. *Physica D: Nonlinear Phenomena* **311**, 37–44.