On the non-local geometry of turbulence

IVÁN BERMEJO-MORENO AND D. I. PULLIN

Graduate Aeronautical Laboratories, 205-45, California Institute of Technology,
Pasadena, CA 91125, USA
ibermejo@caltech.edu

A multi-scale methodology for the study of the non-local geometry of eddy structures in turbulence is developed. Starting from a given three-dimensional field, this consists of three main steps: extraction, characterization and classification of structures. The extraction step is done in two stages. First, a multi-scale decomposition based on the curvelet transform is applied to the full three-dimensional field, resulting in a finite set of component three-dimensional fields, one per scale. Second, by iso-contouring each component field at one or more iso-contour levels, a set of closed iso-surfaces is obtained that represents the structures at that scale. The characterization stage is based on the joint probability density function (p.d.f.), in terms of area coverage on each individual iso-surface, of two differential-geometry properties, the shape index and curvedness, plus the stretching parameter, a dimensionless global invariant of the surface. Taken together, this defines the geometrical signature of the iso-surface. The classification step is based on the construction of a finite set of parameters, obtained from algebraic functions of moments of the joint p.d.f. of each structure, that specify its location as a point in a multi-dimensional 'feature space'. At each scale the set of points in feature space represents all structures at that scale, for the specified iso-contour value. This then allows the application, to the set, of clustering techniques that search for groups of structures with a common geometry. Results are presented of a first application of this technique to a passive scalar field obtained from 512^3 direct numerical simulation of scalar mixing by forced, isotropic turbulence ($Re_\lambda = 265$). These show transition, with decreasing scale, from blob-like structures in the larger scales to blob- and tube-like structures with small or moderate stretching in the inertial range of scales, and then toward tube and, predominantly, sheet-like structures with high level of stretching in the dissipation range of scales. Implications of these results for the dynamical behaviour of passive scalar stirring and mixing by turbulence are discussed.

1. Introduction

Observation of natural fluid flows indicates the presence of structures with apparent repeating geometries. Vortical structures in multiphase flows are commonly observed. The roll-up of an ocean wave before it breaks (pressure- and gravity-driven flow), the swirling motion of a hurricane around its centre (pressure-driven flow affected by Coriolis and friction forces) or that of the stellar gas accretion disk occurring during the formation of galaxies (gravitational-driven flow), and the Kelvin–Helmholtz wave clouds formed between two layers of air of different density and speeds (shear-driven flow) are just a few examples. Turbulent fluid flows are no exception, often adding

levels of complexity to the structure geometry owing to the multiple scales that comprise such flows.

Visualization experiments have provided means for the systematic study of geometrical structures in fluid flows and have substantially increased the number of known flows where repeating geometrical patterns are present. Experimental study of the flow past cylinders and spheres led to the discovery of the Kármán vortex street while experiments in turbulent mixing layers resulted in an exhaustive study of 'coherent' vortical structures (see Brown & Roshko 1974). This work has stimulated theoretical analysis of pattern formation, for example the description of eddying motions and flow patterns based on critical-point theory (see Perry & Chong 1987).

Direct numerical simulations (DNS) have also proven to be a valuable tool in the search for geometrical structures in fluid flows, for example the organized cylindrical elongated vortices (so-called 'worms') found in the intense vorticity of isotropic turbulence (Siggia 1981; Kerr 1985; Jiménez et al. 1993). The 'worms', however, remain a puzzle; their contribution to the kinetic energy dissipation is almost negligible and their role in turbulence dynamics remains an open question.

Structures in turbulent flows can be considered a consequence of the forces and boundary conditions driving the flow, but also can be seen as themselves producing some intrinsic properties of the turbulence. In the multi-scale ansatz based on self-similarity and the idea of energy cascade (Richardson 1922; Kolmogorov 1941, 1962), the external forces and the boundary conditions affect mainly large energy-containing scales, with diminished influence on progressively smaller eddies. The energy-containing scales then depend strongly on the external forces and boundary conditions and are not expected to be universal, while small-scale structures may be related to universal properties of turbulence, thereby exhibiting a generic geometric signature that may be characteristic of efficient cascade dynamics. A geometrical characterization of those structures could provide improved understanding of cascade mechanics and dissipation-range dynamics, contributing potentially to the development of structure-based models of turbulence fine scales (see Townsend 1951; Tennekes 1968; Lundgren 1982; Pullin & Saffman 1993), subgrid-scale models for large-eddy simulation (see Misra & Pullin 1997) and simulation methods based on multi-resolution decomposition by means of the wavelet transform (see Farge 1992; Meneveau 1991; Farge et al. 1996; Farge, Schneider & Kevlahan 1999). Further, a better understanding of eddy structure at large Reynolds number may provide important insight into possibly singular or near-singular structures in the dynamics of the Euler equations (see Hou & Li 2006) by elucidating the geometrical characterization of sites within the turbulent field where extreme dissipative or vortical events occur, and which are candidates for singularity formation in the limit of vanishing viscosity.

Prior work on the identification of structures in turbulence addresses mainly the identification of vortex tube- and sheet-like structures with emphasis on vortex tubes. But the importance of sheet-like structures, where significant turbulent kinetic dissipation may be concentrated owing to their high amplitude of strain rate, and which may produce tubes by roll-up instabilities, has led to renewed interest in sheets. Most methods either for tubes, or sheets or both, are based on local measures of scalar fields obtained from the velocity-gradient tensor and/or the pressure field. Chong, Perry & Cantwell (1990) classified regions with complex eigenvalues of the velocity-gradient tensor as vortex tubes (since the local streamlines are then closed or spiral in a reference frame moving with the fluid). The second-order invariant, Q, of the velocity-gradient tensor was used by Hunt, Wray & Moin (1988), to define a

vortex tube as the region with a positive value of Q, and the condition of a pressure lower than the ambient, while Ashurst *et al.* (1987) based their identification criterion on the sign of the intermediate eigenvalue of the rate-of-strain tensor, **S**. Analysis of the relation between the vorticity field and the eigenvector associated with the intermediate eigenvalue of the strain-rate tensor showed a tendency for both vector fields to be statistically aligned. Tanaka & Kida (1993) extended the identification criterion based on Q for the extraction of both tubes and sheets. Jeong & Hussain (1995) proposed a method based on the second largest eigenvalue, λ_2, of the tensor **L** formed by summing the products of the symmetric, **S**, and antisymmetric, **Ω**, parts of the velocity-gradient tensor with themselves, $\boldsymbol{L} = \boldsymbol{SS} + \boldsymbol{\Omega\Omega}$. They define a vortex core as the region where λ_2 is negative. Horiuti (2001) combined this methodology with the physical explanations of the alignment of vorticity and the eigenvector associated with the intermediate eigenvalue of **S** (Andreotti 1997) to develop a new method in which the eigenvalues and eigenvectors of **L** are reordered based on their alignment with the vorticity; then, regions are classified into vortex tubes, and so-called flat vortex sheets and curved vortex sheets depending on the relations of those reordered eigenvalues. Horiuti & Takagi (2005) proposed an improved method for the eduction of vortex sheet structures, based on local values of the largest eigenvalue of the tensor $A_{ij} = S_{ik}\Omega_{kj} + S_{jk}\Omega_{ki}$, once the eigenvalue corresponding to the eigenvector maximally aligned with the vorticity is removed. Based solely on the pressure field, Miura & Kida (1997) developed a methodology for extracting axes of tubular vortices as the loci of sectionally local minima of the pressure field (obtained by means of the sign of the second largest eigenvalue of the pressure Hessian evaluated at each point; positive values indicate pressure minima).

Most of the existing methods of identification are, nonetheless, local, and are based on point-wise quantities used to discriminate whether the point belongs to one type of structure or another (or none). Regions of points sharing a common identity based on the local criterion applied can then be formed, but often that local analysis is the end of the identification process. Visualization of such regions has proved a helpful tool in its analysis, but here we seek a more automated, systematic approach to structure characterization. An extended structural and fractal description of turbulence was proposed by Moisy & Jiménez (2004) who applied a box-counting method to sets of points of intense vorticity and strain-rate magnitude (educed by thresholding). They also analyse geometrically individual structures, defined as a connected set of points satisfying the threshold criterion (thus, considering the spatial extent of such structures), based on their volume and spatial distribution, finding that intense vorticity and dissipation structures are concentrated in clusters of inertial size.

The present approach is based on a non-local multi-scale methodology for the extraction, characterization and classification of structures in turbulence. It is non-local, focusing on the spatial extent of structures. Conventional turbulence quantities such as properties of the local velocity-gradient tensor are not used. The multi-scale analysis is performed through the curvelet transform, a higher dimensional generalization of the wavelet transform. In the present illustrative application, the structures are defined as iso-surfaces, extracted at different scales, from a passive scalar field convected and diffused by isotropic turbulence. The characterization and classification steps are based on measures of the geometry of iso-surfaces. The problem of shape analysis of free-form surfaces has been widely studied in the fields of computer graphics, computer vision, and image understanding (see Campbell & Flynn 2001; Iyer *et al.* 2005; Dorai & Jain 1997; Osada *et al.* 2001; Zaharia & Prêteux 2001). Our method characterizes each individual structure in terms of local differential-geometry

properties. Structure identification in terms of non-local characterization is done via area-based probability density functions of those geometrical properties. Classification is based on this geometrical characterization of individual structures and is enhanced via clustering techniques. Clustering algorithms allow the eduction of groups of structures without the need for strong *a priori* assumptions about their properties.

This paper is organized as follows: a description of the three main steps of this methodology – extraction, characterization and classification – is presented in §2, §3 and §4 respectively, with emphasis on the conceptual basis and on some particular implementation details. In §5, a system test that validates the methodology applied to a virtual world of modelled structures is presented. In §6, we show the results of its application to the eduction of extended passive scalar structures obtained from a direct-numerical simulation (DNS) database. We emphasize that the tools developed here – the multi-resolution analysis, geometric characterization, spectral projection and clustering algorithms – can be applied to many scalar and tensor fields in turbulence, and in fields beyond fluid mechanics.

2. Extraction of structures

The main requirement imposed on the extraction process is to enable eduction of structures associated with different ranges of scales. Although scale decomposition is commonly defined in Fourier space, the nature of Fourier basis functions, that are localized in wavenumber but not in physical space, makes top-hat window filtering in Fourier space inappropriate for the purpose of educing structures that are extended but compact in physical space. Thus, a transformation with basis functions that are localized both in Fourier space, where the ranges of scales are defined, and in physical space, where the structures are to be educed, is required. For this purpose, the curvelet transform (Candès & Donoho 2003*a*,*b*) in its three-dimensional discretized version (Ying, Demanet & Candès 2005; Candès *et al.* 2005) is used. Owing to the multi-dimensional character of their definition, curvelets, unlike wavelets, are naturally suited for detecting, organizing, and providing a compact representation of intermediate multi-dimensional structures.

2.1. *The curvelet transform*

Curvelets, the basis functions of the curvelet transform, are localized in scale (frequency/Fourier space), position (physical space) and orientation (unlike wavelets). The frequency space is smoothly windowed in radial and angular spherical coordinates, providing the decomposition in different scales and orientations, respectively. For a given scale†, j, the radial window smoothly extracts the frequency contents near the dyadic corona $[2^{j-1}, 2^{j+1}]$. A low-pass radial window is introduced for the coarsest scale, j_0. The unit sphere representing all directions in $\mathbb{R}^3$ is partitioned, for each scale $j > j_0$, into $O(2^{j/2} \cdot 2^{j/2}) = O(2^j)$ smooth angular windows, each with a disk-like support of radius $O(2^{-j/2})$. In a discrete three-dimensional data field, of uniform grid of size n^3, the last scale, which extracts the highest frequency content, is given by $j_e = \log_2(n/2)$.

Denoting by $f(n_1, n_2, n_3)$ the scalar field, where $0 \leqslant n_i < n$, being n the number of grid points in each direction, the discrete version of the curvelet transform (see Ying

---

† The term scale, when referred to the index j in curvelet space, denotes in fact the range of scales in physical space that results from the radial window filter in Fourier space.
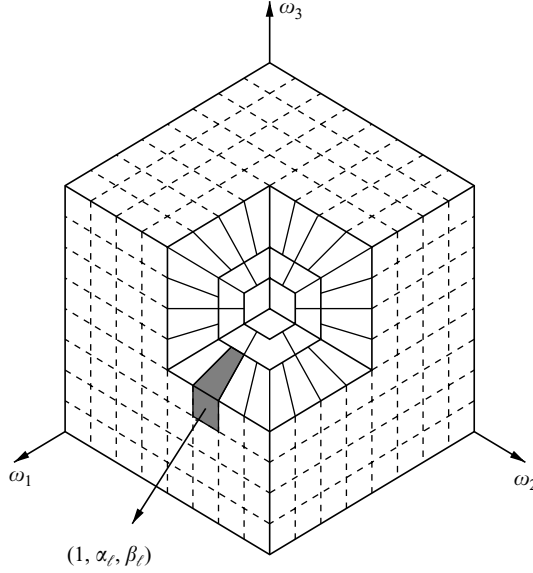
FIGURE 1. Frequency window $\tilde{U}_{j,\ell}$ (darkened region) defined in the three-dimensional discrete curvelet transform, extracting the frequency content near the wedge with centre slope $(1, \alpha_\ell, \beta_\ell)$ (figure based on Ying *et al.* 2005).

et al. 2005) provides a set of coefficients $c^D(j, \ell, k)$ defined as

$$c^D(j, \ell, k) \equiv \sum_{n_1, n_2, n_3} f(n_1, n_2, n_3) \overline{\varphi^D_{j,\ell,k}(n_1, n_2, n_3)} \tag{2.1}$$

where $j, \ell \in \mathbb{Z}$, $k = (k_1, k_2, k_3)$ (j represents the scale, ℓ the orientation and k the spatial location); $\varphi^D_{j,\ell,k}(n_1, n_2, n_3)$ are the curvelets, which are defined in Fourier space by

$$\hat{\varphi}^D_{j,\ell,k}(\omega) \equiv \tilde{U}_{j,\ell}(\omega) \exp\left(\frac{-2\pi i \sum_{i=1}^{3} \dfrac{k_i \omega_i}{L_{i,j,\ell}}}{\sqrt{\prod_{i=1}^{3} L_{i,j,\ell}}} \right) \tag{2.2}$$

where $\tilde{U}_{j,\ell}(\omega)$ is the frequency window $\tilde{U}_{j,\ell}(\omega) = \tilde{W}_j(\omega) \tilde{V}_{j,\ell}(\omega)$, with $\tilde{W}_j(\omega)$ and $\tilde{V}_{j,\ell}(\omega)$ the radial and angular windows. A Cartesian coronae is used, so that

$$\tilde{W}_{j_0}(\omega) = \Phi_{j_0}(\omega); \qquad \tilde{W}_j(\omega) = \sqrt{\Phi^2_{j+1}(\omega) - \Phi^2_j(\omega)}, \qquad j > j_0, \tag{2.3}$$

where $\Phi_j(\omega_1, \omega_2, \omega_3) = \phi(2^{-j}\omega_1) \phi(2^{-j}\omega_2) \phi(2^{-j}\omega_3)$, and ϕ is a smooth function such that $0 \leqslant \phi \leqslant 1$; it is equal to 1 on $[-1, 1]$ and zero outside $[-2, 2]$. The angular window for the ℓth direction is defined (for example, in the $\omega_1 > 0$ face of the unit cube) as

$$\tilde{V}_{j,\ell}(\omega) = \tilde{V}\left(2^{j/2} \frac{\omega_2 - \alpha_\ell \omega_1}{\omega_1} \right) \tilde{V}\left(2^{j/2} \frac{\omega_3 - \beta_\ell \omega_1}{\omega_1} \right) \tag{2.4}$$

where $(1, \alpha_\ell, \beta_\ell)$ is the direction of the centreline of the wedge (see figure 1) defining the centre slope for the ℓth wedge. Wherever three smooth angular windows $\tilde{V}_{j,\ell}$, $\tilde{V}_{j,\ell'}$ and $\tilde{V}_{j,\ell''}$ overlap, they are redefined as $(\tilde{V}_{j,\ell}, \tilde{V}_{j,\ell'}, \tilde{V}_{j,\ell''})/\sqrt{\tilde{V}^2_{j,\ell} + \tilde{V}^2_{j,\ell'} + \tilde{V}^2_{j,\ell''}}$.

$\tilde{W}_j(\omega)$ and $\tilde{V}(\omega)$ satisfy

$$\sum_{j \geqslant j_0} \tilde{W}_j^2(\omega) = 1, \qquad \sum_{\ell=-\infty}^{\infty} \tilde{V}^2(t - 2\ell) = 1. \qquad (2.5)$$

Curvelets form a tight-frame in $L^2(\mathbb{R}^3)$. Any function $f \in L^2(\mathbb{R}^3)$ can be expanded as $f = \sum_{j,\ell,k} \langle \phi_{j,\ell,k}, f \rangle \phi_{j,\ell,k}$, where $\phi_{j,\ell,k}$ is the curvelet at scale j, orientation ℓ and position $k = (k_1, k_2, k_3)$. Parseval's identity holds: $\sum_{j,\ell,k} | \langle f, \phi_{j,\ell,k} \rangle |^2 = \| f \|_{L^2(\mathbb{R}^3)}^2$. The effective longitudinal and cross-sectional dimensions (length and width), of curvelet basis functions in physical space follow the relation width$\approx$length^2 (parabolic scaling). As a consequence of this parabolic scaling, curvelets are an optimal (sparse) basis for representing surface-like singularities of codimension one. These are three of the most remarkable properties of the curvelet transform.

Here we apply the curvelet transform to a passive scalar field, at an instant in time, that has resulted from mixing by a forced isotropic turbulent velocity field. In what follows, we cast the discussion in terms of this example application but again emphasize its broader applicability to other fields. The curvelet transform, when applied to the passive scalar field, allows a multi-scale decomposition by filtering in curvelet space the different scales of interest $j = j_0, \ldots, j_e$, individually or in groups. In addition, for anisotropic fields with privileged direction(s) (e.g. shear flows), a multi-orientation decomposition may be useful for studying structures according to their directionality (by using the angular window filtering in frequency space of the curvelet transform (index ℓ in curvelet space)). Throughout this paper, only the multi-scale decomposition is used, which could be also attained by other multi-resolution techniques sharing the same choice of sub-band radial filtering decompostion in Fourier space. Nevertheless, those capabilities that set curvelets apart from other multi-resolution techniques, e.g. multi-orientation decomposition and compact representation of surface-like singularities, justify its early implementation within the frame of this methodology, enhancing its potential applications and possibilities of expansion.

For each scale $j = j_0, \ldots, j_e$, a new passive scalar field is obtained after filtering all other scales ($j\prime \neq j$) in curvelet space and inverse transforming to physical space. Thus, a set of $j_e - j_0 + 1$ filtered scalar fields results from the original field. The volume-based probability density functions of the filtered fields are, in general, different from each other and from the original field; their comparison can be useful in determining how the original scalar field is distributed among the different scales.

After this multi-scale analysis, a second step is applied in the extraction process, by which the structures of interest associated with each relevant range of scales are educed. Currently those structures of interest are defined as the individual disconnected surfaces obtained by iso-contouring each filtered scalar field at particular contour values (for example, the mean value of that filtered scalar field plus a multiple of its standard deviation). See Appendix A for a physical interpretation of the educed structures following this multi-scale decomposition plus iso-contouring procedure.

2.2. *Periodic reconnection*

In the case of scalar fields with periodic boundaries, an additional step is included in the extraction process, to reconnect those structures intersecting boundaries with their periodic continuation on the opposite boundaries.

3. Characterization of structures

We seek a geometrical characterization able to distinguish structures based on their shape. A two-step method is used: first, a suitable set of differential-geometry properties is obtained locally (at all points of the surface), and then, area-based probability density functions of those local properties are calculated, making the transition from local to non-local (in the surface sense) possible.

3.1. *Shape index and curvedness*

Shape index, Υ, and curvedness, Λ, (see Koenderink & van Doorn 1992) are the differential-geometry properties chosen to represent locally the geometry of the surface. They are related to the *principal curvatures* $\{\kappa_1, \kappa_2\}$ of a surface at a given point by

$$\Upsilon \equiv -\frac{2}{\pi} \arctan\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right), \qquad \Lambda \equiv \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}. \tag{3.1}$$

Υ is scale-independent, whereas Λ is scale-dependent, having the dimensions of a reciprocal length. The scaling is such that, for example, Λ at every point on a sphere equals the absolute value of its reciprocal radius, $1/R$, whereas the cylinder of radius R presents $\Lambda = 1/(\sqrt{2}R)$ for all points. The principal curvatures, $\{\kappa_1, \kappa_2\}$, are obtained as the maximum and minimum values of the normal curvature, κ_n, in all possible directions of the tangent plane defined at the point P of the surface of study. The normal curvature, κ_n, at a point P in a given direction $\boldsymbol{a}$ of the tangent plane, defined as the division of the second and first fundamental forms of differential geometry applied in that direction, $\boldsymbol{a}$, can also be interpreted as the inverse of the radius of curvature, R, of the curve obtained as the intersection of the surface and the plane defined by the direction $\boldsymbol{a}$ and the normal $\boldsymbol{N}$ to the surface at the point P. Thus higher values of the curvedness correspond to smaller radius of curvature (and, therefore, more locally curved surface at P). All regular patches of a regular surface M map on the domain $(\Upsilon, \Lambda) \in [-1, +1] \times \mathbb{R}^+$, except for the planar patch, which has an indeterminate shape index and nil curvedness (since $\kappa_1 = \kappa_2 = 0$).

The mapping $(\kappa_1, \kappa_2) \rightarrow (\Upsilon, \Lambda)$ represents (see figure 2) a transformation from Cartesian coordinates (κ_1, κ_2) to non-standard polar coordinates (Υ, Λ). For any point in the (κ_1, κ_2)-plane, Υ contains the information on the direction (measured as the angle, ϕ, with respect to the axis $\kappa_1 - \kappa_2$, rescaled into the range $[-1, +1]$ by $\Upsilon = -2\phi/\pi$) whereas Λ contains the information on the distance, ρ, to the origin (rescaled as $\Lambda = \rho/\sqrt{2}$). The convention chosen when ordering the principal curvatures ($\kappa_1 \geqslant \kappa_2$) implies that only the region $\kappa_1 - \kappa_2 \geqslant 0$ of the (κ_1, κ_2)-plane is accessible (see figure 2). Therefore, the polar angle ϕ can only have values in the range $[-\pi/2, +\pi/2]$, and, consequently, $\Upsilon = -\phi/\frac{1}{2}\pi \in [-1, +1]$ covers all the possible cases, thus making the mapping $(\kappa_1, \kappa_2) \rightarrow (\Upsilon, \Lambda)$ injective (excluding the point $(\kappa_1, \kappa_2) = (0, 0)$ from its domain) by eliminating the multi-valuedness of the arctan function used in the definition of the shape index. The absolute value of the shape index $S \equiv |\Upsilon|$ represents the local shape of the surface at the point P, with $0 \leqslant S \leqslant 1$. Its sign indicates the direction of the normal, distinguishing, for example, convex from concave elliptical points. Figure 3 shows the range of values of Υ and sketches of the local shapes associated with representative values, with the names of corresponding points. Figure 4 shows the mapping of both Υ and Λ in the plane of principal curvatures.
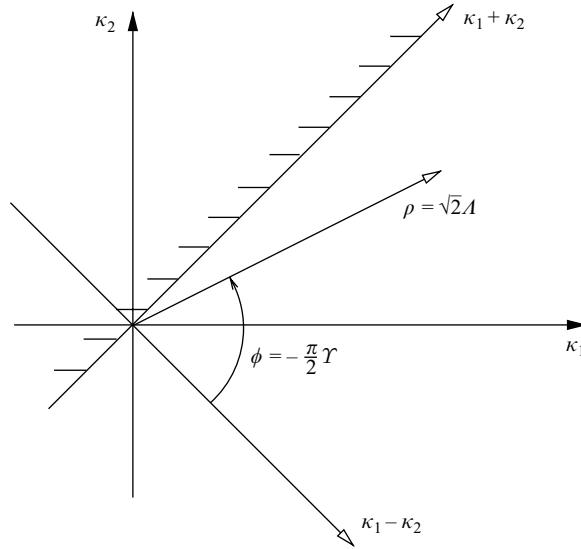
FIGURE 2. Transformation from (κ_1, κ_2) to (Υ, Λ).
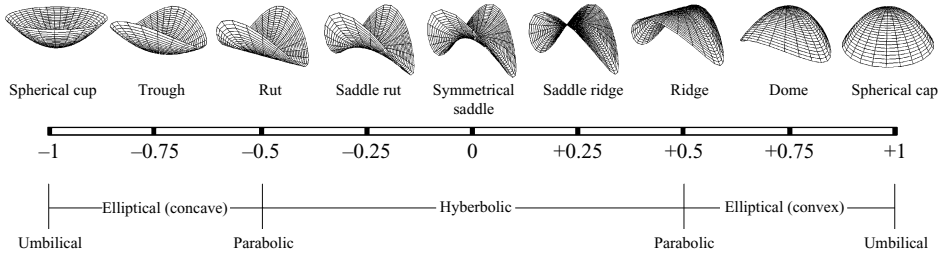

FIGURE 3. Range of shape index, Υ, with its most representative associated local shapes.

In terms of the Gaussian and mean curvatures, K and H, the shape index and curvedness can be expressed as

$$\Upsilon = -\frac{2}{\pi} \arctan\left(\frac{H}{\sqrt{H^2 - K}}\right), \qquad \Lambda = \sqrt{2H^2 - K}. \qquad (3.2)$$

3.2. *Joint probability density function (p.d.f.)*

From the pointwise Υ and Λ, a two-dimensional area-based joint probability density function in the space of (S, Λ) can be obtained. Since Λ is scale dependent, in order to compare the shape of surfaces of different sizes, a non-dimensionalization is required for each surface. Selection of the appropriate length scale for this purpose is critical; several can be obtained from global geometrical invariants of the surface, such as the square root of its area (A), the cubic root of its volume (V), etc. Here we define

$$C \equiv \mu\Lambda, \qquad \mu \equiv \frac{3\,V}{A}. \qquad (3.3)$$

For the sphere, $C = 1$. The definition of a volume implies that the structure under consideration is a closed surface. Thus, only closed surfaces educed from the passive scalar field are studied. For homogeneous isotropic turbulence in a periodic domain, for example, only those structures with infinite extent will not be closed.
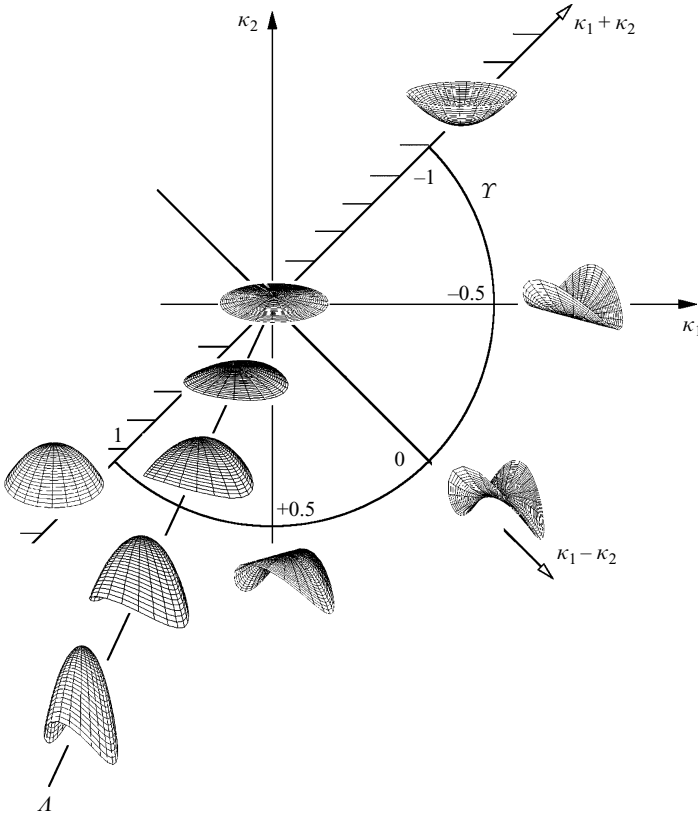
FIGURE 4. Representative local shape in the combined (κ_1, κ_2) plane.

All others, following periodic reconnection, will be closed. Another dimensionless global parameter useful in the characterization of the geometry of the educed closed structures is

$$\lambda \equiv \sqrt[3]{36\pi}\,\frac{V^{2/3}}{A} \tag{3.4}$$

which represents the stretching of the structure; the lower its value, the more stretched the structure is. For the sphere $\lambda = 1$.

The area-based joint p.d.f. $\mathscr{P}(S, C)$, $\int \int \mathscr{P}(S, C)\,\mathrm{d}S\,\mathrm{d}C = 1$, contains non-local information on the geometry of the surface. $\mathscr{P}(S, C)$ can be geometrically interpreted as a representation of how the local shape, S, is distributed across the different (relative) length scales present on the surface, given by C, in terms of area coverage. For closed surfaces, their geometry and topology are related by the Gauss–Bonnet theorem, which imposes an integral constraint on the area-based joint probability density function of S and C (see Appendix B for details).

3.3. *Signature of a structure*

We consider $\mathscr{P}(S, C)$ plus its associated one-dimensional marginal p.d.f.s,

$$\mathscr{P}_S(S) = \int \mathscr{P}(S, C)\,\mathrm{d}C, \qquad \mathscr{P}_C(C) = \int \mathscr{P}(S, C)\,\mathrm{d}S, \tag{3.5}$$

to be the signature of the structure. This is complemented with its area A and λ, representing the stretching of the structure. We find useful to display $\mathscr{P}(S, C)$ mapped
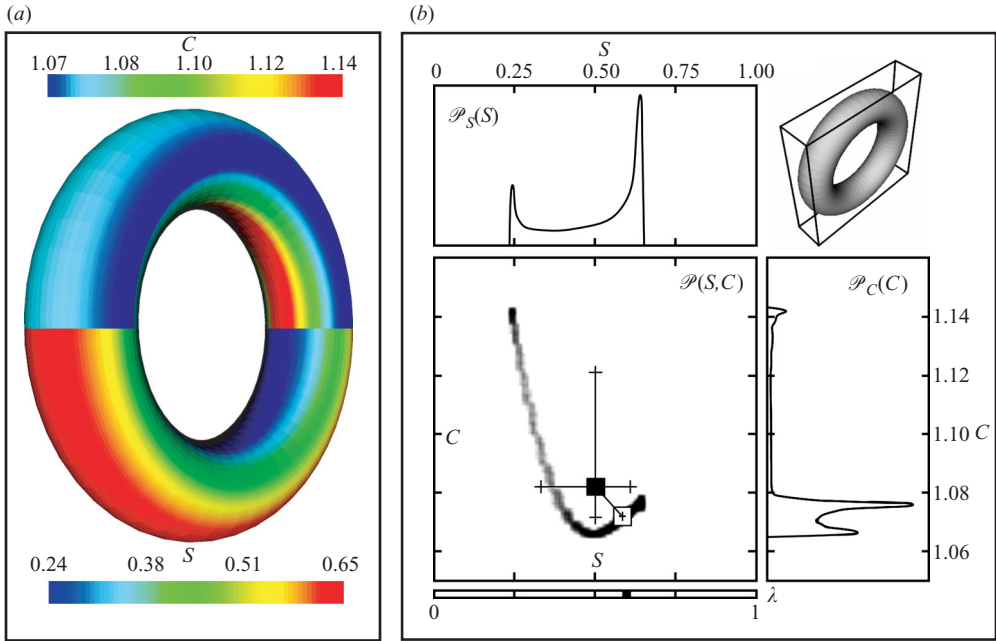
FIGURE 5. Example of a three-dimensional surface (*a*) with *S* and *C* mapped onto it (bottom and upper halves respectively) and its corresponding signature (*b*), for which a projection of the three-dimensional physical structure is shown at the top right corner, its joint two-dimensional probability density function (in terms of *S* and *C*) is presented in the bottom left area, while the marginal probability density functions of both *S* and *C*, are drawn at its top left and right, respectively. The value of the stretching parameter, λ, is represented below the joint p.d.f. by a black bar (in a scale from 0 to 1). Mean and feature centres, as well as upper and lower distances for each variable of the joint p.d.f., are superimposed to the joint p.d.f., as the filled and hollow squares respectively (refer to §4).

onto the (S, C)-plane with greyscale rendering of $\mathscr{P}$; white $\equiv 0$, black $\equiv \max(\mathscr{P})$. Additionally, we plot $\mathscr{P}_S(S)$ and $\mathscr{P}_C(C)$ on the *S* (top) and *C* (right) axes respectively; see figure 5 for an example. This geometrical characterization is based on properties of the structure that are invariant with respect to translations and rotations of the reference system, and therefore, are suited for comparing structures based on their geometry, the basis of the next step of this methodology: the classification of structures.

Several methods have been proposed in the computer graphics literature for estimating curvatures of a discretized surface (such as the ones that represent our structures in the computational domain). A subset of them, applied to the case in which the discretized surface is a triangular mesh, was here implemented and tested (Chen & Schmitt 1992; Dong & Wang 2005; Taubin 1995; Meyer *et al.* 2003). Finally, a modification of the algorithm proposed by Dong & Wang (2005) (based on Chen & Schmitt 1992) is used. The only modification is the way in which the normal vector to each face of the discretized surface is computed, following the method proposed by Chen & Wu (2004).

4. Classification of structures

A process of classification assigns different elements of a given set to groups based on the similarities of their signatures. In our system, the elements to be classified

are the educed structures, and the signatures are given by $\mathscr{P}(S, C), \mathscr{P}_S(S), \mathscr{P}_C(C)$ and λ, obtained in the characterization step for each structure. Among the different approaches to the problem of classification, we seek those involving as little *a priori* knowledge as possible of the relationships governing the different groups and of the number of groups present in the set of elements under evaluation. This leads to the utilization of learning-based clustering techniques. The idea behind this approach is to be able to detect other types of geometries apart from the known tube-like and sheet-like structures in turbulence data bases, should they exist, by not imposing strong assumptions on the groups.

4.1. *Clustering algorithm*

The clustering algorithm used in this classification step combines several techniques found in the data mining, pattern recognition and artificial intelligence literature (see, for example, Berkhin 2002 for a survey of such clustering techniques). It is a locally scaled spectral partitional clustering algorithm that automatically determines the number of clusters. Its main steps are summarized below using the notation proposed by Ng, Jordan & Weiss (2001) in their NJW algorithm, that conforms the core of our technique. Additions, particularizations and modifications to the NJW algorithm are also described below. In what follows we denote a set of N structures at a particular scale by the N elements $E = \{e_1, \ldots, e_N\}$. For each member of this set we construct a set of parameters $\{p[k], k = 1, \ldots, N_p\}$ which will serve as the contracted computational signature of the structure. These will be a finite set of moments of $\mathscr{P}(S, C), \mathscr{P}_S(S), \mathscr{P}_C(C)$, to be defined subsequently, together with λ. The $p[k]$ will also define a *feature space of parameters* in which the elements e_i are mapped. Typically $N = O(10^2 - 10^5)$, depending on the scale, and it will be seen that $N_p = 7$.

(*a*) Start from a set of N elements $E = \{e_1, \ldots, e_N\}$ and their corresponding contracted signatures $\{p_{e_i}[k], k = 1, \ldots, N_p\}$.

(*b*) Construct the *distance matrix*, $d_{ij} = d(e_i, e_j)$, $e_i, e_j \in E$. The element d_{ij} of the distance matrix measures dissimilarity between the two elements e_i and e_j of E, based on their signatures. Here we define the distance

$$d_{ij} = F(\{p_{e_i}[k] - p_{e_j}[k], k = 1, \ldots, N_p\}) \tag{4.1}$$

where $p_{e_i}[k]$ is the kth parameter associated with element e_l. The weighting function F defines a distance in that space of parameters. For example, a functional dependence of F of the form $F(\boldsymbol{x}) = (\sum_i x_i^2)^{1/2}$ defines a Euclidean distance in the feature space of parameters.

(*c*) Construct a *locally scaled affinity matrix* $\hat{\boldsymbol{A}} \in \mathbb{R}^{N \times N}$ defined by

$$\hat{A}_{ij} = \exp\left(-\frac{d_{ij}^2}{\sigma_i \sigma_j}\right) \tag{4.2}$$

where σ_l is a *local scaling parameter* introduced by Zelnik-Manor & Perona (2004) and defined as the distance of the element e_i to its Rth closest neighbour, denoted by $e_{R,i}$, $\sigma_i = d(e_i, e_{R,i})$ (a value of $R = 7$ is used, following Zelnik-Manor & Perona 2004). The purpose of introducing a local scaling parameter is to take into consideration the multiple scales that can occur in the clustering process, which is important, for example, when tight clusters are embedded within more sparse background clusters. Note that the elements of the diagonal of $\hat{\boldsymbol{A}}$ are null.

(*d*) Normalize $\hat{\boldsymbol{A}}$ with a diagonal matrix $\boldsymbol{D}$ such that $D_{ii} = \sum_{j=1}^{N} \hat{A}_{ij}$, obtaining the *normalized locally scaled affinity matrix* $\boldsymbol{L} = \boldsymbol{D}^{-1/2} \boldsymbol{A} \boldsymbol{D}^{-1/2}$

(*e*) For N_C varying between the minimum and maximum number of clusters considered, do the following loop:

(i) Find the N_C largest eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_C}$ of $\boldsymbol{L}$ and form the matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_C}] \in \mathbb{R}^{N \times N_C}$. This step constitutes the spectral part of the algorithm. It is intended to map the elements e_i onto a different eigenspace where clusters can be better identified. It can be considered a pre-clustering step, that combined with the local scaling explained in one of the previous steps allows clustering of elements with more complicated relationships among them (and to other clusters) than traditional clustering techniques that do not use these features. For example, concentric clusters can be easily educed by means of spectral clustering.

(ii) Re-normalize the rows of $\boldsymbol{X}$ so that they have unitary length, obtaining the matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times N_C}$ as $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)^{1/2}$.

(iii) Treat each row of $\boldsymbol{Y}$ as a point in $\mathbb{R}^{N_C}$ and cluster them into N_C clusters via K-means algorithm.

(iv) Assign the original element e_i to cluster k iff row i of $\boldsymbol{Y}$ was assigned to cluster k in the previous step.

(v) Obtain *optimality score* for this number of clusters N_C (see §4.3).

(*f*) After the previous step has been done for all the possible numbers of clusters under evaluation, determine the optimum number of clusters based on the minimization of the optimality score for each one of the possible numbers of clusters (as will be described in §4.3).

The K-means clustering algorithm mentioned above is one of the simplest partitional clustering techniques available. It first initializes the cluster centres (for a given number of them). Then it assigns each element to the cluster with the closest centroid to that element. After all elements have been assigned, it recalculates the position of the cluster centres. The last two steps are repeated until the cluster centres no longer move. Different implementations of the K-means clustering algorithm differ mainly in the initialization of the cluster centres: we choose the initial position of the first cluster centre randomly among all the elements; initial positions of subsequent cluster centres are obtained as the farthest elements to the previously assigned cluster centres.

4.2. *Feature and visualization spaces; definition of the* $p[k], k = 1, \ldots, N_p$

The selection of the $p[k]$ used to define the *feature space* plays a decisive role in the classification step. Each structure will be represented by a point in that feature space and its distance to the other points will define the similarity to their corresponding structures. The number of parameters (dimensions of feature space) should be sufficiently large to distinguish satisfactorily relevant groups of structures, but at the same time, it should be kept as small as possible to avoid the so-called 'curse of dimensionality' (see Bellman 1961) that affects unsupervised learning algorithms, like the clustering method used in this methodology, compromising its success by making the points too disperse in such high-dimensional space. The set of (seven) parameters chosen here for each element e_i of E is

$$\{p[k], k = 1, \ldots, 7\} \equiv \left\{ \hat{S}, \hat{C}, \lambda, d_u^S, d_l^S, d_u^C, d_l^C \right\} \tag{4.3}$$

where $\hat{S}, \hat{C}$ denotes the feature centre of $\mathscr{P}(S, C)$ and $d_u^S, d_l^S, d_u^C, d_l^C$ are the upper and lower distances of the joint p.d.f. in each variable. The feature centre $(\hat{S}, \hat{C})$ takes into account the possible asymmetry of the joint p.d.f., correcting the mean centre $(\bar{S}, \bar{C})$ so that the feature centre lies closer to the region of higher density of the joint p.d.f.. The upper and lower distances, d_u and d_l, can be regarded as the r.m.s. of the part
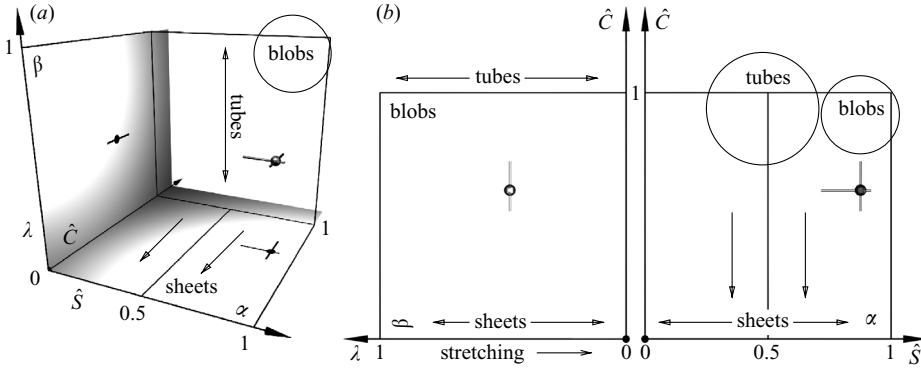
FIGURE 6. Projections of the *visualization space* with the predominantly blob-, tube- and sheet-like regions sketched: three-dimensional perspective projection (*a*), two-dimensional orthogonal projections (*b*) of the planes β (formed by the axes $\hat{C}$ and λ) and α (formed by $\hat{S}$ and $\hat{C}$). An example of a glyph consisting of a sphere and four bars along the $\pm\hat{S}$, $\pm\hat{C}$ axes can represent nine parameters of the characterization of the corresponding structure: $\hat{S}, \hat{C}, \lambda$ given by the centre of the sphere, upper and lower distances of S and C given by each bar, the surface area A of the associated structure, given by the size of the glyph, and the group to which the structure belongs, given by the colour of the glyph.

of the p.d.f. above and below, respectively, its mean value. A graphical example can be seen in figure 5, where the mean and feature centres have been superimposed to its corresponding joint p.d.f.. Definitions of feature centre, upper and lower distances, together with a representative one-dimensional example can be found in Appendix C.

Based on the idea of the *feature space* of parameters used for educing clusters of similar structures, we define a *visualization space*, intended to provide a graphical representation of the distribution of individual structures in a three-dimensional space, providing qualitative and quantitative information. In general, the higher-dimensional character of the feature space prevents its use as visualization space, but the utilization of *glyphs*, scaling and colouring allows more than just three dimensions to be represented in the visualization space.

We define the three axes of the visualization space by $\hat{S}$, $\hat{C}$ and λ. Owing to the choice of non-dimensionalization of the curvedness and the normalization factors (see §3), as well as the intrinsic meaning of the shape index, curvedness and stretching parameter, it is possible to identify regions in the visualization space with a particular geometrical meaning for those structures whose representation lies in them. For example, blob-like structures occupy the region near the point $(1, 1, 1)$ (which corresponds to spheres); tube-like structures are localized near the $(1/2, 1, \lambda)$† axis (λ being an indication of how stretched the tube is) and the transition to sheet-like structures occurs as the curvedness and λ decrease. The plane $\hat{C} = 0$ is the limiting case of planar structure; furthermore, any structure composed of (predominantly) planar regions, thus featureless in the curvature sense, will have a (nearly) nil $\hat{C}$, independently of its relative aspect ratios, that will nevertheless affect its λ value. See Appendix D for an analysis of these limiting values. Throughout this paper, the visualization space is presented by a set of two-dimensional projections (see figure 6 for an example).

† Note that a value of the shape index equal to $1/2$ corresponds to locally cylindrical shapes, that are predominant in tube-like structures. The dimensionless curvedness of a straight elongated circular cylinder of radius R reduces to $C = 3V/A\sqrt{2}R \approx 3/2\sqrt{2} \approx 1.06$.

4.3. *Optimality score: silhouette coefficient*

The determination of the optimum number of clusters is based on the minimization of an *optimality score*. Different approaches have been considered. Among them, probabilistic criteria that consider the relative increment of complexity of a model (set of clusters) when another parameter (cluster) is added, such as the *Bayesian Information Criterion* (Schwarz 1978), were found to provide unsatisfactory results. This is mainly due to the use of spectral techniques, since they map the elements to be clustered onto a different eigenspace whose dimensions change with the number of clusters considered, complicating the task of comparing the goodness-of-fit for different number of clusters by such probabilistic methods. Instead, the *silhouette coefficient* (Rousseeuw 1987), SC, is used. It is a confidence indicator of the membership of an element to the cluster it was assigned. It is defined, for each element e_i, as $SC_i = (b_i - a_i)/\max(a_i, b_i)$, where a_i is the average distance between element e_i and other elements in its cluster, and b_i is the average distance to the items in the closest cluster. It varies from -1 (lowest membership) to $+1$ (highest membership). Being a dimensionless quantity, the mean and variance throughout all the clustered elements can be used as indicators of the optimality of the clustering, and compared among results for different numbers of clusters to determine the optimum number of them. High values of the mean silhouette indicate a high degree of membership of the elements being clustered to the clusters they were assigned, and low values of its variance indicate that the majority of elements have a similar value of the silhouette. The combination of both indications reflects a successful clustering.

Once the cluster centres have been obtained, it is also possible to retrieve the closest elements to those cluster centres among the elements being classified. These closest elements to the cluster centres can be considered as representative elements of each cluster.

5. Application to a virtual set of structures

The validity and applicability of the last two steps (characterization and classification) of the proposed methodology were tested on a virtual set of nearly 200 surfaces created using computer three-dimensional modelling tools. The extraction step was not included in this test, since the starting point is the set of surfaces itself. Nevertheless, modelled structures of very different sizes were included, to emulate the multiple scales that would result from the extraction step, had it been included. Also, the shapes of the modelled structures are all different. They could be visually classified into three main groups with a common geometry: blobs, tubes and sheets. The target of this test was to educe those three main groups automatically and without any *a priori* knowledge of the possible geometries of the structures present in the dataset or of the number of groups among them, that is, simply based on the characterization and classification steps of the methodology previously explained. Among the modelled sheet-like structures, approximately one third were given a certain rolling geometry (spiral-like sheets).

Figure 7 shows the visualization space with the results of the test. Each sphere in that space represents a structure of the virtual set (some examples are projected onto the planar sides). The colour of each sphere in the visualization space indicates the cluster to which its corresponding structure has been automatically assigned by the clustering algorithm during the classification step of the methodology, and its
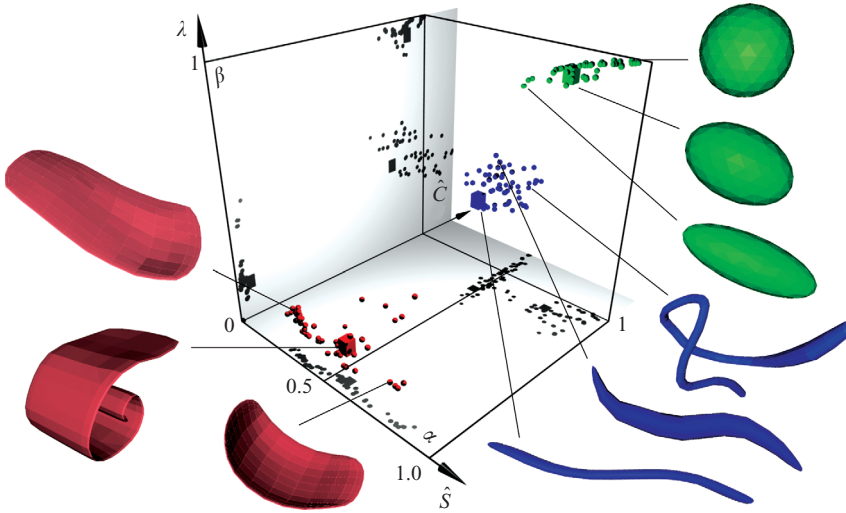
FIGURE 7. Visualization space with clustering results for the virtual set of modelled structures, with representative examples shown at the sides.

diameter is scaled using its associated silhouette coefficient, that represents, for each structure, the degree of membership to the cluster to which it was assigned (refer to §4.3), renormalized to have positive values that allow a comparison among structures. For reference, the closest elements to the cluster centres have been highlighted using cubes of slightly bigger size.

Three clusters were automatically educed in the classification step and each structure was 'correctly' assigned by the algorithm to the appropriate group corresponding to the previously constructed geometry. This can be seen in figure 7 from the relative locations of the centres of the glyphs defined by $\hat{S}$, $\hat{C}$ and λ for each structure. We emphasize that neither information on the previously constructed shapes nor the number of groups to be educed formed any part of the clustering algorithm (for example as pre-conditioning). The results of figure 7 are a consequence of the geometric characterization and automatic classification in the feature space of parameters.

We note also that in figure 7 (as was sketched in figure 6), the sheet-like structures can spread over a large region near the plane $\hat{C} = 0$ in the visualization space. This region could be narrowed by means of a transformation of the $(\hat{S}, \hat{C})$-plane to Cartesian coordinates ($\hat{X} = \hat{C} \cos[\pi(\hat{S}-1/2)]$, $\hat{Y} = \hat{C} \sin[\pi(\hat{S}-1/2)]$). This would bring sheet-like structures to the axis $(0, 0, \lambda)$ in the new visualization space. Nevertheless, it is helpful to keep the original visualization and feature spaces, since that allows a possible distinction of the different shapes of the structures that fall into the broadly defined sheet-like geometry. For example, in the test case of the virtual set of modelled structures presented here, the second optimum automatic clustering result was such that four groups were educed: the blob-like and tube-like clusters remained the same as in the optimum case of three clusters described above, but the sheet-like cluster was split into two, with one of these clusters containing a large proportion of structures with a rolling geometry (spiral-like sheets). This also suggests that further post-processing (ideally also automatic) of the educed clusters can be helpful in refining the results.
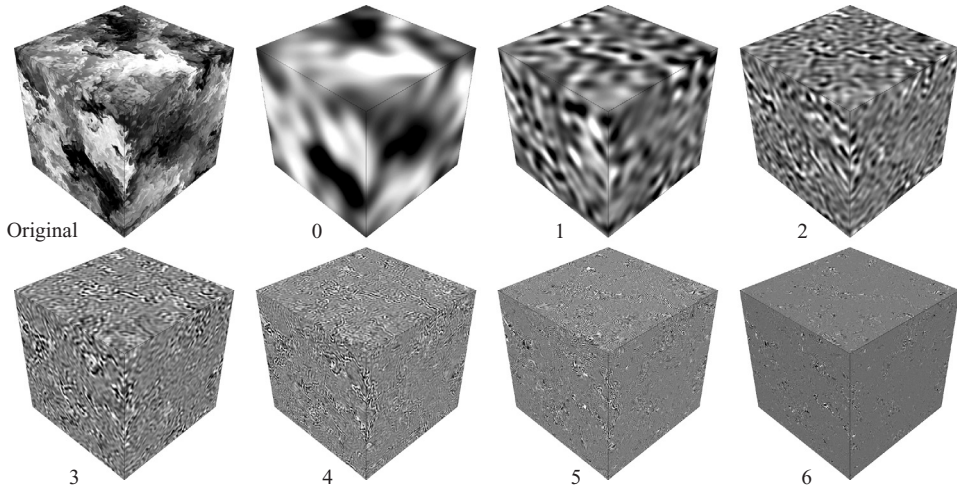
FIGURE 8. Cubic plane cuts of the passive scalar fluctuation field for the original database (top left) and each of the filtered scales resulting from the multi-scale analysis (filtering in curvelet domain) (increasing scale number from left to right and top to bottom).

6. Application to the passive scalar fluctuation field of a numerical turbulence database

6.1. *DNS database*

We use a numerical database obtained from a DNS with 512^3 grid points. The incompressible Navier–Stokes equations for the velocity field and the advection-diffusion equation for the passive scalar were solved by means of a Fourier–Galerkin pseudo-spectral method. The domain is a cube of side 2π with periodic boundary conditions. The velocity field was forced at large scales, becoming statistically stationary in time. A mean scalar gradient was imposed so that the scalar fluctuation field became also statistically stationary in time. Despite the mean scalar gradient applied, the scalar fluctuation is statistically homogeneous. The Reynolds number based on the integral length scale is 1901, whereas the Taylor Reynolds number is $Re_\lambda = 265$. The Schmidt number of the simulation is 0.7. The product of the largest dynamically significant wavenumber, k_{max}, and the Kolmogorov length, η, is $k_{max}\eta = 1.05$. More specific details of the database can be found in O'Gorman & Pullin (2004).

6.2. *Multi-scale diagnostics*

We apply our methodology to the passive scalar field given by the passive scalar fluctuation at an instant in time. For the given resolution of 512^3 grid points, and a coarsest scale $j_0 = 2$, the curvelet transform provides seven scales. They will be named by a scale number, from 0 to 6; increasing values of the scale number correspond to smaller scales. Thus, scale 0 captures the largest scales and 6 the smallest. Figures 8 and 9 show the result of the multi-scale analysis based on the curvelet transform. Plane cuts of the original database and each of the filtered scales (filtered in the curvelet domain and then inverse transformed to the physical domain) are shown. Three-dimensional views with plane cuts in the three directions of the volume data are presented in figure 8, and more detailed plane cuts normal to the z-direction at half the length of the cube are shown in figure 9. Volume-data p.d.f.s obtained for the scalar field associated with the original database and for each of the filtered scales,
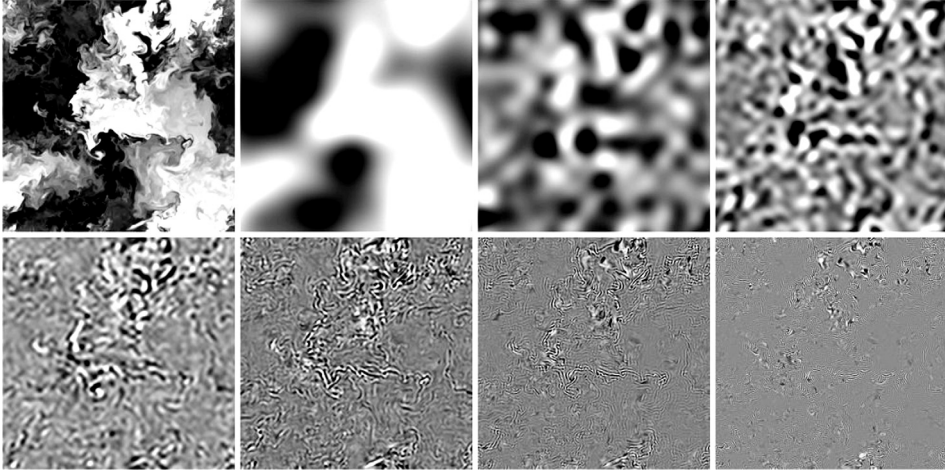
FIGURE 9. Plane cuts normal to the *z*-axis at its midpoint of the passive scalar fluctuation field for the original database (top left) and each of the filtered scales resulting from the multi-scale analysis (filtering in curvelet domain) (increasing scale number from left to right and top to bottom).
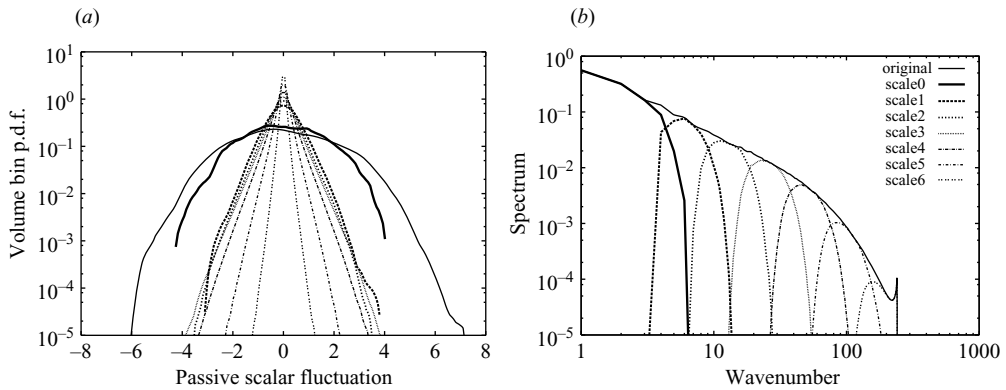

FIGURE 10. Volume-data p.d.f.s of the passive scalar fluctuation field (*a*) and corresponding spectra (*b*), associated with the original database and each of the filtered scales.

figure 10(*a*), give insight into the distribution of the scalar values at the different scales and their contribution to the total field (original data base). For this scalar field, the p.d.f.s tend to become narrower for increasing values of the scale number, that is, for smaller scales. Scalar fluctuation spectra are also computed for the original volume data field and each filtered scale and are shown in figure 10(*b*). The effect of the curvelet filtering in the Fourier domain can be observed. This differs from a top-hat window filtering in that domain, in order to preserve the localization in the physical domain. It can be noticed that scales 1, 2 and 3 correspond mainly to the inertial range of scales, whereas scales 4 and 5 are mainly dissipation scales. From this observation and from figure 10(*a*) we note that those p.d.f.s associated with the scales corresponding to the inertial range (1, 2 and 3) are very similar, almost collapsing in that plot.

Additionally, an equivalent multi-scale decomposition is done for the velocity field. That allows us to define characteristic squared integral velocities, $\overline{u_i^2}$, and integral

Scale	$\overline{u_i^2}/\overline{u^2}$	L_i/η	L_i'/η
original	1.000	249.6	249.6
0	0.591	226.9	383.8
1	0.155	14.68	96.1
2	0.113	5.235	46.2
3	0.085	1.927	22.8
4	0.044	0.519	11.9
5	0.011	0.070	6.3
6	0.001	0.004	3.3

TABLE 1. Breakdown of characteristic integral velocities and length scales
for the filtered scales.

length scales, L_i and L_i', for each filtered scale, i, in the same terms in which they are defined for the original velocity field. For the case of isotropic turbulence, they can be expressed as

$$\overline{u_i^2} = \frac{2}{3} \int_0^\infty E_i(k)\,\mathrm{d}k, \tag{6.1}$$

$$L_i = \frac{\pi}{2\overline{u^2}} \int_0^\infty \frac{E_i(k)}{k}\,\mathrm{d}k, \tag{6.2}$$

$$L_i' = \frac{\pi}{2\overline{u_i^2}} \int_0^\infty \frac{E_i(k)}{k}\,\mathrm{d}k, \tag{6.3}$$

where $E_i(k)$ is the energy spectrum associated with scale i, and the absence of a subindex refers to the original velocity field. It follows from (2.5) that the sums of the energy spectra and the characteristic squared integral velocities of the all filtered scales are equal to those of the original velocity field, $E(k)$ and $\overline{u^2}$ respectively:

$$E(k) = \sum_i E_i(k), \tag{6.4}$$

$$\overline{u^2} = \sum_i \overline{u_i^2}. \tag{6.5}$$

Table 1 shows, for the original velocity field and for each filtered scale, the characteristic squared integral velocity and the characteristic integral length scale, and how they compare to the total characteristic squared integral velocity, $\overline{u^2}$, and the Kolmogorov length scale, η.

6.3. *Geometry of passive scalar iso-surfaces*

After the multi-scale analysis, iso-surfaces are obtained for each of the filtered scales. The contour values are, for each filtered scale, the mean value of the scalar field plus two times the standard deviation of that field (mean and standard deviation values can be obtained from the first- and second-order moments of the volume p.d.f.s presented before) (see figure 11). Those iso-surfaces corresponding to the same relative contour value at each scale will be characterized and classified and their results compared among the different scales. We also remark that an additional step in the extraction is applied to periodically reconnect those structures intersecting boundaries with their continuation on the opposite boundaries; this reconnection is performed for each individual filtered scale. Both the largest scale (0) and the smallest scale (6) are not
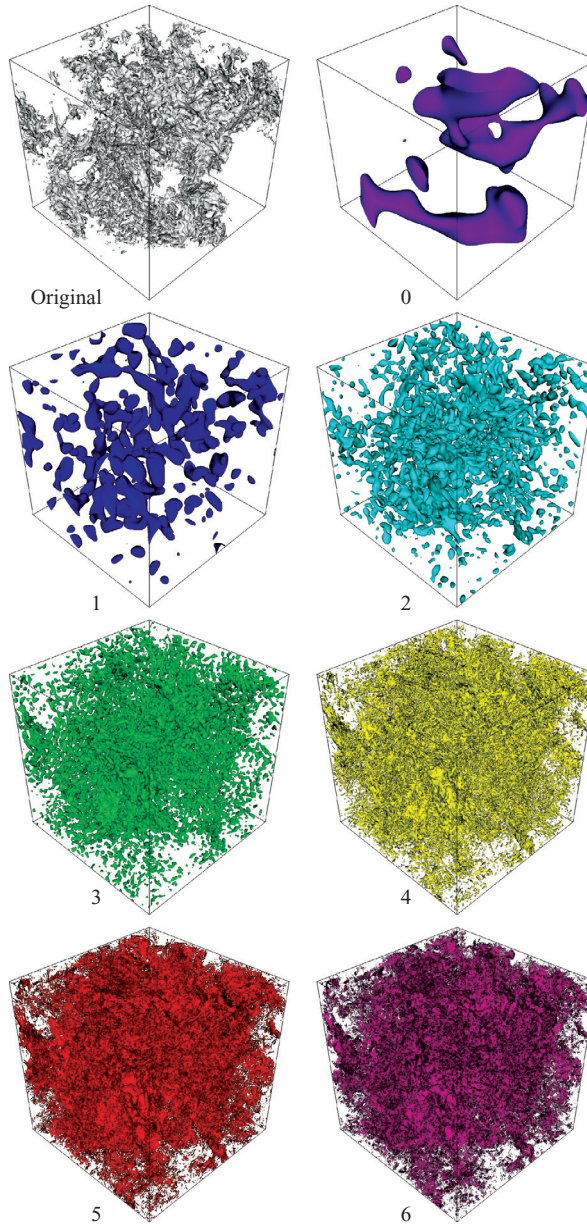
FIGURE 11. Iso-contours of the passive scalar fluctuation field for the original database (top left) and each of the filtered scales resulting from the multi-scale analysis (filtering in the curvelet domain) (increasing scale number from left to right and top to bottom). A contour value equal to the mean plus two times the standard deviation of each resulting scalar field was used.

considered in the process: the largest scale is of less relevance in this analysis since its structure is expected to depend on the boundary conditions and external forces applied. The smallest scale is excluded to avoid the extraction of spurious structures and/or an erroneous geometrical characterization that could result from the lack of grid resolution, or aliasing effects at that scale. Thus, the scale numbers under analysis
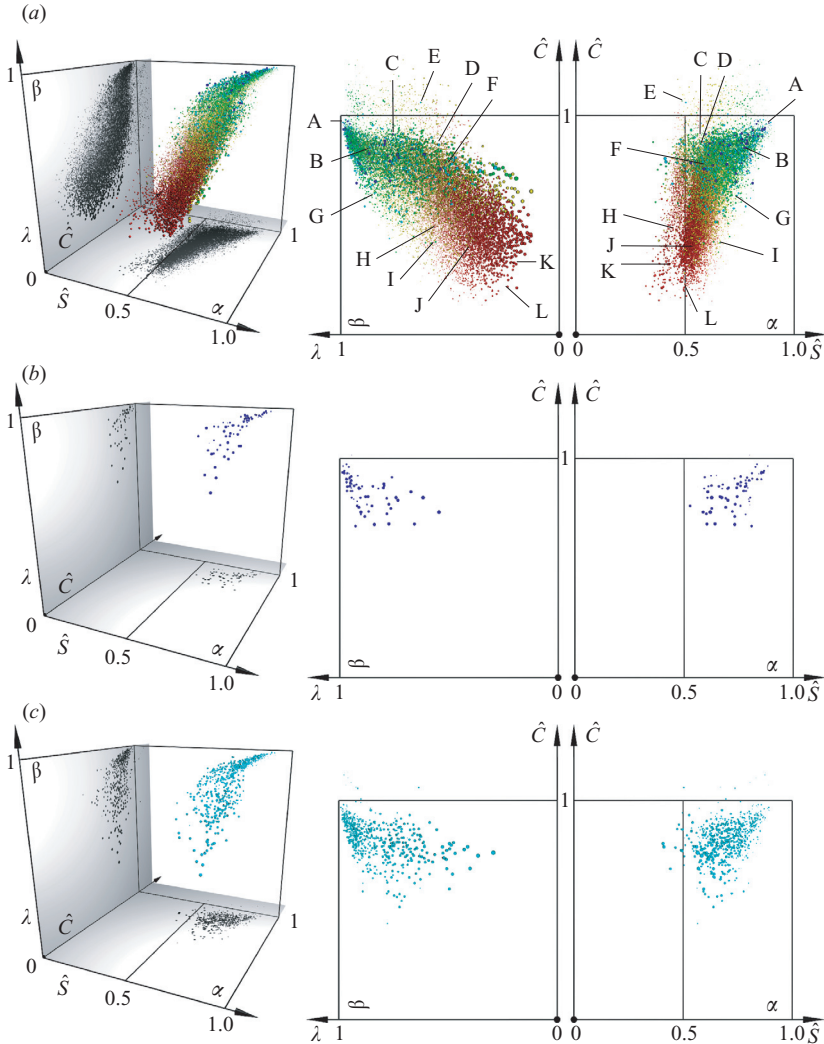
FIGURE 12(*a–c*). For caption see facing page.

are 1–5. In the same spirit, a minimum number of points (300) was considered for a structure to be analysed, so that it is smooth enough for a reliable calculation of its differential-geometry properties, the basis of the characterization step.

Then, each structure is geometrically characterized as described in §3 and, based on the parameters extracted from its signature, it can be represented in the visualization space referred in §4. Figure 12 shows the distribution of glyphs representing each structure, for the different scale numbers 1–5 considered. In this case, glyphs are spheres whose centres correspond to the $\hat{S}, \hat{C}, \lambda$ parameters and whose radii are scaled according to the surface area of the structure, and their colour is assigned based on the scale number to which the structure belongs. First, structures of all scales are shown and then the progression for individual scales is presented. As can be seen, the structures go from predominantly blob-like and tube-like at scales 1–3 toward more sheet-like structures at the smaller scales 4 and 5, The stretching of
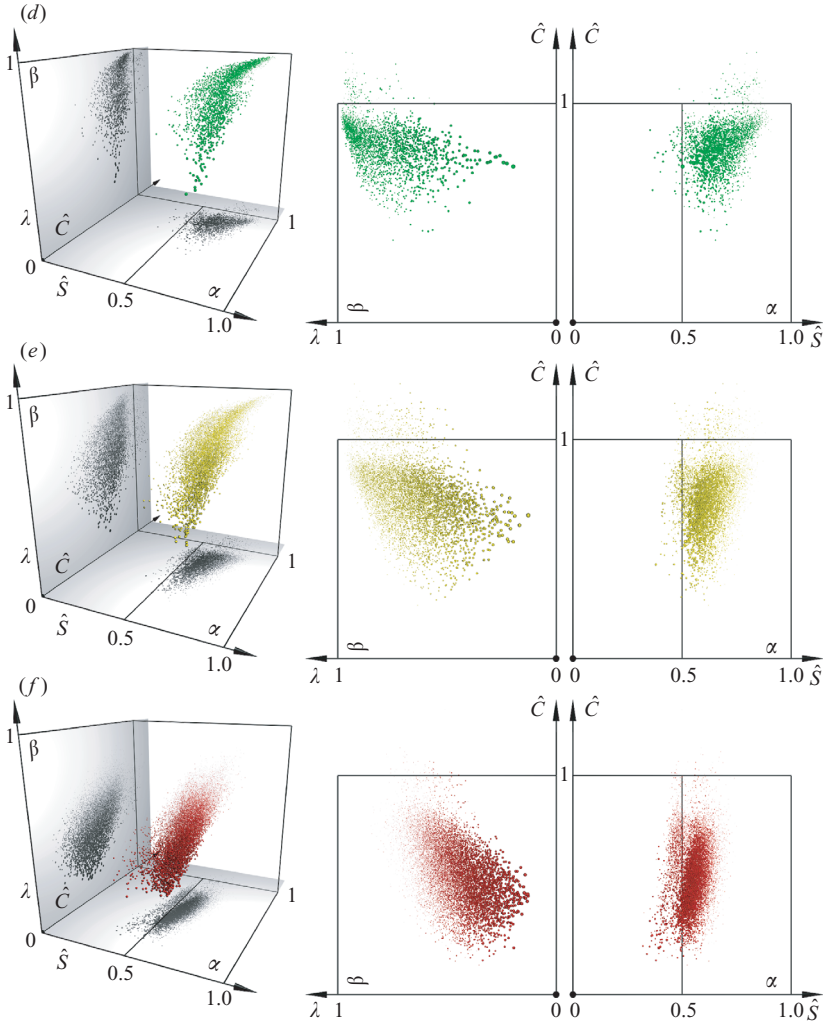
FIGURE 12. Visualization space with spheres representing the structures educed from the passive scalar fluctuation field at all scales (*a*) and scale 1 (*b*), scale 2 (*c*), scale 3 (*d*), scale 4 (*e*) and scale 5 (*f*). Radii of spheres represents (in normalized logarithmic scale) the surface area of each structure. Colour of the spheres represents the scale to which the structure belongs: dark blue (1), light blue (2), green (3), yellow (4), red (5).

the structures increases with the scale number, that is, the parameter λ decreases for smaller and smaller scales.

Some representative structures, named A–L, have been selected (see figure 12*a*); their corresponding signatures are shown in figure 13. The cascade in the passive scalar fluctuation spectrum is thus translated into a cascade of the representation of structures in the visualization space, that starts near the point $(1, 1, 1)$ (sphere) and evolves toward highly stretched sheet-like structures whose geometry tends to be complex (see for example their corresponding signatures in the last few points in figure 13).

The clustering algorithm is then applied to the structures. The set of parameters $\{\hat{S}, \hat{C}, \lambda, d_l^S, d_u^S, d_l^C, d_u^C\}$ is used to form the feature space where each structure is
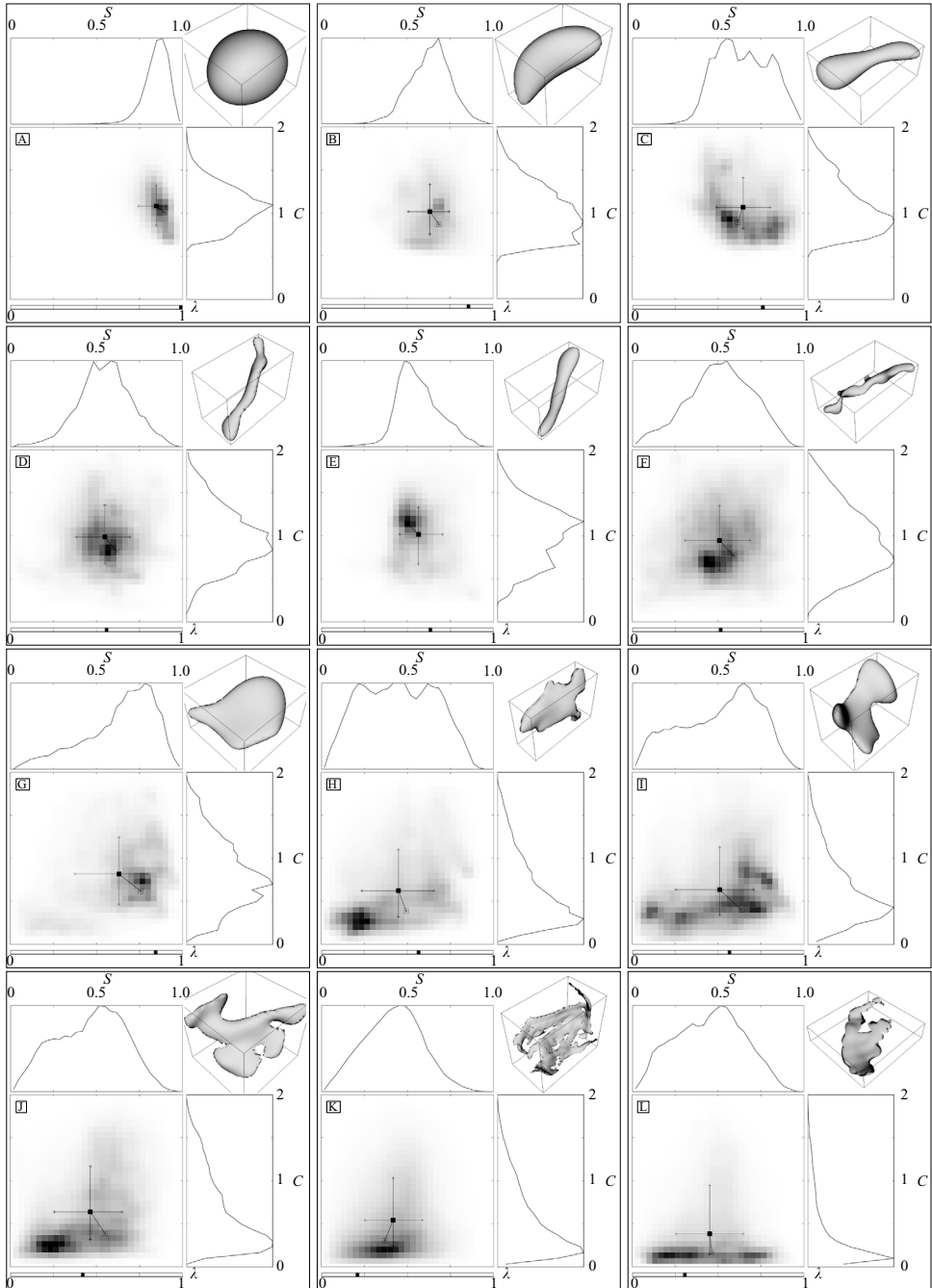
FIGURE 13. Signatures of representative structures (refer to figure 12(*a*) for the location of the corresponding points (A–L) in the visualization space).

represented by a point. Three groups of structures are obtained, and the result can be seen in figure 14. That figure shows a visualization space with the same three spatial coordinates as in earlier plots ($\hat{S}, \hat{C}, \lambda$), and the structures represented by glyphs consisting of spheres (coloured by the cluster ID and with radius scaled by
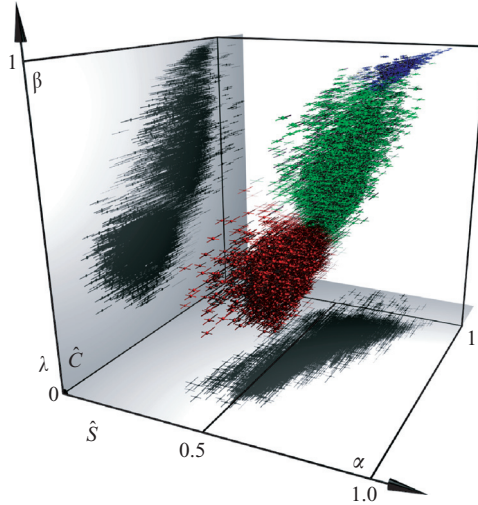
FIGURE 14. Result of the clustering algorithm in a visualization space showing spheres representing structures with radii scaled by their silhouette value (renormalized to have only positive values) and coloured by the cluster to which they belong. An optimum number of three clusters was automatically found. The seven clustering parameters used to define the feature space are shown for each structure by the centre of its representing sphere (of coordinates $\hat{S}$, $\hat{C}$, λ) and the four bars scaled by the value of d_u, d_l of S and C.

the silhouette coefficient, defined in §4.3, that indicates the level of membership to the cluster to which it has been assigned) and horizontal bars with origin at the centre of the sphere and lengths proportional to the other four parameters used for clustering (distances d_u and d_l in $\pm S$ and $\pm C$ directions). The thickness of these bars is also scaled by the silhouette value. Although the clustering algorithm captures the main trends, the structure geometries appear continuously distributed across the main groups, rather than separating into well-differentiated groups. This translates into the distribution of points and glyphs representing structures in the feature and visualization spaces: in particular, glyphs associated with the educed structures are organized as a cloud in the visualization space, transitioning from one region to other regions. For example, a comparison of the clustering results for the passive scalar field (figure 14) and the test case of modelled structures (figure 7) previously presented in §5 clearly shows the difference between the continuously distributed geometries of the structures educed for the passive scalar field and the well-distinct groups of geometries found in the test case. As a result, for the case of the passive scalar, some of the structures classified as belonging to one group but lying on the overlapping regions in the feature space will not necessarily be significantly different (geometrically) from other structures that belong to other groups but with a similar location in the feature space. The degree of membership to the educed clusters (measured by the silhouette coefficient) of those structures in the overlapping regions between clusters will therefore be lower than that of structures near the cluster centres. Note how the glyphs in figure 14, scaled by the renormalized silhouette coefficient of the associated structures, are smaller in the overlapping regions (compare for example, figure 12(*a*), where the density of points is much more homogeneous throughout the whole distribution, since the scaling factor in that case was the area of the structure, not its silhouette coefficient).

6.4. *Discussion and physical interpretation*

We discuss first the smallest scales. Figure 12(*f*) shows that highly stretched sheet-like structures are predominant at those scales. From the size of the spheres representing the structures (related to the surface area of the structures in a normalized logarithmic scale), we conclude that the larger of these structures appear more stretched (lower λ) and tend to be more sheet-like (lower $\hat{C}$). This trend is in agreement with the results of Schumacher & Sreenivasan (2005), who, using a conventional box-counting method, found that passive scalar isolevel sets at the smallest scales become smooth sheets. They did not find fractal structure in the passive scalar field within their range of Sc and Re_λ, which differ from present values. Schumacher & Sreenivasan also used the area-to-volume ratio of the isolevels, which is essentially the parameter μ used in our methodology to form the dimensionless curvedness $\hat{C}$: they use this in a global sense rather than applied to individual structures as is done here. In the study of intense strain structures in homogeneous isotropic turbulence by Moisy & Jiménez (2004), the dominance of sheet-like structures in the smallest scales, suggested by a similar box-counting method, was confirmed by the geometrical study of aspect ratios of individual structures.

Multi-scale decompositions of the vorticity field in turbulent flows have been previously applied using orthogonal wavelets in two and three dimensions (see Farge *et al.* 1999, 2001, 2003), where thresholding of the wavelet coefficients based on denoising theory separates the vorticity into two orthogonal fields, denoted as *coherent* and *incoherent* vorticity. It is found that the coherent field is responsible for most of the energy transfer in the large and inertial scales. In three-dimensional homogeneous isotropic turbulence, the p.d.f. of the coherent vorticity is found to be stretched exponential while the incoherent vorticity is exponential. We find that the p.d.f. of the scalar fluctuation (see figure 10) is Gaussian, with sub-Gaussian tails, in agreement with previous results for scalar fields (Overholt & Pope 1996; Celani *et al.* 2001). In our multi-scale decomposition (see figure 10), the variance of the p.d.f.s also decreases for smaller scales (resulting in narrower p.d.f.s), quickly transitioning from Gaussian (with slightly sub-Gaussian tails) to exponential. Our multi-scale analysis does not include any assumptions about the 'coherence' of the educed structures through thresholding of the multi-scale coefficients, since it has a different purpose, mainly as a diagnostics tool.

It has been suggested that exponential tails of the scalar fluctuation p.d.f.s are linked to metrics of anomalous mixing (see the discussion in Warhaft 2000). There have been attempts to clarify the conditions under which sub-Gaussian/exponential tails appear based on various factors that include relative simulation box size (Overholt & Pope 1996; Schumacher & Sreenivasan 2005), Reynolds number (distinguishing between soft and hard turbulence (Jayesh & Warhaft 1992)), flow initial conditions and forcing (Jaberi *et al.* 1996) and, for the vorticity, the structure of intense portions of the field (Siggia 1981; Kerr 1985; Jiménez *et al.* 1993). The transition, seen in figure 10, from Gaussian p.d.f. in the larger scales to predominantly exponential p.d.f.s in the smaller scales may be related to the geometry of individual structures present at each scale.

The presence of ramp–cliff structures in the scalar field (plateau–cliff in the scalar fluctuation) (Antonia *et al.* 1979) has been associated with anisotropy of the passive scalar field in the presence of a mean gradient (Celani *et al.* 2001; Overholt & Pope 1996; Warhaft 2000). These features are seen in the large scales (plateau regions) as well as in the small scales (cliffs of fronts), where steep changes in the values of the passive scalar occur (see figure 9). The highly stretched, sheet-like structures found

here at the smaller scales could be related to these fronts. A study of their spatial distribution with respect to proximity to the structures of the larger scales would be needed to confirm this. Further, the predominant orientation of the sheet-like regions of such structures could help to clarify their role. This could be obtained within the framework of the present methodology using the multi-orientation decomposition of the curvelet transform (not applied in this paper).

For the present *Sc* and *Re* we know of no previous reports of blob- and tube-like (with moderate stretching) structures in the intermediate scales of the passive scalar fluctuation field. Theoretical developments in physical models of passive scalar mixing have utilized tube-like structures, stretched by large-scale strain fields, to analyse cascade and dissipation dynamics for a passive scalar (Pullin & Lundgren 2001). Small-scale scalar mixing is modelled as a two-dimensional blob (a tube in three-dimensions) convecting, deforming and diffusing in the presence of the swirling motion of a stretched spiral vortex. The blob is drawn out into rolled-up sheets whose azimuthally averaged structure remains tube-like. Within this compound tube-sheet structure, the derived scalar spectrum comprises two parts in the form of Batchelor (1959) k^{-1} and Obukov–Corrsin $k^{-5/3}$ (see Tennekes & Lumley (1974)) components. The $k^{-5/3}$ contribution arises from the non-axisymmetric scalar field which tends to be sheet-like. This dynamical model is not inconsistent with the present findings of tube and sheet structures at the smallest scales. We can hypothesize that blob-like structures, similar to structure A shown in figure 13, are created first. These are then strained and stretched by the action of vortex tubes (D, E, F) to form vortex sheets (J, K, L). Further vortex tubes are then created by rolling-up of the sheets. Additional support for the validity of this picture as a physical mechanism of the cascade would probably require (at least) local correlation in tube and sheet structure locations and perhaps orientations, for adjacent scales in the sense of the curvelet transform. This is beyond the scope of the present study.

7. Conclusions

A methodology for the identification of structures based on their geometry, and its present application to a numerical database of the mixing of a passive scalar in homogeneous isotropic turbulence, have been presented. Our goal has been to develop a methodology that can compensate for the computational bottleneck of DNS computing for turbulent flows, and to provide a solid mathematical framework for non-local characterization of the flow structures based on existing data sets. The main characteristics of this methodology, in comparison with previously existing ones, are its multi-scale and non-local character. The multi-scale nature, implemented by means of the curvelet transform, provides the framework for studying the evolution of the structures associated to the main ranges of scales defined in Fourier space, while keeping the localization in physical space that enables a geometrical study of such structures. We note that the multi-orientation decomposition included in the curvelet transform, not used in this study, can be useful when analysing other flows in which the directionality of the structures can play a significant role, such as channel flow. The non-local character of the methodology is achieved through the calculation of area-based probability functions of the differential-geometry properties of the surface under consideration. It is also a generic methodology, not intended to educe a particular kind of geometry, but able to manage and classify all possible geometries. There are three main steps involved: extraction, characterization and classification

of structures. Individual structures (considered as closed surfaces disconnected from each other) are studied.

From its application to the passive scalar fluctuation field advected and diffused in homogeneous isotropic turbulence, the following conclusions can be drawn. First, the multi-scale decomposition resulted in a set of scalar fields (associated with the different ranges of scales extracted) with volumetric probability density functions of decreasing width for smaller scales. In addition, those probability density functions corresponding to scales approximately in the inertial range tend to overlap. Secondly, the study of the structures educed for the different scales shows a transition of their geometry from predominantly the blob-like and tube-like kind in the inertial range of scales toward sheet-like structures in the dissipation range. The dominant structures become more and more stretched for the smaller scales. This transition of geometry is smooth, complicating the automatic classification of structures. There are not clearly distinct groups of structures with a common geometry, but a continuous distribution of them filling the spectrum of present geometries instead. Thus, the application of the clustering algorithms is more challenging. In this case, three groups were educed automatically by applying the clustering technique implemented, and their projection in the visualization space and the identified cluster centres agree with the comments stated above. Nevertheless, clustering results are to be used with care in these conditions in which the points are so continuously distributed in the feature space used for clustering. Alternatives to the currently implemented K-means clustering algorithm, such as *fuzzy c-means* clustering or *density-based* clustering, more oriented toward educing intermingled clusters without clear boundaries can be considered. Also, the addition of other relevant parameters in the clustering process may be useful to allow more separation.

Future work could include the application of this technique to other scalar fields, such as the vorticity magnitude and other local Galilean invariants derived from the velocity-gradient tensor. The study of the relation of the passive scalar structures to these and other fields such as scalar dissipation, local enstrophy and local mechanical dissipation may illuminate local dynamical processes. Applications in the tracking and evolution in time of individual Lagrangian structures and their developmental geometry may also be useful.

The modular character of the implementation of the present methodology, a consequence of its conceptual division into the three main steps of extraction, characterization and classification, should facilitate future algorithmic improvements corresponding to each step. Also, the addition of other modules, such as a time tracker for studying the temporal evolution of the geometry of structures, can be achieved by adding a new layer on top of the existent implementation, When combined with the visualization space already presented in this paper, this may facilitate the search for potential geometrical attractors in a suitable feature space.

The requirements of our implementation and application of this methodology to the 512^3 passive scalar database do not exceed the computational resources offered by a normal desktop or laptop computer.
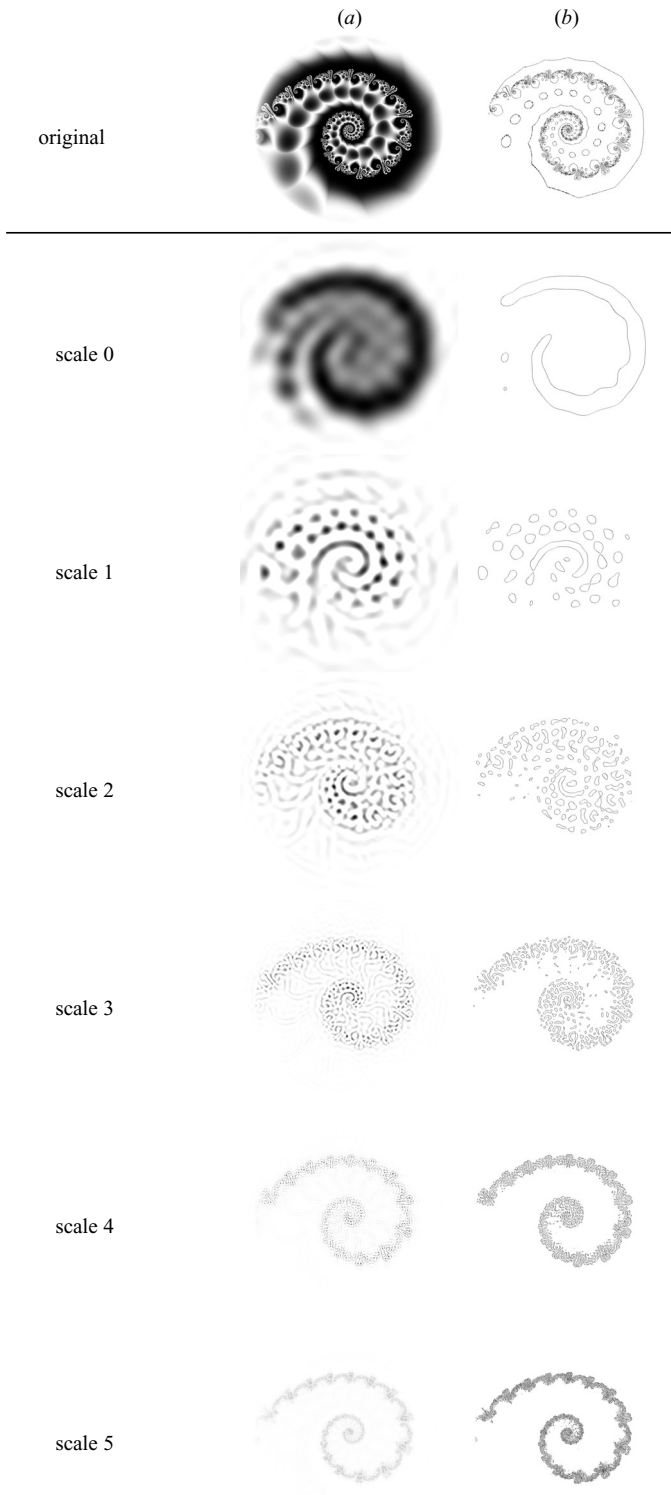
FIGURE 15. Fields (*a*) and corresponding iso-contours (*b*) for original and filtered scales.
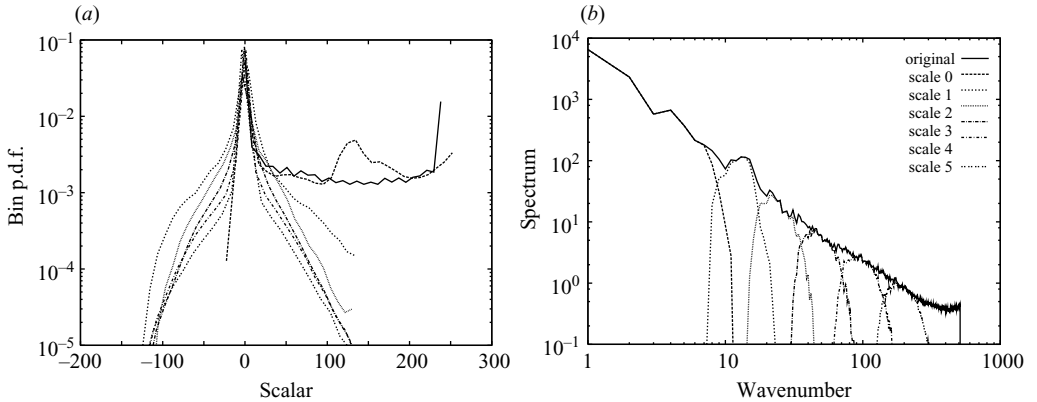
FIGURE 16. P.d.f.s (*a*) and spectra (*b*) associated with the original and each
of the filtered scales

Appendix A. Interpretation of extracted structures

Here we give an interpretation of the physical meaning of the educed structures at different scales resulting from the extraction step of the present methodology. For simplicity and clarity, a two-dimensional scalar field is used. A 1024×1024 grey-scale image (with values in the range $[0 - 255]$) (see top left of figure 15) obtained from a particular realization of a Julia set has been chosen as the scalar field. This set is of interest since it contains self-similar structures at different scales. The outer region has been faded to white so that all boundaries have the same value.

The extraction procedure described in the main text is applied to the two-dimensional image. The result of the multi-scale decomposition provided by the curvelet transform can be seen on the left images of figure 15. The effect in Fourier space is shown by the spectra on figure 16(*b*), while figure 16(*a*) shows the bin p.d.f.s of the original and filtered fields, in physical space. Note that, in this case, the low-pass filter used for the coarsest scale is (in logarithmic scale) wider than the others. It can be thought as two scales merged into one (the coarsest scale, in this case), and could be done also for other groups of scales.

Each filtered field (image) corresponding to each scale is then iso-contoured at a value equal to its mean plus 3/2 times its standard deviation: see right plots on figure 15. The original field has also been iso-contoured (top right) for comparison.

From the way in which the decomposition is done, as observed in the spectra, the structures educed for each filtered scale have a correspondence with the different 'energetic' bands of the field, as defined by frequency corona in Fourier space. But furthermore, they also have a direct correspondence in physical space with the structures of the original field (features of the image). First, we notice that the spatial localization of the features (structures) educed for each scale is retained, with respect to the original image. As expected, they vary in relative sizes (scales), from one filtered scale to the next. Some features of the original image that span across different scales are split as a result of the decomposition. See, for example, the dark continuous arm of the spiral: scale 0 captures its largest portion, but the remainder can be seen also in the rest of the scales. The geometry of each part resembles that of the structure from which it was derived. Shape is preserved and thus a geometrical analysis of the educed structures is meaningful in this context. The iso-contour obtained from the original field (top right of figure 15) contains a large individual structure, rich in features, and

a few simpler structures, but is missing many other features of the original image. In contrast, contours of the filtered fields tend to contain many more (simpler) structures that capture the essential features of the original field at that scale. The fact that the spatial localization is kept can be used for the study of relative positioning, clustering and other organizational aspects of the sets of structures.

The extension of this reasoning to three dimensions is immediate. The complexity of the structures that can be found increases. For example, structures that appear as circular in two dimensions could become either blob- or tube-like, while elongated structures in two dimensions could become either tube- or sheet-like structures. We note that an alternative to the multi-scale decomposition of the scalar field applied here is to perform a multi-resolution analysis applied to the iso-contours extracted from the original database. Since there is a loss of information by iso-contouring, we choose to perform the multi-scale decomposition first over the entire field and then iso-contour each of the filtered scales.

As an analogy, consider the decomposition of a tree into its trunk, branches, leaves, etc. The outer surface of the tree, containing all those elements, would correspond to the iso-surface of the original field. It is generally too rich and complex to study as a whole. By applying a multi-scale decomposition before iso-contouring, we can separate the tree into its individual components, ranged by the scale. Then, iso-contouring extracts structures at those different levels, whose properties can be studied individually. This is the philosophy applied in our methodology for the study of structures in turbulence. In the same manner that the geometry of the elements of a tree has a relation to their physical functionality, perhaps that is also the case for those structures present in turbulent flows. A multi-scale decomposition followed by surface identification (by iso-contouring based on global contour values or other means) seems an appropriate framework for this study. Its current form can be considered a starting point, but there is much room for refinement: use of additional multi-resolution capabilities (as outlined in the body of the paper), such as multi-orientation decomposition, selection of locally adapted contour levels for optimal feature extraction are, for example, two possible paths for improvement.

Appendix B. Gauss–Bonnet theorem in the shape index, curvedness space

For any compact two-dimensional Riemann manifold without boundaries, M, the Gauss–Bonnet theorem states that the integral of the *Gaussian curvature*, K, over the manifold with respect to area, A, equals 2π times its *Euler characteristic*, χ:

$$\int_M K \, dA = 2\pi\chi(M). \tag{B1}$$

This formula relates the geometry of the manifold (given by the integration of the Gaussian curvature, a differential geometry property) to its topology (given by the Euler characteristic). The Euler characteristic of a surface is related to its *genus*† by $\chi = 2 - 2g$. From the relation among shape index, curvedness and mean and Gaussian curvatures stated in §3 (see equation (3.2)), the following relation can be obtained:

$$K = -\Lambda^2 \cos(\pi\Upsilon). \tag{B2}$$

† The genus of an orientable surface is a topological invariant (as is the Euler characteristic) defined as the largest number of non-intersecting simple closed curves that can be drawn on the surface without disconnecting it.

Then, the Gauss–Bonnet theorem can be restated in terms of the shape index and curvedness as

$$\int_M \Lambda^2 \cos(\pi \Upsilon)\, \mathrm{d}A = 4\pi[g(M) - 1]. \tag{B 3}$$

Furthermore, considering the non-dimensionalization of the curvedness also introduced in §3, $C = \mu\Lambda$ ($\mu \equiv 3V/A$, where V is the volume and A the area of the surface) and taking into account that cosine is a symmetric function and thus $\cos(\pi\Upsilon) = \cos(\pi|\Upsilon|) \equiv \cos(\pi S)$, then (B 1) can be rewritten as

$$\int_M C^2 \cos(\pi S)\, \mathrm{d}A = 4\pi\mu^2[g(M) - 1]. \tag{B 4}$$

The left-hand side of (B 4) can be expressed in terms of the $\{S, C\}$ area-based joint probability density function of the surface, $\mathscr{P}(S, C)$:

$$\int_M C^2 \cos(\pi S)\, \mathrm{d}A = A \cdot \iint C^2 \cos(\pi S)\mathscr{P}(S, C)\, \mathrm{d}S\, \mathrm{d}C. \tag{B 5}$$

Considering the *stretching parameter*, $\lambda \equiv \sqrt[3]{36\pi}(V^{2/3}/A)$, also introduced in §3, the Gauss–Bonnet theorem finally results in an integral relation between the $\{S, C\}$ area-based joint probability density function, $\mathscr{P}$, the stretching parameter, λ, and the genus of the surface, g:

$$\iint C^2 \cos(\pi S)\mathscr{P}(S, C)\, \mathrm{d}S\, \mathrm{d}C = \lambda^3[g(M) - 1]. \tag{B 6}$$

Appendix C. Definition of feature centre and upper and lower distances of a probability density function

Consider a real-valued random variable X with probability density function $f(x)$, $x \in \mathbb{R}$. We define the *feature centre* $\hat{x}$ as

$$\hat{x} \equiv \begin{cases} \bar{x} - d_l \sqrt{1 - (d_l/d_u)^2} & \text{if } d_l < d_u \\ \bar{x} + d_u \sqrt{1 - (d_u/d_l)^2} & \text{if } d_l > d_u \end{cases} \tag{C 1}$$

where $\bar{x}$ is the *mean* or *expected value* of X, $\bar{x} \equiv \int xf\, \mathrm{d}x$. The <u>lower</u> and <u>upper</u> *distances* are defined by

$$d_l \equiv \sqrt{\frac{\int_{x \leqslant \bar{x}} (\bar{x} - x)^2 f\, \mathrm{d}x}{\int_{x \leqslant \bar{x}} f\, \mathrm{d}x}}, \qquad d_u \equiv \sqrt{\frac{\int_{x \geqslant \bar{x}} (\bar{x} - x)^2 f\, \mathrm{d}x}{\int_{x \geqslant \bar{x}} f\, \mathrm{d}x}}. \tag{C 2}$$

The feature centre can be interpreted as a correction to the mean that accounts for the asymmetry (skewness) of the density function $f(x)$ with respect to its mean, defining a new point closer to the region of higher density. When the probability density function $f(x)$ is symmetric, the feature centre and mean coincide ($\hat{x} = \bar{x}$). The upper and lower distances, d_u and d_l, can be regarded as the r.m.s. of the part of the p.d.f. above and below, respectively, its mean value. A graphical example is shown in figure 17, for a probability density function $f(x) = x^2 \exp(-\sqrt{x})/\int_0^\infty \xi^2 \exp(-\sqrt{\xi})\, \mathrm{d}\xi$ that shows a long tail in one direction. The mean, $\bar{x}$, feature centre, $\hat{x}$, and lower and upper distances, d_l and d_u, are superimposed on the probability density function.
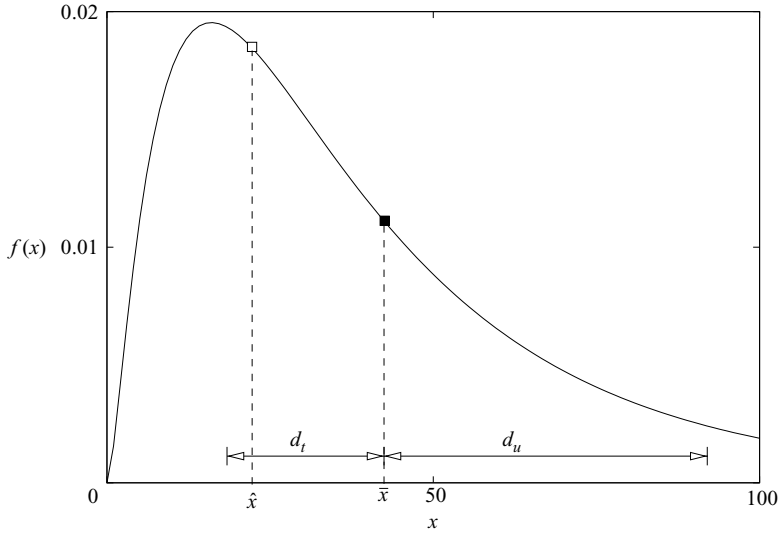
FIGURE 17. Mean ($\bar{x}$) and feature ($\hat{x}$) centres and upper (d_u) and lower (d_l) distances for a sample asymmetric probability density function, $f(x)$.
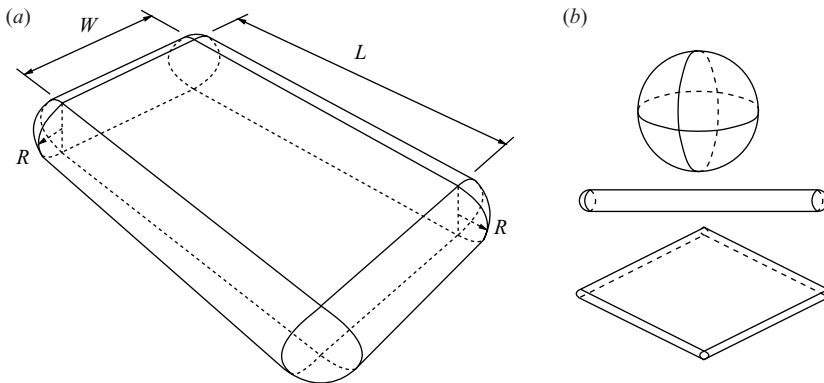

FIGURE 18. Generic structure (*a*) and limiting cases (*b*).

These definitions can be immediately extended to higher-dimensional probability density functions.

Appendix D. Analytic geometric characterization of limiting surfaces

Consider the generic surface in figure 18(*a*). It consists of two planar parallel sheets of area LW separated a distance of $2R$, four halves of circular cylinders of radii R and lengths L and W by pairs, tangent to the planar sheets that they connect and four quarters of a sphere of radius R tangent to the circular cylinders. The resulting surface is closed. The surface is $\mathscr{C}^1$ along the curves of tangency among its parts (across which curvature is discontinuous) and $\mathscr{C}^2$ everywhere else. The area-based joint p.d.f. of S and C is thus still applicable.

Define $\xi \equiv L/R$, $\eta \equiv W/R$. Note that for $\xi = \eta = 0$ the surface is a sphere, for $\xi \gg 1$ and $\eta = 0$ (and vice versa) the surface is a circular tube with spherical
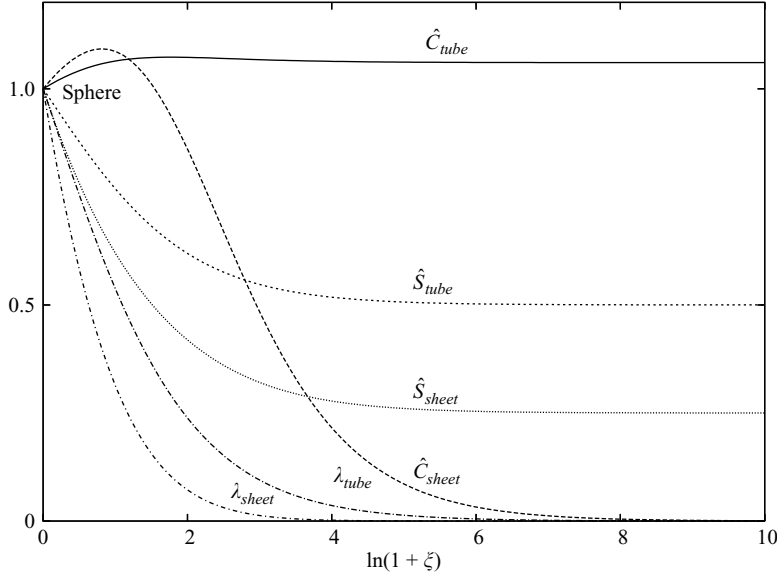
FIGURE 19. $\bar{S}$, $\bar{C}$ and λ as a function of ξ for the tube-like and sheet-like limits, evolving from the sphere limit ($\xi = 0$). Note that the abscissa has been rescaled as $\ln(1 + \xi)$ to show more clearly the transition region.

caps (more stretched as ξ increases), and for $\xi, \eta \gg 1$ the surface is predominantly sheet-like (see figure 18(b)).

The area and volume of this surface are:

$$
A = 4\pi R^2 \left[1 + \frac{1}{2}(\xi + \eta) + \frac{1}{2\pi}\xi\eta \right], \qquad V = \frac{4}{3}\pi R^3 \left[1 + \frac{3}{4}(\xi + \eta) + \frac{3}{2\pi}\xi\eta \right]. \quad \text{(D 1)}
$$

Therefore

$$
\mu \equiv \frac{3V}{A} = R \frac{1 + \frac{3}{4}(\xi + \eta) + \frac{3}{2\pi}\xi\eta}{1 + \frac{1}{2}(\xi + \eta) + \frac{1}{2\pi}\xi\eta}, \qquad \lambda \equiv \sqrt[3]{36\pi}\frac{V^{2/3}}{A} = \frac{\left[1 + \frac{3}{4}(\xi + \eta) + \frac{3}{2\pi}\xi\eta \right]^{2/3}}{1 + \frac{1}{2}(\xi + \eta) + \frac{1}{2\pi}\xi\eta}.
$$
$$
\text{(D 2)}
$$

The principal curvatures, κ_1 and κ_2, are both $1/R$ in the spherical regions, $1/R$ and 0 respectively in the circular cylindrical regions, and both nil in the planar regions of such surface. Thus the dimensionless curvedness associated with each region is $C_{sph} = \mu/R$, $C_{cyl} = \mu/\sqrt{2}R$, $C_{pla} = 0$, respectively. The absolute value of the shape index is $S_{sph} = 1$ for the spherical regions and $S_{cyl} = 1/2$ for the circular cylindrical regions, while its value is undefined for the planar regions. For the purpose of this illustrative example, define such a value as $\gamma \in [0, 1]$.

The mean values of S and C for the surface, in terms of the dimensionless parameters ξ and η, result:

$$
\bar{S} = \frac{1}{A}[S_{sph}A_{sph} + S_{cyl}A_{cyl} + S_{pla}A_{pla}] = \frac{1 + \frac{1}{4}(\xi + \eta) + \frac{1}{2\pi}\xi\eta\gamma}{1 + \frac{1}{2}(\xi + \eta) + \frac{1}{2\pi}\xi\eta}, \qquad \text{(D 3)}
$$

$$\bar{C} = \frac{1}{A}[C_{sph}A_{sph} + C_{cyl}A_{cyl} + C_{pla}A_{pla}] = \frac{\left[1 + \frac{3}{4}(\xi + \eta) + \frac{3}{2\pi}\xi\eta\right]\left[1 + \frac{1}{2\sqrt{2}}(\xi + \eta)\right]}{1 + \frac{1}{2}(\xi + \eta) + \frac{1}{2\pi}\xi\eta}.$$

$$(D\,4)$$

In the limiting cases:
  (i) for a sphere ($\xi = \eta = 0$): $\bar{S} = \bar{C} = 1$;
  (ii) for a predominantly tube-like surface ($\xi \gg 1$, $\eta = 0$): $\bar{S} \approx 1/2$, $\bar{C} \approx 3/2\sqrt{2} \approx 1.06$;
  (iii) for a predominantly sheet-like surface ($\xi = \eta \gg 1$): $\bar{S} \approx \gamma$, $\bar{C} \approx 0$.

Figure 20 shows the dependence on ξ of $\bar{S}$, $\bar{C}$ and λ for the two last cases (surfaces becoming, as ξ increases, tube-like ($\eta = 0$) and sheet-like (with $\eta = \xi$ for simplicity)), starting from the sphere limit ($\xi = 0$). A particular value of γ has been chosen, without loss of generality, in order to represent the limit $\bar{S}_{sheet}$ graphically. In a general sheet-like surface, γ can take any value between 0 and 1, depending on its particular configuration. In the limiting cases ($\xi = 0$ and $\eta = 0$; $\xi \gg 1$ and $\eta = 0$; $\xi = \eta \gg 1$) $\hat{S} \approx \bar{S}$, $\hat{C} \approx \bar{C}$. Thus, a surface predominantly blob-, tube- or sheet-like can be distinguished based on its values of $\hat{S}, \hat{C}, \lambda$.

REFERENCES

ANDREOTTI, B. 1997 Studying Burgers' models to investigate the physical meaning of the alignments statistically observed in turbulence. *Phys. Fluids* **9**, 735–742.

ANTONIA, R. A., CHAMBERS, A. J., FRIEHE, C. A. & VAN ATTA, C. W. 1979 Temperature ramps in the atmospheric surface layer. *J. Atmos. Sci.* **36**, 99–108.

ASHURST, W. T., KERSTEIN, A. R., KERR, R. M. & GIBSON, C. H. 1987 Alignment of vorticity and scalar gradient with strain rate in simulated Navier–Stokes turbulence. *Phys. Fluids* **30**, 2343–2353.

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

BELLMAN, R. 1961 *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

BERKHIN, P. 2002 Survey of clustering data mining techniques. *Tech. Rep.* Accrue Software, San Jose, CA.

BROWN, G. L. & ROSHKO, A. 1974 On density effects and large structure in turbulent mixing layers. *J. Fluid Mech.* **64**, 775–816.

CAMPBELL, R. J. & FLYNN, P. J. 2001 A survey of free-form object representation and recognition techniques. *Comput. Vis. Image Underst.* **81** (2), 166–210.

CANDÈS, E., DEMANET, L., DONOHO, D. & YING, L. 2005 Fast discrete curvelet transforms. *Multiscale Model. Simul.* **5**, 861–899.

CANDÈS, E. J. & DONOHO, D. L. 2003a Continuous curvelet transform: I Resolution of the wavefront set. *Appl. Comput. Harmon. Anal.* **19**, 162–197.

CANDÈS, E. J. & DONOHO, D. L. 2003b Continuous curvelet transform: II Discretization and frames. *Appl. Comput. Harmon. Anal.* **19**, 198–222.

CELANI, A., LANOTTE, A., MAZZINO, A. & VERGASSOLA, M. 2001 Fronts in passive scalar turbulence. *Phys. Fluids* **13**, 1768–1783.

CHEN, S.-G. & WU, J.-Y. 2004 Estimating normal vectors and curvatures by centroid weights. *Comput. Aided Geom. Des.* **21**, 447–458.

CHEN, X. & SCHMITT, F. 1992 Intrinsic surface properties from surface triangulation. In *ECCV '92: Proc. Second European Conference on Computer Vision*, pp. 739–743. Springer.

CHONG, M. S., PERRY, A. E. & CANTWELL, B. J. 1990 A general classification of three-dimensional flow field. *Phys. Fluids* A **2**, 765–783.

DONG, C. & WANG, G. 2005 Curvatures estimation on triangular mesh. *J. Zhejiang Uni. SCIENCE* **6A** (I), 128–136.

DORAI, C. & JAIN, A. K. 1997 Shape Spectrum Based View Grouping and Matching of 3D Free-Form Objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 1139–1146.

FARGE, M. 1992 Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.* **24**, 395–457.

FARGE, M., KEVLAHAN, N., PERRIER, V. & GOIRAND, E. 1996 Wavelets and turbulence. *IEEE Proc., Special Issue on Wavelets* (ed. I Daubechies & J. Kovasevic) **84**(4), 639–669.

FARGE, M., PELLEGRINO, G. & SCHNEIDER, K. 2001 Coherent vortex extraction in 3D turbulent flows using orthogonal wavelets. *Phys. Rev. Lett.* **87**, 054501.

FARGE, M., SCHNEIDER, K. & KEVLAHAN, N. 1999 Non-Gaussianity and Coherent Vortex Simulation for two-dimensional turbulence using an adaptive orthogonal wavelet basis. *Phys. Fluids* **11**, 2187–2201.

FARGE, M., SCHNEIDER, K., PELLEGRINO, G., WRAY, A. A. & ROGALLO, R. S. 2003 Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Phys. Fluids* **15**, 2886–2896.

HORIUTI, K. 2001 A classification method for vortex sheet and tube structures in turbulent flows. *Phys. Fluids* A **13**, 3756–3774.

HORIUTI, K. & TAKAGI, Y. 2005 Identification method for vortex sheet structures in turbulent flows. *Phys. Fluids* **17**, 121703.

HOU, T. Y. & LI, R. 2006 Dynamic depletion of vortex stretching and non-blowup of the 3-D incompressible Euler equations. *J. Nonlinear Sci.* **16**, 639–664.

HUNT, J. C. R., WRAY, A. A. & MOIN, P. 1988 Eddies, stream, and convergence zones in turbulent flows. *Center for Turbulence Research Report* CTR-S88.

IYER, N., JAYANTIAND, S., LOU, K., KALYANARAMAN, Y. & RAMANI, K 2005 Three dimensional shape searching: state-of-the-art review and future trends. *Computer Aided Design* **37**, 509–530.

JABERI, F. A., MILLER, R. S., MADNIA, C. K. & GIVI, P. 1996 Non-Gaussian scalar statistics in homogeneous turbulence. *J. Fluid Mech.* **313**, 241–282.

JAYESH & WARHAFT, Z. 1992 Probability distribution, conditional dissipation, and transport of passive temperature fluctuations in grid-generated turbulence. *Phys. Fluids* A **4**, 2292–2307.

JEONG, J. & HUSSAIN, F. 1995 On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94.

JIMÉNEZ, J., WRAY, A. W., SAFFMAN, P. G. & ROGALLO, R. S. 1993 The structure of intense vorticity in isotropic turbulence. *J. Fluid Mech.* **255**, 65–90.

KERR, R. M. 1985 Higher-order derivative correlations and the alignment of small-scale structures in isotropic numerical turbulence. *J. Fluid Mech.* **153**, 31–58.

KOENDERINK, J. J. & VAN DOORN, A. J. 1992 Surface shape and curvature scales. *Image Vision Comput.* **10**, 557–565.

KOLMOGOROV, A. N. 1941 The local structure of turbulence in incompressible viscous fluids at very large Reynolds numbers. *Dokl. Nauk. SSSR.* **30**, 301–305.

KOLMOGOROV, A. N. 1962 A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds numbers. *J. Fluid Mech.* **13**, 82–85.

LUNDGREN, T. S. 1982 Strained spiral vortex model for turbulent fine structure. *Phys. Fluids* **25**, 2193.

MENEVEAU, C. 1991 Analysis of turbulence in the orthonormal wavelet representation. *J. Fluid Mech.* **232**, 469–520.

MEYER, M., DESBRUN, M., SCHRÖDER, P. & BARR, A. H. 2003 Discrete differential operators for triangulated 2-manifolds. In *Visualizations and Mathematics III* (ed. H. C. Hege & K. Polthier) pp. 33–57. Springer.

MISRA, A. & PULLIN, D. I. 1997 A vortex-based subgrid stress model for large-eddy simulations. *Phys. Fluids* **9**, 2443.

MIURA, H. & KIDA, S. 1997 Indentification of tubular vortices in turbulence. *J. Phys. Soc. Japan* **66**, 1331–1334.

MOISY, F. & JIMÉNEZ, J. 2004 Geometry and clustering of intense structures in isotropic turbulence. *J. Fluid Mech.* **513**, 111–133.

NG, A., JORDAN, M. & WEISS, Y. 2001 On spectral clustering: Analysis and an algorithm. *Adv. Neural Inform. Proc. Sys.* **14**, 849–856.

O'GORMAN, P. A. & PULLIN, D. I. 2004 On modal time correlations of turbulent velocity and scalar fields. *J. Turbulence* **5**, 035.

OSADA, R., FUNKHOUSER, T., CHAZELLE, B. & DOBKIN, D. 2001 Matching 3D models with shape distributions. In *SMI '01: Proc. Intl Conf. on Shape Modeling & Applications*, pp. 154–166. Washington, DC, USA: IEEE Computer Society.

OVERHOLT, M. R. & POPE, S. B. 1996 Direct numerical simulation of a passive scalar with imposed mean gradient in isotropic turbulence. *Phys. Fluids* **8**, 3128–3148.

PERRY, A. E. & CHONG, M. S. 1987 A description of eddying motions and flow patterns using critical-point concepts. *Annu. Rev. Fluid Mech.* **19**, 125–55.

PULLIN, D. I. & LUNDGREN, T. S. 2001 Axial motion and scalar transport in stretched spiral vortices. *Phys. Fluids* **13**, 2553–2563.

PULLIN, D. I. & SAFFMAN, P. G. 1993 On the Lundgren-Townsend model of turbulent fine scales. *Phys. Fluids* A **5**, 126–145.

RICHARDSON, L. F. 1922 *Weather Prediction by Numerical Process*. Cambridge University Press.

ROUSSEEUW, P. J. 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Maths* **20**, 53–65.

SCHWARZ, G. 1978 Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

SCHUMACHER, J. & SREENIVASAN, K. R. 2005 Statistics and geometry of passive scalars in turbulence. *Phys. Fluids* **17**, 125107.

SIGGIA, E. D. 1981 Numerical study of small-scale intermittency in three-dimensional turbulence. *J. Fluid Mech.* **107**, 375–406.

TANAKA, M. & KIDA, S. 1993 Characterization of vortex tubes and sheets. *Phys. Fluids* A **5**, 2079–2082.

TAUBIN, G. 1995 Estimating the tensor of curvature of a surface from a polyhedral approximation. *ICCV '95: Proc. Fifth Intl Conf. on Computer Vision*, pp. 902–907.

TENNEKES, A. A. 1968 Simple model for the small-scale structure of turbulence. *Phys. Fluids* **11**, 669.

TENNEKES, H. & LUMLEY, J. L. 1974 *A First Course in Turbulence*. The MIT Press.

TOWNSEND, A. A. 1951 On the fine-scale structure of turbulence. *Proc. R. Soc. Lond.* A **208**, 534–542.

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

YING, L., DEMANET, L. & CANDÈS, E. J. 2005 3D Discrete curvelet transform. *Tech. Rep.* Applied and Computational Mathematics, California Institute of Technology.

ZAHARIA, T. & PRÊTEUX, F. 2001 3D shape-based retrieval within the MPEG-7 framework. In *Proc. SPIE Volume 4304: Nonlinear Image Processing and Pattern Analysis XII*, pp. 133–145.

ZELNIK-MANOR, L. & PERONA, P. 2004 Self-tuning spectral clustering. *Eighteenth Annual Conference on Neural Information Processing Systems (NIPS)*.