



**HAL**  
open science

## Cessation and reversals of large-scale structures in square Rayleigh-Bénard cells

Andrés Castillo-Castellanos, Anne Sergent, Bérengère Podvin, Maurice Rossi

► **To cite this version:**

Andrés Castillo-Castellanos, Anne Sergent, Bérengère Podvin, Maurice Rossi. Cessation and reversals of large-scale structures in square Rayleigh-Bénard cells. *Journal of Fluid Mechanics*, 2019, 877, pp.922-954. 10.1017/jfm.2019.598 . hal-01921361v2

**HAL Id: hal-01921361**

**<https://hal.science/hal-01921361v2>**

Submitted on 29 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Cessation and reversals of large-scale structures in square Rayleigh–Bénard cells

Andrés Castillo-Castellanos<sup>1,2</sup>, Anne Sergent<sup>2,3†</sup>, Bérengère Podvin<sup>2</sup>,  
and Maurice Rossi<sup>1</sup>

<sup>1</sup>CNRS, Sorbonne Université, Institut Jean Le Rond d’Alembert, F-75005, Paris, France

<sup>2</sup>CNRS, LIMSI, Université Paris-Saclay, F-91405 Orsay, France

<sup>3</sup>Sorbonne Université, UFR d’Ingénierie, F-75005, Paris, France

We consider direct numerical simulations of turbulent Rayleigh–Bénard convection inside two-dimensional square cells. For Rayleigh numbers  $Ra = 10^6$  to  $Ra = 5 \cdot 10^8$  and Prandtl numbers  $Pr = 3$  and  $Pr = 4.3$ , two types of flow regimes are observed intermittently: consecutive flow reversals (CR), and extended cessations (EC). For each regime, we combine proper orthogonal decomposition (POD) and statistical tools on long-term data to characterise the dynamics of large-scale structures. For the CR regime, centrosymmetric modes are dominant and display a coherent dynamics, while non-centrosymmetric modes fluctuate randomly. For the EC regime, all POD modes follow Poissonian statistics and a non-centrosymmetric mode is dominant. To explore further the differences between the CR and EC regimes, an analysis based on a cluster partition of the POD phase-space is proposed. This data-driven approach confirms the successive mechanisms of the generic reversal cycle in CR as proposed in (Castillo-Castellanos *et al.* 2016). However, these mechanisms may take one of multiple paths in the POD phase-space. Inside the EC regime, this approach reveals the presence of two types of coherent time sequences (weak reversals and actual cessations) and more rarely intense plume crossings. Finally, we analyse within a range of Rayleigh numbers up to turbulent flow, the relation between dynamical regimes and the POD energetic contents as well as the residence time in each cluster.

## 1. Introduction

A remarkable aspect of turbulent flows is the organisation of coherent large-scale flow patterns superposed to small-scale fluctuations. Decomposing the turbulent flow into coherent flow structures and incoherent turbulence, allows to focus on dynamically significant events (Hussain 1986). Such decomposition is usually performed either in terms of predetermined basis functions, like spatial Fourier decomposition (see, for instance, Das *et al.* (2000); Chandra & Verma (2011)), or in terms of basis functions extracted from the data, like Proper Orthogonal Decomposition (POD) (see, for instance Bailon-Cuba *et al.* (2010); Podvin & Sergent (2015, 2017); Faranda *et al.* (2019)), Dynamic Mode Decomposition (DMD) (Schmid 2010; Horn & Schmid 2017), or Koopman eigenfunction analysis (Giannakis *et al.* 2018). From this perspective, the coherent structures correspond to a combination of various such modes. In many instances, different configurations for the large-scale flows coexist and the system rapidly switches from one configuration to the other. Let us mention polarity switches in the Earth’s magnetic field or inside the dynamo laboratory experiments (Wicht *et al.* 2009; Valet *et al.* 2012; Fauve *et al.* 2017), or flow reversals inside decaying and stochastically forced

† Email address for correspondence: anne.sergent@limsi.fr

two-dimensional turbulence in the presence of rigid walls (Van Heijst *et al.* 2006; Molenaar *et al.* 2004). In this regard, flow reversals result from the non-linear interactions between modes.

In turbulent Rayleigh-Bénard (RB) convection, flow reversals are also observed where the large scale circulation (LSC), commonly referred to as the *wind of turbulence*, changes sign intermittently (Niemela *et al.* 2001; Kadanoff 2001; Sreenivasan *et al.* 2002). The structure of the LSC and the nature of its variations depend on the geometry of the container (Grossmann & Lohse 2003; Xi & Xia 2008*b*; van der Poel *et al.* 2011). In cylindrical RB cells, the LSC is subject to erratic reorientations of its vertical circulating plane by means of two different mechanisms: a constant *azimuthal meandering* and a momentary decline of the flow magnitude called *cessation*. Both mechanisms may lead to a change in the direction of the flow circulation, either by half a revolution of the LSC plane around the cylinder axis (a *rotation*-led reversal) or by a LSC restart after cessation in the same vertical plane but in the opposite direction (*cessation*-led reversal) Brown *et al.* (2005); Xi *et al.* (2006). Reorientations by cessation are far more rare than those caused by azimuthal rotation. In turn, *cessation*-led reversals represent only a tiny part of cessations, all the angular changes after cessation having the same probability (Brown & Ahlers 2006; Xi & Xia 2008*a*). In cubic RB cells, the LSC plane tends to align to the diagonals of the cell and sometimes changes its orientation through azimuthal rotation. Little or no evidence of cessation-led reversal has been mentioned in literature (Bai *et al.* 2016; Foroozani *et al.* 2017).

There are different ways to focus on the cessation type events. One may hamper the azimuthal meandering by slightly tilting the convection cell, or prevent it by restricting the experimental configuration to slim (or quasi-2D) rectangular cells (Sugiyama *et al.* 2010; Vasilev & Frick 2011). However, a strong dependence of the reversal dynamics on cell aspect-ratio has been noticed in quasi-2D cells (Ni *et al.* 2015). Another possibility is to consider two-dimensional direct numerical simulations (DNS) (Sugiyama *et al.* 2010; Petschel *et al.* 2011; Chandra & Verma 2011, 2013; Podvin & Sergent 2015, 2017). This approach is interesting for different reasons. First, Sugiyama *et al.* (2010) used 2D DNS to identify a region in the  $(Ra, Pr)$  parameter space in which consecutive reversal events are observed, in good qualitative agreement with experimental observations. For this range of parameters, the LSC flow inside a square cell is mainly composed of a large diagonal roll and two counter-rotating corner-rolls. As pointed out by (Sugiyama *et al.* 2010; Chandra & Verma 2013), in 2D reversals the LSC is temporarily replaced by a quadrupolar mode during the transition. A second flow regime has been observed intermittently in 2D DNS inside the same range of parameters. It is mainly composed either of two counter-rotating horizontally stacked rolls or two vertically stacked rolls (Podvin & Sergent 2015). In the following, we refer to both regimes as the regime of *consecutive reversals* (CR) and the regime of *extended cessations* (EC), respectively. The intermittent switching between the CR and EC regimes features a rich dynamical behaviour. Second, two-dimensional (or quasi-2D RB) flows may also be observed in practical configurations, such as those coupling turbulent convection with a strong magnetic field in typical fusion reactor conditions (Zhang & Zikanov 2015). Third, compared to 3D DNS, 2D DNS have a lower computational cost, thus giving access to long-term statistics.

Due to the wide range of involved temporal scales, the large-scale spatio-temporal dynamics of the CR and EC regimes is far from being completely described. In a previous work (Castillo-Castellanos *et al.* 2016), we applied an amplitude-based filter to separate the CR and EC regimes. For the CR regime, a time-rescaling technique based on a well-defined characteristic time scale (the time between consecutive reversals) was combined with an ensemble average to investigate energetics of flow reversals observed in long-

term 2D DNS data. This technique was previously used for reversals of the geomagnetic fields (Valet *et al.* 2012; Lhuillier *et al.* 2013). Using several hundred realisations of flow reversals, we followed the evolution of the global kinetic and available potential energies (as proposed by Winters *et al.* (1995); Hughes *et al.* (2013)), which led to the identification of a generic reversal mechanism which includes a long phase of LSC strength weakening (called *accumulation*) preceding a rapid transition (called *release*) and a rebound. This process shares common features with observations in cylindrical RB cells (Xi & Xia 2008a), as well as with geomagnetic reversals (Valet *et al.* 2012). In comparison, cessations are much less described, in particular regarding representative flow dynamics. In cylindrical cells, it is commonly admitted that reorientations by cessation (including cessation-led reversals) follow Poissonian statistics (Brown *et al.* 2005; Brown & Ahlers 2006; Xi & Xia 2008a). However, only few selected instantaneous flow patterns are given (Xi & Xia 2008b). Additionally, a decoherence of the LSC has also been mentioned (Xi & Xia 2007).

In the present paper, our objective is to separately characterize the spatio-temporal dynamics of *Consecutive Reversals* and *Extended Cessations* regimes in 2D square cavity using the POD approach. First, as previously mentioned some similarities between flows in 2D square cell and in 3D cylinder have been identified in literature (Xi & Xia, 2008a). Connections are based on phases observed in the generic reversal cycle proposed in Castillo-Castellanos *et al.* (2016). This deterministic feature may appear paradoxical with respect to the well-known stochastic features of 3D reversals. This is one of the points we try to elucidate: we reanalyze the 2D flows in order to determine if only a part of flow in the CR regime may follow Poissonian statistics despite the existence of a generic reversal cycle. This question demands to identify the leading POD modes on the data subset corresponding to the CR regime before characterizing the statistics of waiting times of each POD modes. This leads to reinterpret the generic reversal cycle in terms of POD mode dynamics. By applying the same methodology to the EC regime, we recover the expected Poissonian dynamics of cessations described in cylinders. However we seek to identify typical coherent dynamical events in this regime (specifically by distinguishing weak reversals from actual cessations). The total time periods related to EC regime represent only a small part of the complete time series, and no clear characteristic time can be identified. Thus we cannot apply the previously used time-rescaling leading to the identification of a generic mechanism. A specific methodology is consequently necessary to obtain some kind of classification into the spatio-temporal dynamics. This is performed by first identifying POD dynamical modes in the EC subset, and then analyzing the paths followed by the system in POD state space using statistical approaches. In particular, we use a cluster-based analysis of the POD phase-diagrams as proposed by Kaiser *et al.* (2014), to identify typical dynamical events whatever their probability of occurrence. These are described in terms of representative flow patterns (the centroids of the clusters) and time spent by the system inside each type of event.

The paper is organised as follows. Section 2 introduces the model equations, their numerical implementation and the POD approach. In section 3, we present the criterion used to separate the CR and EC regimes and introduce the data sampling for POD. The following sections explore the differences between both regimes: §4 focuses on the spatial structure of coherent modes, §5 on their temporal evolution, while §6 and §7 introduce and apply cluster analysis to analyse the interactions between competing modes for the CR and EC regimes, respectively. In section 8, we follow the energy contained inside different modes and the residence times in clusters as a function of  $Ra$ . This provides a simplified description of the various flow regimes observed. Finally, we summarize our results in section 9.

$Pr$	$Ra$	$n_x n_y$	$t_s$	$n_e$	$\overline{Nu}$	%Diff.	$Pr$	$Ra$	$n_x n_y$	$t_s$	$n_e$	$\overline{Nu}$	%Diff.
4.3	$1 \cdot 10^6$	$512^2$	4500	418	6.83	0.2	3.0	$1 \cdot 10^6$	$512^2$	4500	520	6.60	0.2
	$3 \cdot 10^6$	$512^2$	4500	248	9.09	0.4		$3 \cdot 10^6$	$512^2$	4500	310	8.82	0.4
	$5 \cdot 10^6$	$512^2$	4500	120	10.47	0.5		$5 \cdot 10^6$	$512^2$	4500	133	10.10	0.4
	$8 \cdot 10^6$	$512^2$	4500	76	12.14	0.7		$8 \cdot 10^6$	$512^2$	4500	22	11.60	0.7
	$1 \cdot 10^7$	$512^2$	9600	3	13.06	0.3		$1 \cdot 10^7$	$512^2$	9600	220	12.50	0.3
	$3 \cdot 10^7$	$512^2$	9600	22	18.29	0.2		$3 \cdot 10^7$	$512^2$	9600	261	17.60	0.5
	$5 \cdot 10^7$	$512^2$	65000	605	20.97	0.7		$5 \cdot 10^7$	$512^2$	65000	754	20.70	0.7
	$8 \cdot 10^7$	$512^2$	4500	66	23.96	2.6		$8 \cdot 10^7$	$512^2$	4500	75	23.79	2.5
	$1 \cdot 10^8$	$1024^2$	4500	56	25.77	0.2		$1 \cdot 10^8$	$1024^2$	4500	45	25.36	0.2
	$3 \cdot 10^8$	$1024^2$	4500	9	35.50	1.6		$3 \cdot 10^8$	$1024^2$	4500	1	35.06	1.5
	$5 \cdot 10^8$	$1024^2$	4500	-	41.09	2.8		$5 \cdot 10^8$	$1024^2$	4500	-	40.76	2.5

TABLE 1. Simulation parameters:  $Ra$  and  $Pr$ , number of grid points  $n_x n_y$ , simulation length  $t_s$  in convective time units and number of events  $n_e$  where the global angular momentum  $A_{2D}(t)$  (defined in equation 3.1) changes sign during the simulation length. Numerical convergence of the Nusselt number: average Nusselt number and maximum relative difference between different definitions, see text.

## 2. Model equations and analysis tools

### 2.1. Physical configuration and governing equations

Consider a fluid contained in a square cell, cooled at the top with constant temperature  $T_{\text{top}}$  and heated at the bottom with constant temperature  $T_{\text{bot}} > T_{\text{top}}$ . The flow equations are based on the Boussinesq approximation. The flow is defined by the *Rayleigh* and *Prandtl* numbers,

$$Ra \equiv \frac{g\beta H^3 (T_{\text{bot}} - T_{\text{top}})}{\kappa\nu}, \quad Pr \equiv \frac{\nu}{\kappa} \quad (2.1)$$

where  $g$  denotes gravity,  $H$  the cell height and  $\beta$ ,  $\kappa$ ,  $\nu$  are respectively volumetric thermal expansion, thermal diffusivity and kinematic viscosity coefficients. As far as notations are concerned, variables  $x$  (resp.  $u$ ) and  $y$  (resp.  $v$ ) stand for the horizontal and vertical directions (resp. velocities). Coordinate vector  $\mathbf{x} = (x, y)$  is equal to  $(0, 0)$  at the cavity centre. One introduces the reduced temperature  $\theta(\mathbf{x}, t) \equiv (T - T_0)/(T_{\text{bot}} - T_{\text{top}})$ , with  $T_0 \equiv (T_{\text{bot}} + T_{\text{top}})/2$ . Given a field  $a(\mathbf{x}, t)$ , quantity  $\bar{a}(\mathbf{x})$  (resp.  $\sigma(a)(\mathbf{x})$ ) denotes the time averaged value (resp. standard deviation) computed using the complete long-term time series. Moreover  $\langle a \rangle(t)$  stands for the volume average of  $a(\mathbf{x}, t)$  over the fluid domain.

Based on the cell height  $H$  as characteristic length scale and  $\frac{\kappa}{H} Ra^{0.5}$  as characteristic velocity scale (one obtained from a balance between the friction and buoyancy forces, equivalent to the free-fall velocity divided by  $Pr^{0.5}$ ) the dimensionless velocity  $\mathbf{u} = (u, v)$  and reduced temperature  $\theta$  satisfy the dimensionless system of equations

$$\begin{cases} \nabla \cdot \mathbf{u} &= 0 \\ \partial_t \mathbf{u} + \nabla \cdot [\mathbf{u} \otimes \mathbf{u}] &= -\nabla p + Pr Ra^{-0.5} \nabla^2 \mathbf{u} + Pr \theta \mathbf{e}_y \\ \partial_t \theta + \nabla \cdot [\mathbf{u} \theta] &= Ra^{-0.5} \nabla^2 \theta \end{cases} \quad (2.2)$$

A no-slip condition for the velocity field is ensured on walls. On top (resp. bottom) walls, one imposes  $\theta = -0.5$  (resp.  $\theta = 0.5$ ) while adiabaticity  $\partial_x \theta = 0$  is satisfied on side-walls. From now on, all quantities are written in dimensionless form only.

## 2.2. DNS: parameter range and numerical method

The values of  $(Ra, Pr)$  used for direct numerical simulations (DNS) covers the transition to a turbulent flow regime where reversals have been reported (Sugiyama *et al.* 2010):  $Ra = 10^6 - 5 \cdot 10^8$  for  $Pr = 3$  and  $Pr = 4.3$ . Sections 3 to 7 are mainly focused on  $Ra = 5 \cdot 10^7$ , while section 8 explores the different dynamical regimes observed as function of  $Ra$ . Whenever both  $Pr$  display similar results, we present in detail only those for  $Pr = 3$ . The length of the simulations range from 4500 to 65000 convective time units as to observe an adequate number of reversals of the LSC (see table 1).

Time integration of the governing equations 2.2 is performed through a second-order semi-implicit scheme. It combines a staggered in time discretisation of the velocity and scalar fields with an implicit treatment of the diffusion terms and the Bell-Colella-Glaz advection scheme (Bell *et al.* 1989) for the non-linear terms. Details of this method can be found on (Popinet 2003, 2009). The time step is variable and verifies the Courant-Friedrichs-Lewy condition  $CFL < 0.5$ . Incompressibility is imposed by a projection method. Numerical implementation is done using Basilisk (Popinet 2015). The code has been validated and verified for Rayleigh–Bénard convection in Castillo-Castellanos *et al.* (2016). Simulations are performed using the finite volume method on a regular Cartesian centred grid with 512 or 1024 points in each direction. We verify the spatial resolution by evaluating the numerical convergence of time-averaged Nusselt numbers obtained by different methods: integrated over the top and bottom plates, over the volume, and derived from exact relations to the viscous and thermal dissipation rates (Shraiman & Siggia 1990). In general, these values converge within 2% of the average Nusselt number, see table 1.

## 2.3. Mode extraction using Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition (Lumley 1967) is a statistical technique which provides a representation of a spatio-temporal vector field  $\boldsymbol{\psi}(\mathbf{x}, t)$  as the superposition of a denumeration of space-dependent structures or *empirical eigenfunctions*  $\phi_k(\mathbf{x}), k \geq 1$ , the amplitude of which  $\alpha_k$  varies in time:

$$\boldsymbol{\psi}(\mathbf{x}, t) = \sum_{k=1}^{\infty} \alpha_k(t) \phi_k(\mathbf{x}) \quad (2.3)$$

Here the field consists of the joint velocity and temperature fields  $\boldsymbol{\psi} = \{\mathbf{u}, \gamma\theta\}$ , where  $\gamma$  is an arbitrary rescaling factor here taken equal to 1 (the influence of  $\gamma$  was studied in Podvin & Sergent (2015)).

The empirical eigenfunctions constitute a hierarchy of patterns most likely to be observed in the simulation and they are solutions of the eigenvalue problem

$$\int \mathbf{R}(\mathbf{x}, \mathbf{x}') \phi_k(\mathbf{x}') d\mathbf{x}' = \lambda_k \phi_k(\mathbf{x}) \quad (2.4)$$

where  $\mathbf{R}(\mathbf{x}, \mathbf{x}') \equiv \frac{1}{N} \sum_{n=1}^N \boldsymbol{\psi}(\mathbf{x}, t_n) \boldsymbol{\psi}(\mathbf{x}', t_n)$  is the spatial autocorrelation tensor computed from  $N$  snapshots obtained at times  $t_i, i = 1 \dots N$ . By construction the eigenfunctions are orthogonal (see (Holmes *et al.* 2012) for more details). It can be shown that the problem can be reformulated into what is called the method of snapshots, following Sirovich (1987), as

$$C_{mn} \alpha_n^k = \lambda_k \alpha_m^k, \text{ with } C_{mn} \equiv \langle \boldsymbol{\psi}(\mathbf{x}, t_m) \boldsymbol{\psi}(\mathbf{x}, t_n) \rangle \quad (2.5)$$

The amplitude  $\alpha_n^k = \alpha^k(t_n)$  represents the amplitude of mode  $k$  at time  $t_n$ . By construction the amplitudes are uncorrelated, and if the eigenfunctions are normalized,

we have  $\overline{\alpha_j \alpha_k} = \delta_{jk} \lambda_k$ ,  $\delta_{jk}$  being the Kronecker symbol. The eigenvalues  $\lambda_k, k \geq 1$  represent the average contribution of mode  $k$  to the total energy of the vector field  $E_{comb} \equiv \langle \overline{\theta^2 + \mathbf{u}^2} \rangle = \sum_{k=1}^{\infty} \lambda_k$  and can be ordered  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_k$ .

The idea is then to superimpose the small set of dominant modes identified by the decomposition to build an approximation to the actual flow field (in other words the flow is replaced by its projection onto the first few most energetic modes). We note that the relationship between empirical eigenfunctions and the standard definition of a coherent structure is not entirely straightforward, as empirical eigenfunctions correspond to linear combinations of individual realizations of the flow over the full spatial domain,

$$\phi_k(\mathbf{x}) = \sum_{n=1}^N \alpha_k(t_n) \psi(\mathbf{x}, t_n) \quad (2.6)$$

while standard definitions of coherent structures (such as the corner rolls) are typically local in space. More on this can be found in (Podvin & Sergent 2017).

For this system, we identify the following symmetry operators which leave the governing equations invariant: a reflection symmetry  $\mathbb{S}_x$  with respect to the vertical axis ( $x = 0$ ),

$$\begin{bmatrix} u(x, y) \\ v(x, y) \\ \theta(x, y) \end{bmatrix} \xrightarrow{\mathbb{S}_x} \begin{bmatrix} -u(-x, y) \\ v(-x, y) \\ \theta(-x, y) \end{bmatrix} \quad (2.7)$$

and a reflection symmetry  $\mathbb{S}_y$  with respect to the horizontal axis ( $y = 0$ ).

$$\begin{bmatrix} u(x, y) \\ v(x, y) \\ \theta(x, y) \end{bmatrix} \xrightarrow{\mathbb{S}_y} \begin{bmatrix} u(x, -y) \\ -v(x, -y) \\ -\theta(x, -y) \end{bmatrix} \quad (2.8)$$

The combination of both  $\mathbb{R}_\pi = \mathbb{S}_x \circ \mathbb{S}_y = \mathbb{S}_y \circ \mathbb{S}_x$  represents the centrosymmetry.

$$\begin{bmatrix} u(x, y) \\ v(x, y) \\ \theta(x, y) \end{bmatrix} \xrightarrow{\mathbb{R}_\pi} \begin{bmatrix} -u(-x, -y) \\ -v(-x, -y) \\ -\theta(-x, -y) \end{bmatrix} \quad (2.9)$$

which in addition to the identity form a symmetry group (Podvin & Sergent 2015). In this work, we extract a set of POD modes  $M$ . In the next paragraph, we show that each mode  $M$  possesses symmetries related to elements  $\mathbb{E}$  of the symmetry group. Each mode can be symmetric, i.e.  $\mathbb{E}M = M$ , or antisymmetric, i.e.  $\mathbb{E}M = -M$ . In keeping the naming scheme proposed by Podvin & Sergent (2017), we introduce the following notations for coherent modes: (i) two modes  $Q$  and  $Q^*$ , which are symmetric with respect to all the elements of the group; (ii) two modes  $L$  and  $L^*$ , which are symmetric with respect to  $\mathbb{R}_\pi$ , but antisymmetric with respect to  $\mathbb{S}_x$  and  $\mathbb{S}_y$ ; and (iii) two modes  $S$  and  $S^*$ , also known as symmetry-breaking modes, which are antisymmetric with respect to  $\mathbb{R}_\pi$ . However, mode  $S$  is symmetric with respect to  $\mathbb{S}_y$  and antisymmetric with respect to  $\mathbb{S}_x$ , while mode  $S^*$  has the opposite symmetries (see table 2).

To validate our approach, POD was applied a set of 1200 snapshots for ( $Ra = 5 \cdot 10^7, Pr = 4.3$ ). Snapshots were taken every 5 convective time units (t.u.) from the complete series. Using a longer series or a smaller sampling time did not affect results, giving some confidence in the robustness of our mode extraction. Results are in good agreement with (Podvin & Sergent 2017), see figure 1.

	Mode-Q	Mode-Q*	Mode-L	Mode-L*	Mode-S	Mode-S*
Symmetry $R_\pi$	S	S	S	S	AS	AS
Symmetry $\mathbb{S}_x$	S	S	AS	AS	AS	S
Symmetry $\mathbb{S}_y$	S	S	AS	AS	S	AS

TABLE 2. Definition of modes based on symmetry properties. In each case, we indicate whether the mode is symmetric (S) or antisymmetric (AS) with respect to  $\mathbb{S}_x$ ,  $\mathbb{S}_y$ , and  $R_\pi$ .

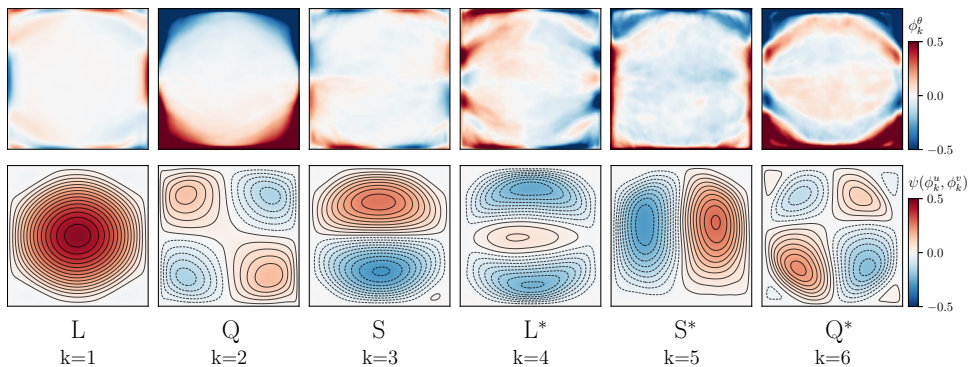


FIGURE 1. Leading POD modes obtained from the complete series for  $Ra = 5 \cdot 10^7$  and  $Pr = 4.3$ . For each figure, colour corresponds to the temperature modes  $\phi_k^\theta$  (top row) and to the streamfunction of the velocity modes  $(\phi_k^u, \phi_k^v)$  (bottom row). Streamlines are also presented in bottom row with dashed lines for negative streamfunction values.

### 3. Data Processing: Identification of regimes and conditional sampling

#### 3.1. Identification of the two flow regimes

For the range of  $(Ra, Pr)$  considered, two different dynamical regimes have been identified by following the global angular momentum

$$A_{2D}(t) \equiv -\frac{1}{2} \int \mathbf{x}^2 \omega(\mathbf{x}, t) \, d\mathbf{x}, \quad \text{with } \omega(\mathbf{x}, t) = \partial_x v - \partial_y u \quad (3.1)$$

as a function of time (figure 2a). Indeed, this quantity is a measure of the organised rotation around the centre of the cavity, i.e the strength of the LSC (Molenaar *et al.* 2004). A criterion based on the amplitude of  $A_{2D}$  was proposed to discriminate the two regimes (Podvin & Sergent 2015; Castillo-Castellanos *et al.* 2016). If  $|A_{2D}| > A_{thres}$  (here set to  $A_{thres} = \overline{|A_{2D}|} + \sigma(|A_{2D}|)$ ) during the interval between two consecutive sign changes in  $A_{2D}$ , then the time interval is assigned to the regime of *consecutive reversals* (CR), corresponding to white areas in figure 2a. A time interval where such threshold is not reached can be of two kinds: weak reversals if only one consecutive sign change is observed, and a cessation of the LSC otherwise. For simplicity, both kinds are assigned to the regime of *extended cessations* (EC), corresponding to grey areas in figure 2a. On the one hand, consecutive reversals are characterised by alternating positive and negative plateaus in  $A_{2D}$ . Positive (resp. negative) plateaus are associated to a large counter-clockwise (resp. clockwise) diagonal roll with two corner-rolls (see figure 2b). On the other hand, extended cessations are characterised by a more erratic dynamics and by the absence of a central vortex (see figure 2c).



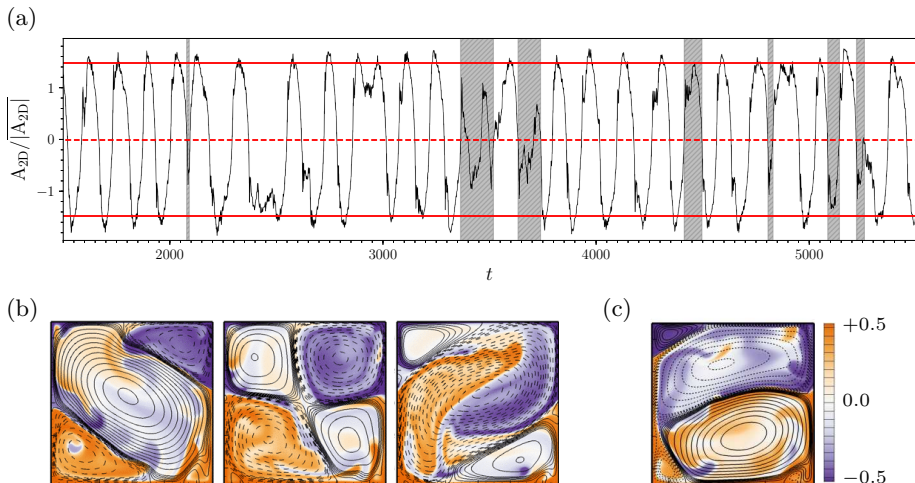


FIGURE 2. Figure (a) Time evolution of  $A_{2D}(t)$  scaled by  $|A_{2D}|$  for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . The threshold separating the CR and EC regimes set to  $A_{thres} = |A_{2D}| + \sigma(|A_{2D}|)$  is shown in red. White areas correspond to CR, while grey areas correspond to EC. Figure (b) (resp. (c)) displays snapshots taken inside the CR (resp. EC) regime. Colour indicates temperature field  $\theta(\mathbf{x}, t)$  and lines are streamlines (negative values shown in dashed).

### 3.2. Conditional sampling for POD

Conditional sampling allows to distinguish and provide quantitative information about interesting spatial and/or temporal regions of a turbulent flow (Antonia 1981). Here, we use the aforementioned criterion to separate our data into two subsets: a first one which corresponds to the CR regime only and a second one which excludes this regime, leaving mostly the EC regime. The idea is now to apply a POD analysis to each subset separately and investigate in detail the dynamics of each regime. In §4 to §7, we focus on the values  $(Ra = 5 \cdot 10^7, Pr = 3)$  and  $(Ra = 5 \cdot 10^7, Pr = 4.3)$  since the system displays in that two cases, an intermittency between the CR and EC regimes. For reference, at  $(Ra = 5 \cdot 10^7, Pr = 3)$  (resp.  $(Ra = 5 \cdot 10^7, Pr = 4.3)$ ) the probability of being inside the CR regime is 79.6% (resp. 83.0%) of the complete series. It is thus possible to highlight the different roles played by the large-scale modes from different viewpoints. These observations will be used in §8 to follow the large-scale flow dynamics over a wider range of  $Ra$  numbers.

For  $(Ra = 5 \cdot 10^7, Pr = 3)$ , the CR subset is restricted to 108 flow reversals, corresponding to 1521 snapshots taken every 4 convective time units. The sampling frequency of snapshots has been adjusted to ensure the different phases of the flow reversals are represented and the number of snapshots was progressively increased until the convergence of POD modes is reached. The EC subset is composed of 1500 snapshots taken every convective time unit. The sampling frequency was increased due to the generally faster dynamics and short duration of this regime. For this regime it was more difficult to obtain converged POD modes and small differences are still observed when considering different sets of snapshots. For  $(Ra = 5 \cdot 10^7, Pr = 4.3)$  the CR subset is composed of 100 reversals, corresponding to 3157 snapshots taken every 4 time units, while the EC subset is composed of 3800 snapshots taken every time unit.

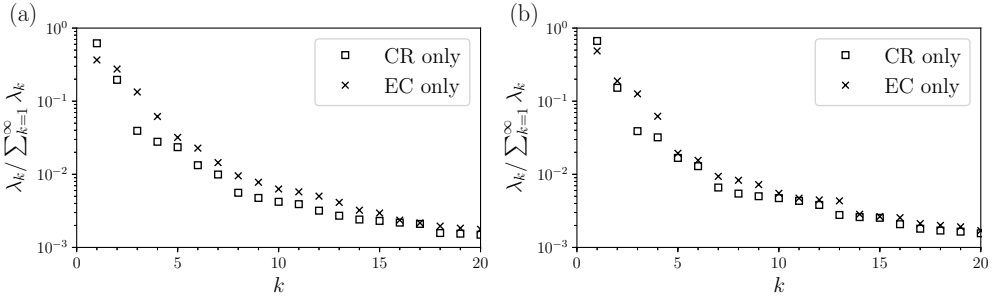


FIGURE 3. Conditional POD analysis of the CR and EC subsets. Normalised eigenvalues for the first 20 modes for  $Ra = 5 \cdot 10^7$  for (a)  $Pr = 3.0$  and (b)  $Pr = 4.3$ .

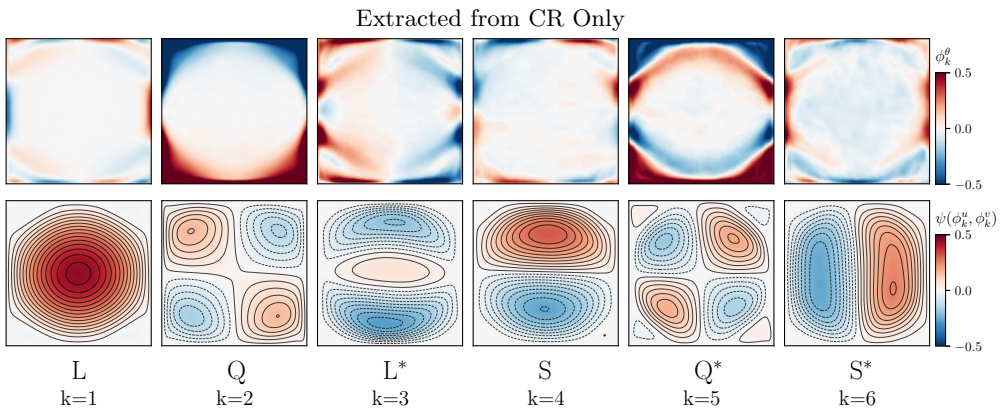


FIGURE 4. Leading POD modes of the CR subset obtained for  $Ra = 5 \cdot 10^7$  and  $Pr = 3.0$ . For each figure, colour corresponds to the temperature modes  $\phi_k^\theta$  (top row) and to the streamfunction of the velocity modes ( $\phi_k^u, \phi_k^v$ ) (bottom row). Streamlines are also presented in bottom row with dashed lines for negative streamfunction values.

#### 4. Comparison of the spatial structures extracted from each regime

For  $(Ra = 5 \cdot 10^7, Pr = 3)$  and  $(Ra = 5 \cdot 10^7, Pr = 4.3)$ , and both CR and EC subsets, the POD spectra exhibits at least two decades of decay between modes  $k = 1$  and  $k = 20$  (figure 3). The two first modes of the CR regime represent around 80% percent of the combined energy  $E_{comb}$ , with an additional 10% contained inside modes  $k = 3$  to 7 for  $Pr = 3$  (resp. modes  $k = 3$  to 6 for  $Pr = 4.3$ ). In comparison, the first two and six modes of the EC regime represent 70% and 86% of  $E_{comb}$ , respectively. Generally EC modes  $k = 2$  to 20 are more energetic than their CR counterparts, indicating a stronger competition between POD modes for the EC subset. In sections 4.1 and 4.2, we study only on the six most energetic POD modes for each regime. In the following and whenever there is no ambiguity, we refer to the amplitude  $\alpha_k(t)$  of the mode  $k$  as the mode itself. For instance if  $k = 1$  mode is a mode  $Q$ ,  $\alpha_1$  is written as  $Q$ .

##### 4.1. Leading POD modes inside the regime of ‘consecutive reversals’ (CR)

Figure 4 displays the POD modes extracted from the CR subset for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Each mode displays a clear symmetry. Mode  $k = 1$  corresponds to  $L$ , associated to a single-roll mode with small recirculation cells in the four corners. In the following, this mode will be used as a marker of the strength of the LSC. Mode  $k = 2$  corresponds to

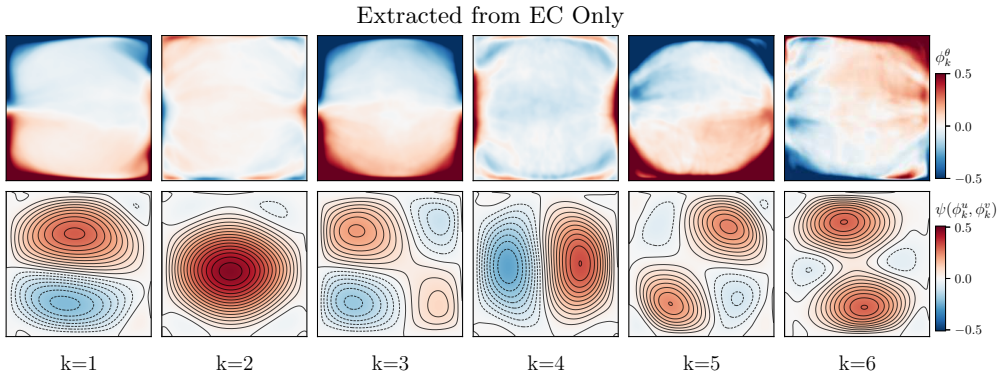


FIGURE 5. Most energetic POD modes of the EC subset obtained for  $Ra = 5 \cdot 10^7$  and  $Pr = 3.0$ . Temperature and velocity modes are displayed as in figure 4.

$Q$ , associated to a four-roll mode and to the mean temperature field. Indeed, a linear combination of centrosymmetric modes  $L$  and  $Q$  roughly corresponds to the LSC with added corner-rolls, the most recognisable flow structure of the CR regime. Mode  $k = 3$  corresponds to  $L^*$ , composed of three vertically stacked rolls;  $k = 4$  to mode  $S$ , composed of two vertically stacked rolls, analogous to the short-lived double-roll mode observed in cylindrical cells (Xi & Xia 2008b). Mode  $k = 5$  corresponds to mode  $Q^*$ , similar to mode  $Q$ , but with four additional rolls carrying cold (resp. hot) fluid close to the bottom (resp. top) plate; and finally  $k = 6$  corresponds to mode  $S^*$ , composed of two horizontally stacked rolls. For  $(Ra = 5 \cdot 10^7, Pr = 4.3)$  one obtains a similar set of POD modes but with modes  $k = 3$  and  $k = 4$  interchanged.

#### 4.2. Leading POD modes inside the regime of ‘extended cessations’ (EC)

Figure 5 displays POD modes extracted from the EC subset for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Such modes are reminiscent of those extracted from the CR subset:  $k = 1$  is similar to the symmetry-breaking mode  $S$ , but the temperature field suggests this mode is a linear combination of modes  $S$  and  $Q$ . Similarly,  $k = 2$  can be seen as a linear combination of centrosymmetric modes  $L$  and  $L^*$  with  $LL^* > 0$ , and corresponds to a vortex detached from the top and bottom plates (Podvin & Sergent 2017). Both leading modes have comparable energetic contents (figure 3). Unlike the CR regime, the EC regime displays a strong competition between modes with different symmetries. Modes  $k = 3$ ,  $k = 4$ , and  $k = 5$ , are reminiscent of modes  $Q$ ,  $S^*$ , and  $Q^*$ , respectively, while  $k = 6$  is a new mode specific to the EC regime. The latter one is composed of two co-rotating vortices surrounded by six smaller counter-rotating rolls. A similar set of POD modes is obtained for  $(Ra = 5 \cdot 10^7, Pr = 4.3)$ .

Given the similarities between the modes extracted from EC and CR subsets (the sole exception being the mode  $k = 6$  of the EC regime), we conclude that both subsets are accurately represented by the same set of coherent spatial structures. This suggests that differences between the CR and EC regimes concern the time evolution of modal amplitudes. In the following sections, we hence project the complete time series into the set of modes extracted from the CR regime as to compare the temporal evolution of the modal amplitudes for both regimes.

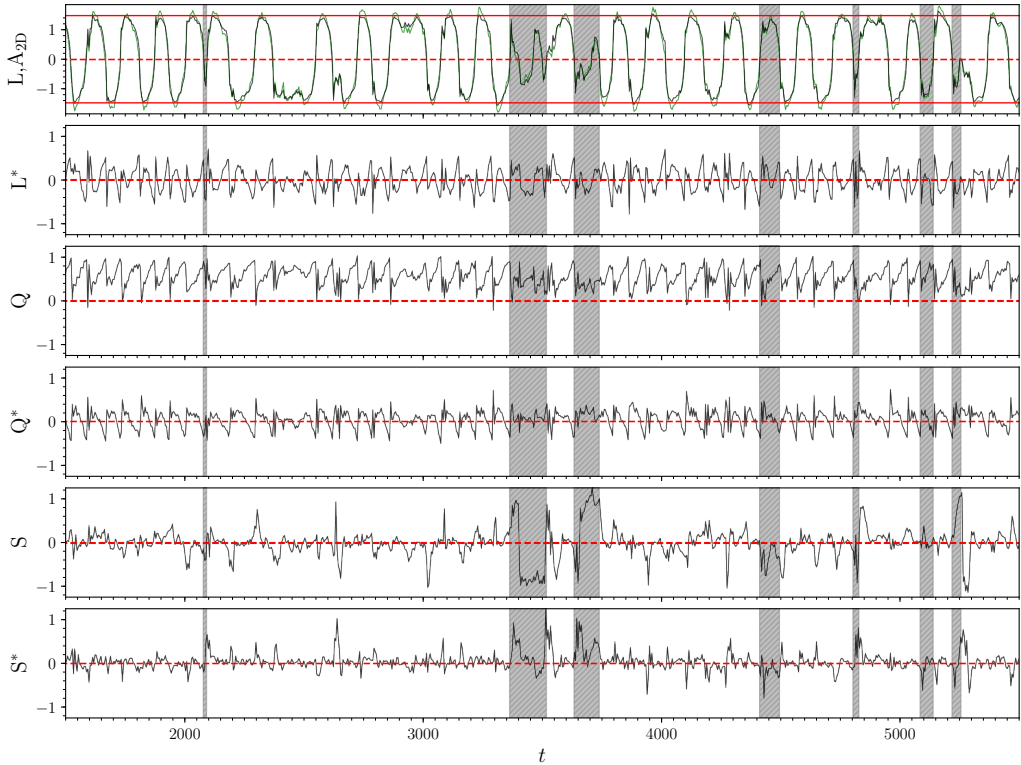


FIGURE 6. Scaled amplitudes  $\alpha_k(t)/|\overline{L}|$  and the scaled angular impulse  $A_{2D}/|A_{2D}|$  for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . White (resp. shaded) regions belong to the CR (resp. EC) subset. Red solid lines on the top figure pinpoints the threshold separating the CR and EC subsets. The presented time series correspond to roughly 1/16 of the total simulation length.

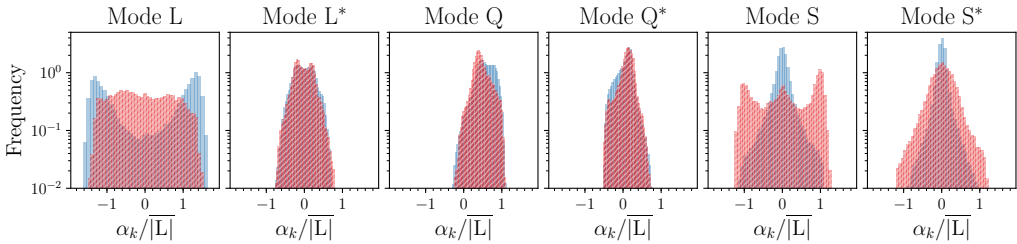


FIGURE 7. Probability distributions of the scaled amplitudes  $\alpha_k(t)/|\overline{L}|$  for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . PDFs inside CR regime are displayed in blue, while conditional PDFs inside EC regime are displayed in red.

## 5. Comparison of temporal evolution in the CR and EC regimes

We analyse here the temporal evolution in both regimes and focus on the coherent or stochastic nature of the dynamics of each POD mode.

### 5.1. General temporal evolution and PDF of modal amplitudes

Let us project the fields  $\psi(\mathbf{x}, t) = (\theta, u, v)$  into the set of POD modes  $\phi_k(\mathbf{x})$  extracted from the CR subset to obtain the modal amplitudes  $\alpha_k(t)$ . For simplicity, these amplitudes are then scaled by the common factor  $1/|\overline{L}|$ . Figure 6 displays their time evolution,

and figure 7 their corresponding probability distribution functions (PDF). A different behaviour is observed for the CR and EC regimes.

During the CR regime (white areas), the dominant mode  $L$  alternates between positive and negative plateaus. Both  $A_{2D}$  (shown in green) and  $L$  are very well correlated and are considered as excellent indicators of the presence of flow reversals. As observed by Podvin & Sergent (2017), the time evolution of mode  $L^*$  is different for positive and negative values of  $L$ . Modes  $S$  and  $S^*$  display positive and negative peaks during transitions, but otherwise fluctuate around zero. Mode  $Q$  increases progressively during the plateaus and suddenly drops during transitions, while  $Q^*$  is centred around zero and displays the opposite behaviour. As for as PDFs are concerned, modes  $L$  and  $L^*$  display a bimodal distribution, modes  $S$  and  $S^*$ , a double exponential distribution, which are almost symmetric. On the contrary, modes  $Q$  and  $Q^*$  have skewed probability distributions. This behaviour can be explained as follows. If a set of amplitudes  $M(t) \equiv \{L, L^*, Q, Q^*, S, S^*\}$  is obtained by the governing equations, then  $\mathbb{S}_x M$ ,  $\mathbb{S}_y M$ , and  $R_\pi M$  are also data sets associated to these governing equations. Furthermore, contrary to modes  $Q$  and  $Q^*$  which are symmetric under operators  $\mathbb{S}_x$ ,  $\mathbb{S}_y$ , and  $R_\pi$ , the modes  $\{L, L^*, S, S^*\}$  are changed into their opposite by the action of at least one of these operators (see table 2). Consequently their PDFs are expected to symmetric with respect to zero. This is not completely the case in figure 7 due to limited number of records.

During the EC regime (shaded areas), the dominant mode  $S$  is characterized by positive and negative plateaus and its PDF takes the form of a symmetric bimodal distribution. This is analogous to the evolution of mode  $L$  during the CR regime. Mode  $S$  is therefore a good marker for this regime. Mode  $S^*$  becomes more energetic than in the CR regime, but does not exhibit such clear plateaus. The time evolution of the other POD modes is also modified by the change of regime: modes  $\{L, L^*\}$  are found to oscillate around zero or may display smaller positive and negative plateaus, while  $\{Q, Q^*\}$  fluctuate around a non-null positive value.

For the above symmetry considerations, the PDFs of modes  $\{L, L^*, S, S^*\}$  are symmetric with respect to zero. As such, these modes are prone to exhibit a reversing dynamics (see figure 6), and the dynamics of each mode is similar for its positive and negative values. By contrast, the PDFs of modes  $Q$  and  $Q^*$  are skewed: mode  $Q$  does not reverse, whereas  $Q^*$  does change sign, but evolves differently for its positive and negative values. Consequently, in order to characterise the temporal dynamics of each regime more precisely, we study the process governing the reversing modes  $\{L, L^*, S, S^*\}$ .

## 5.2. Distribution of waiting times for reversing POD modes

To characterise the reversal process of modes  $\{L, L^*, S, S^*\}$ , we compute for each such mode the waiting time  $\tau$  between consecutive sign changes and evaluate its probability distribution  $P(\tau)$ . The quantity  $h(\tau)d\tau$  where  $h(\tau) \equiv P(\tau)/(1 - \int_0^\tau P(\tau')d\tau')$ , stands for the probability of the occurrence of another reversal inside time interval  $[\tau, \tau+d\tau]$  where  $\tau$  is defined as the time elapsed since the last reversal. The probability distribution  $P(\tau)$  and  $h(\tau)$  are often employed for polarity reversals of Earth's magnetic field. Polarity changes are typically assumed to originate from a general renewal process, in which reversals are independent from one another (see, for instance (Merrill *et al.* 1998; Lhuillier *et al.* 2013)). To verify whether or not the independence hypothesis is correct, the probability distribution  $P(\tau)$  is compared to a Gamma distribution

$$P(\tau) = \left(\frac{k}{m}\right)^k \frac{\tau^{k-1}}{\Gamma(k)} \exp\left(-k\frac{\tau}{m}\right), \quad \text{with } \Gamma(k) = \int_0^\infty x^{k-1} \exp(-x)dx \quad (5.1)$$

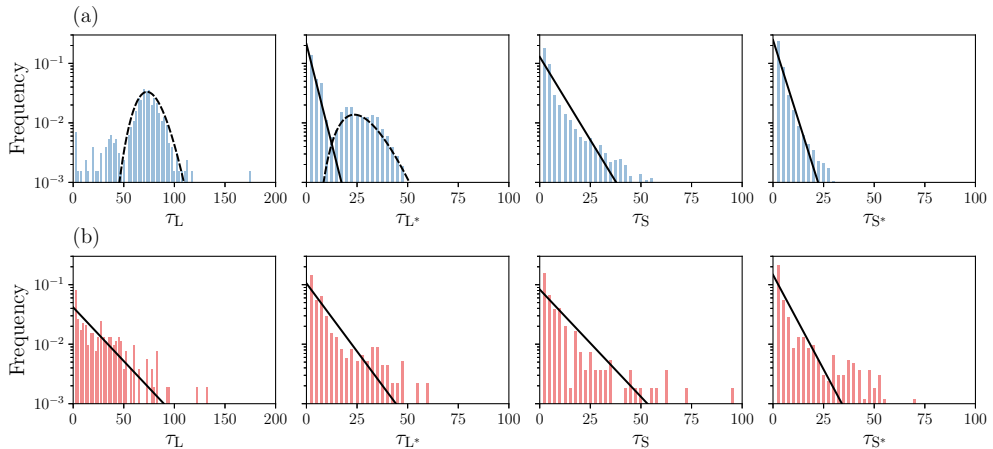


FIGURE 8. Distribution of waiting times  $\tau_M$  of modes  $M = \{L, L^*, S, S^*\}$  for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Figure (a) displays the CR regime and (b) the EC regime. PDFs are obtained using histograms with fixed bins of 2.5 t.u. width. Exponential (solid) and Gamma distributions with  $k \neq 1$  (dashed) are drawn using parameters from table 3.

where  $m = \bar{\tau}$  denotes the mean value of  $\tau$  and the shape parameter  $k$  provides the standard deviation  $m/\sqrt{k}$ . For a Gamma distribution,  $h$  reaches a constant  $\lim_{\tau \rightarrow \infty} h(\tau) = \frac{k}{m}$  at long time intervals (McFadden & Merrill 1986). When  $k = 1$ , the probability  $h$  is found always constant  $h(\tau) = \frac{1}{m}$ . This exponential distribution is interpreted as a lack of memory of the system: reversals are thus generated by a Poisson process characterized by an affine profile in a log plot, i.e.  $\log P(\tau) = -\frac{\tau}{m} - \log m$ . On the contrary, a memory effect appears when  $k \neq 1$ : on the one hand, for  $k > 1$  the distribution is such that  $\lim_{\tau \rightarrow 0} h(\tau) = 0$ , indicating a temporary inhibition of future reversal events at short times. On the other hand, for  $k < 1$ ,  $h$  diverges for  $\tau \rightarrow 0$  indicating an encouragement at short times. Figure 8 displays the distribution of waiting times from the CR and EC subsets for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . PDFs are compared either to Gamma distributions with  $k > 1$  (dashed lines) or to exponential distributions (solid lines) using the parameters listed in table 3. To illustrate the goodness of each fit, table 3 also includes the quantity  $R^2 \equiv 1 - \frac{\sum_{bin} e^2}{\sum_{bin} P_{hist}^2}$  where  $P_{hist}$  is the estimated density distribution using histograms and  $e$  is the error between the fit and the histogram. In this context, a value  $R^2 = 1$  indicates that the Gamma distribution perfectly fits the histogram.

For the EC subset, the distribution of waiting times for all reversing modes fit reasonably well to exponential distributions (figure 8), so indicating a Poisson-like dynamics for all modes studied: the cessation regime seems to be composed of memoryless events. In the EC regime, the lack of memory was identified previously by (Brown & Ahlers 2005, 2006) in three-dimensional flows. Our POD analysis simply confirms that all modes have Poissonian statistics for waiting times and no memory in the EC regime.

For the CR subset, the distributions of  $\tau_S$ , and  $\tau_{S^*}$  also fit well exponential distributions. By contrast, the distributions  $P(\tau_L)$  for mode  $L$  are fitted to Gamma distributions with  $k = 38$ , with large dispersion and a peak value around  $\tau_L = 70$  t.u., several times longer than the large eddy turnover time  $t_E \approx 3.5$  t.u. The leading mode  $L$  thus strongly deviates from Poissonian statistics, suggesting that the system keeps a form of memory of the preceding reversals. This highlights the coherent behaviour of the large-scale structures specific to the CR regime. The distribution of  $\tau_{L^*}$  displays elements of the

(a)				(b)						
		Exp. dist.		Gamma distrib.					Exp. dist.	
$\tau_M$	$n$	$m$	$R^2$	$k$	$m$	$R^2$	$\tau_M$	$n$	$m$	$R^2$
$\tau_L$	523	-	-	38.1	74.9	0.89	$\tau_L$	231	24.1	0.73
$\tau_{L^*} \geq 11.1$	1059	-	-	7.8	27.1	0.96	$\tau_{L^*}$	675	9.5	0.83
$\tau_{L^*} < 11.1$	2287	3.2	0.97	-	-	-				
$\tau_S$	5045	7.8	0.79	-	-	-	$\tau_S$	285	12.0	0.72
$\tau_{S^*}$	8970	4.0	0.83	-	-	-	$\tau_{S^*}$	947	6.8	0.71

TABLE 3. (a) CR subset and (b) EC subset for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Values  $k$  and  $m$  define the fit of the probability distributions of waiting times  $\tau_M$  to either exponential (in that case  $k$  is set to 1) or Gamma distributions with  $k \neq 1$ . The coefficient  $R^2$  quantifies the goodness of the fit (see text). Finally  $n$  denotes the number of events where a mode changes signs.

exponential and Gamma distributions: if one separates the values less than a cut-off value around 11 t.u. (this value is comparable to the duration of the transition from a positive to negative plateau for  $A_{2D}$ , see (Castillo-Castellanos *et al.* 2016)). The shorter timescales  $\tau_{L^*} < 11$  follow an exponential distribution, whereas the longer timescales  $\tau_{L^*} \geq 11$  are governed by a Gamma distribution with  $k = 9$  for  $Pr = 3$  (resp.  $k = 8$  for  $Pr = 4.3$ ). In this respect, the distribution of  $\tau_{L^*}$  reflects the interplay between long-term coherent dynamics and short-term random fluctuations typical of this regime.

As a final remark, evaluating the PDFs of the waiting times considering both subsets at once would mask the more coherent nature of the CR regime.

### 5.3. Generic reversal in the CR regime

The time coherence displayed by some POD modes ( $L$  and  $L^*$ ) in the CR regime can be further studied by revisiting the statistical approach proposed in (Castillo-Castellanos *et al.* 2016). This procedure extracts a generic reversal cycle from an ensemble of realisations. First, one identifies the set of consecutive times  $t_i$  at which a marker of reversal changes sign. Here, we use the amplitude  $L(t)$  as the marker of the reversal. Consider a time interval  $[t_i, t_{i+1}]$  in which the marker  $L(t) \geq 0$ ; a quantity  $a(t)$  in this interval is represented in the re-scaled time  $t_o = (t - t_{i+1})/(t_{i+1} - t_i)$ , so that  $a(t_o|t_o \in [-1, 0])$ . Conversely, for an interval  $[t_i, t_{i+1}]$  in which the marker  $L(t) \leq 0$ ,  $a(t)$  is represented in the re-scaled time  $t_o = (t - t_i)/(t_{i+1} - t_i)$ , so that  $a(t_o|t_o \in [0, 1])$ . The ensemble average over the rescaled curves represents the evolution of  $a(t_o)$  during a generic reversal in the interval  $t_o \in [-1, 1]$ . This treatment may exhibit the presence of a dynamical path in  $a(t)$  that repeats over the ensemble of realisations. In contrast, if  $a(t)$  displays random fluctuations throughout the reversal cycle, its generic curve is flat.

The idea is to use this averaging procedure on the POD mode amplitudes as to characterise (in the statistical sense) their dynamical evolution and to connect it back to the generic reversal mechanism proposed by Castillo-Castellanos *et al.* (2016). In said work, the marker of reversal was the global angular impulse  $A_{2D}(t)$  while quantity  $a(t)$  was the global angular impulse itself or the global kinetic energy  $E_{\text{kin}}(t) \equiv \int \frac{1}{2} \mathbf{u}^2 d\mathbf{x}$ , or else the available potential energy  $E_{\text{apot}}(t)$  (for a precise definition of  $E_{\text{apot}}(t)$ , see Winters *et al.* (1995); Hughes *et al.* (2013)). From this viewpoint, the generic reversal mechanism is composed of 3 consecutive phases. First, an *accumulation* phase characterised by a progressive decrease in  $E_{\text{kin}}$  as  $E_{\text{apot}}$  builds up inside growing corner-rolls. This is followed by a *release* phase where energy is exchanged from  $E_{\text{apot}}$  to  $E_{\text{kin}}$ , complete with a rebound. Last, the *acceleration* phase where  $E_{\text{kin}}$  increases while  $E_{\text{apot}}$  remains fairly

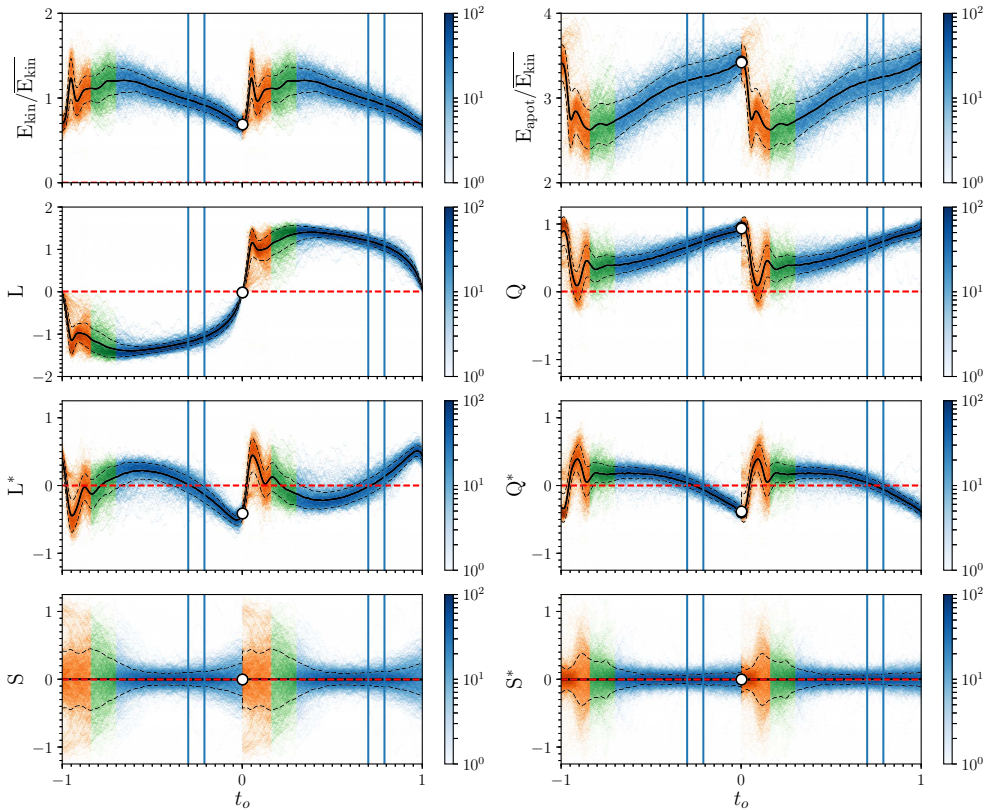


FIGURE 9. Bivariate histograms of the global kinetic energy  $E_{\text{kin}}$ , available potential energy  $E_{\text{apot}}$ , and of the modal amplitudes as function of re-scaled time  $t_o$  during a generic reversal for ( $Ra = 5 \cdot 10^7, Pr = 3$ ). Generic curves are displayed in solid lines, while dotted lines indicate one standard deviation. Colour intensity indicates the probability density (here shown for the blue phase, but equivalent for the other colours), while the colour indicates the 3 phases of a generic reversal in energetic viewpoint: accumulation (blue), release (orange), and acceleration (green). A precursory region to the flow reversal is enclosed between vertical blue lines, see text.

constant during the recovery of the LSC. Additionally, a precursory region was identified, indicating the transition from a LSC dominant to a corner-roll dominant flow.

Now, let us consider the amplitudes of the POD modes during a generic reversal (see figure 9). For the sake of comparison with the energy analysis presented in [Castillo-Castellanos \*et al.\* \(2016\)](#) the accumulation phase is identified in figure 9, by a blue colour, the release phase by an orange colour and the acceleration phase by a green colour. Finally the two vertical lines indicate the position of the precursory region.

Modes  $\{S, S^*\}$  do not display any particularly coherent shape (flat generic curves with larger amplitudes observed from the precursory region until the end of the acceleration phase). By contrast, the modes  $\{L, L^*, Q, Q^*\}$  show some very coherent dynamical patterns. Modes  $\{L, L^*\}$  carry the flow reversal. Given the spatial structure of  $\{L, L^*\}$ , the main circulation is reinforced when both modes have amplitudes with opposite signs. At the beginning of the accumulation phase,  $L$  and  $L^*$  have opposite signs. Consider the case with  $L(t_o) > 0$  and  $L^*(t_o) < 0$  (the opposite case being equivalent). During the accumulation phase,  $L^*$  increases progressively and becomes positive, thus weakening the circulation along the upper and lower boundary layers. A peak in  $L^* > 0$  coincides



with the change in sign of mode  $L$ . Modes  $\{Q, Q^*\}$  carry the thermal energy. Mode  $Q$  progressively increases during the plateau state, then suddenly drops, before returning to its initial value, reminiscent of the generic reversal of  $E_{\text{apot}}$ . This could be expected: by construction, the contribution to  $E_{\text{apot}}$  from modes antisymmetric with respect to  $\mathbb{S}_x$  is null thus leaving symmetric modes  $\{Q, Q^*\}$  as the main carriers of  $E_{\text{apot}}$  since  $S^*$  is negligible. It can be shown by computation that contributions to  $E_{\text{apot}}$  from the spatial eigenfunction related to mode  $Q$  and  $Q^*$  are positive, which can be understood in terms of thermal stratification, see figure 4. As a function of time, mode  $Q^*$  is centred around zero and is anti-correlated to  $Q$ . During the generic cycle, the contribution to  $E_{\text{apot}}$  of mode  $Q^*$  compensates in part that of  $Q$ . Finally, note that modes  $Q^*$  and  $L^*$ , change signs closely to the precursory region. This coincides with a change in slope for  $E_{\text{apot}}$ . Our observations place the precursory time at around 75% of  $t_o$  (or 56 t.u.). By switching signs, these modes, because of their spatial structure, mark the detachment of the LSC from the top/bottom plates and the progressive redistribution of thermal energy towards the bulk which precedes the flow reversal.

For the case ( $Ra = 5 \cdot 10^7, Pr = 4.3$ ), similar trends are found. The change in sign of modes  $L^*$  and  $Q^*$  is observed around 72% of  $t_o$  (or 89 t.u.). The most noticeable difference with respect to  $Pr = 3$  concerns the rebound which is less intense, and the evolution of  $E_{\text{apot}}$  which displays a steeper slope inside the precursory region. Our observations are also in good agreement with the limit cycle obtained from the 5-mode model by Podvin & Sergent (2017). The precursor event and the precursor duration as defined in the same work are also recovered in the generic cycle of POD modes.

## 6. Exploring the CR regime using the cluster-based analysis

One way to explore further the differences between CR and EC regimes, is to study the statistical correlations between modes. However, it is not an easy matter to exhibit connections between the complete set of modes at the same time. To achieve this goal, we provide a drastically simplified description of the spatio-temporal dynamics in terms of network dynamics with  $K$  representative states. The temporal dynamics is then reduced to two elements: the trajectories followed between these  $K$  points and the residence time, which is the time spent inside each point. To define this network, we seek a decomposition of a 6-dimensional phase-space made up of the POD amplitudes recorded every  $\delta t$  into a finite set of clusters of points ( $c_k, k = 1, \dots, K$ ). Each cluster  $c_k$  is characterised by a centroid  $\mu_k$ , i.e. the average of the ensemble of 6-dimensional points in  $c_k$ . Centroids become the points of the network, each associated to a representative spatial pattern, which may be seen as the most common superposition of POD modes inside the cluster. In the following and whenever there is no ambiguity, we refer to the cluster  $c_k$  corresponding to the centroid  $\mu_k$  by the centroid  $\mu_k$  itself.

An adequate sampling frequency  $\delta t$  and a large number of records are key to obtain correct results. For the present work, the POD amplitudes  $\alpha_k(t)$  are taken every  $\delta t = 0.05$  time units. Additionally, we apply the operators  $\mathbb{S}_x$ ,  $\mathbb{S}_y$ , and  $\mathbb{R}_\pi$  (defined in equations 2.7 to 2.9) to the amplitudes  $\alpha_k$  to increase the number of records since such data set is a possible one for the dynamics. This greatly facilitates the interpretation by enforcing the natural symmetries of the system. The partitioning of the phase-space is done using the  $K$ -means clustering algorithm as described in Kaiser *et al.* (2014). The algorithm is tasked to find a partition that minimizes the total within-cluster variance, i.e. the squared distance between points inside a cluster and the corresponding centroid. This requires fixing the number  $K$  of clusters. Further details are provided in appendix A.

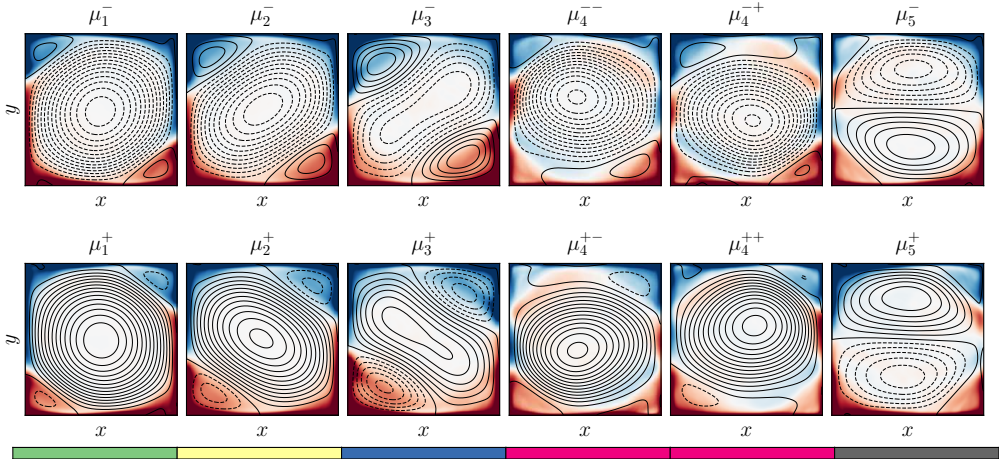


FIGURE 10. Flow patterns corresponding to the  $K = 12$  cluster centroids extracted from the CR subset for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Colour indicates the temperature field, and lines correspond to the streamlines (negative values of streamfunction are in dashed). The five groups are identified by a different colour of the bottom ruler.

In the present section we validate our approach on the CR regime. The study of the EC regime is presented in section 7.

### 6.1. Flow patterns associated to cluster centroids in the CR regime

For  $(Ra = 5 \cdot 10^7, Pr = 3)$ , partitioning the CR subset using  $K = 12$  clusters is found to be satisfactory (see appendix A). These clusters are organised in 5 groups  $(G_k^{cr}, k = 1, \dots, 5)$ , each associated to a different colour in figures 10 to 12:

- Three groups  $G_k^{cr}$ ,  $(k = 1, 2, 3)$ , each characterized by two clusters which are  $\mathbb{S}_x$  and  $\mathbb{S}_y$  antisymmetric. The cluster centroids are denoted  $\{\mu_k^+, \mu_k^-\}$ , superscripts indicating the sign in mode L. Colours are green, yellow and blue, respectively.
- A single group  $G_4^{cr}$  with four clusters. Centroids are denoted  $\{\mu_4^{++}, \mu_4^{+-}, \mu_4^{-+}, \mu_4^{--}\}$ , superscripts indicating the quadrant in the (L,S) plane. Colour is red.
- A single group  $G_5^{cr}$  with two clusters which are  $\mathbb{S}_y$ -symmetric. The cluster centroids are denoted  $\{\mu_5^+, \mu_5^-\}$ , superscripts indicating the sign in mode S. Colour is grey.

The flow patterns associated to each cluster centroid are displayed in figure 10. They indeed capture the underlying physics. For members of  $G_1^{cr}$  (green),  $G_2^{cr}$  (yellow) and  $G_3^{cr}$  (blue), the flow pattern exhibits a large diagonal roll with counter-rotating corner-rolls of different sizes typical of the growth of corner-rolls. For members of  $G_4^{cr}$  (red) the flow pattern exhibits a central vortex partially detached from the top/bottom plates which is typical of the rearrangement of the LSC. Members of  $G_1^{cr}$  to  $G_4^{cr}$  can be easily recognised as rough approximations of the LSC during a generic reversal (see, for instance figure 7 in Castillo-Castellanos *et al.* (2016)). For members of  $G_5^{cr}$  (grey), the flow pattern corresponds to a vertically stacked double-roll which forms an horizontal jet. A horizontal jet impacts one of the side-walls along which two small corner-rolls are placed, reminiscent of the leading mode of the EC regime (see figure 5). This flow pattern is predominantly observed during cessations (Podvin & Sergent 2015).

### 6.2. Analysis of reversals in cluster space

The simplified representation *via clusters* is also capable to retrieve the three phases of the reversal cycle (accumulation, release, and acceleration). Figure 11 displays the clusters

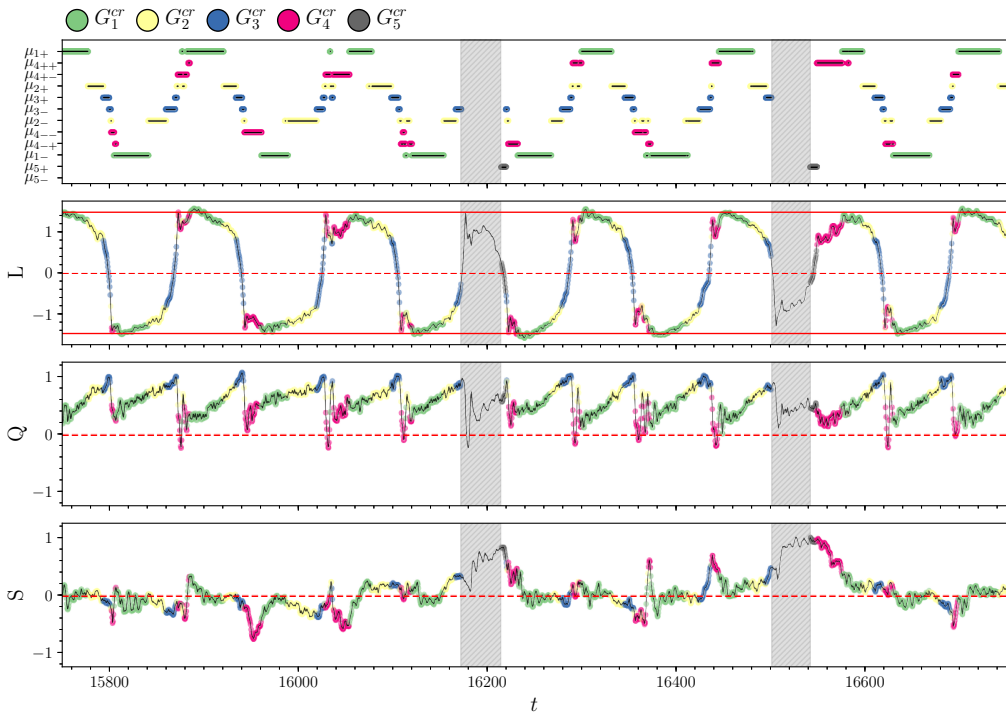


FIGURE 11. Time series of the cluster  $\mu_k$  visited at time  $t$  and the corresponding amplitudes of modes  $L$ ,  $Q$ , and  $S$ , during a series of particular flow reversals observed for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Data-points inside the EC subset correspond to the shaded areas, whereas data-points from the CR subset are coloured to match the groups  $G_k^{cr}$ , see colour code at the top part of the figure.

visited as a function of time are shown alongside a time series of the amplitudes of modes  $L$ ,  $Q$ , and  $S$ . The time series covers a series of standard flow reversals interspersed by two weak reversals, which are assigned to the EC subset and will be discussed in §7. For all standard reversals, the clusters are almost always visited in the same sequence. Starting from a positive  $L$ , the sequence  $(\mu_1^+, \mu_2^+, \mu_3^+)$  illustrates the growing corner-rolls observed during the accumulation phase, where  $\mu_2^+$  roughly corresponds to a precursory region identified in §5.3. This is followed by a sequence  $(\mu_3^+, \mu_3^-, \mu_2^-)$  which splits into two branches (either  $\mu_4^{--}$  or  $\mu_4^{+-}$ ) and corresponds to the beginning of the release phase (sign change in  $L$ ) leading to the rebound event. Finally, during the complex re-organisation process typical of the acceleration phase, the system alternates between different clusters  $(\mu_4^{--}, \mu_1^-, \mu_4^{+-})$ , before eventually settling down on  $\mu_1^-$ .

Figure 12 displays in the planes  $(L, Q)$  and  $(L, S)$ : (i) a random sample of data-points coloured according to the corresponding cluster groups, (ii) the position of the cluster centroids, and (iii) the generic reversal cycle from §5.3. In these planes, all of cluster centroids, are located along the generic cycle with the exception of centroids  $\{\mu_5^+, \mu_5^-\}$  and centroids  $\{\mu_4^{++}, \mu_4^{+-}, \mu_4^{-+}, \mu_4^{--}\}$ . It is inside the  $(L, S)$  plane that the two branches leading to the rebound event are best seen. Since both branches have equal probability of occurrence, by averaging over the ensemble of realisations, the generic cycle inevitably ‘smoothes out’ both branches. Consequently, cluster analysis is better suited to work whenever multiple transition paths are observed.

For each group  $G_k^{cr}$ , we compute the residence time  $\%p_{cr}$ , as well as the conditional average for different global quantities (table 4a). The system spends 70% of time inside

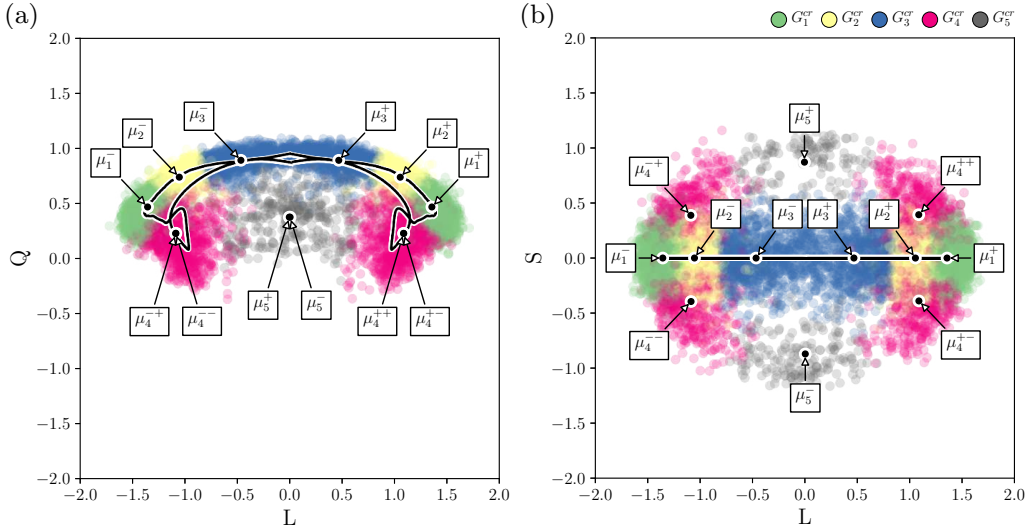


FIGURE 12. Phase diagram in the  $(L, Q)$  and  $(L, S)$  planes for  $(Ra = 5 \cdot 10^7, Pr = 3)$  inside the CR subset. For reference, the generic reversal from §5.3 is displayed in solid lines. Data-points are coloured to match the groups  $G_k^{cr}$ , see colour code at the top of the figure. Solid marks indicate the corresponding centroids.

(a)	$\%p_{cr}$	$E[ A_{2D} ]$	$E[E_{kin}]$	$E[E_{aprot}]$	(b)	$\%p_{ec}$	$E[ A_{2D} ]$	$E[E_{kin}]$	$E[E_{aprot}]$
$G_1^{cr}$	43.5	1.49	1.15	0.96	$G_1^{ec}$	21.9	1.03	1.02	0.97
$G_2^{cr}$	26.0	0.99	0.95	1.04	$G_2^{ec}$	5.8	0.28	0.75	1.06
$G_3^{cr}$	13.6	0.36	0.76	1.09	$G_3^{ec}$	4.6	0.31	1.03	0.84
$G_4^{cr}$	14.5	0.99	1.05	0.91	$G_4^{ec}$	49.0	0.57	0.88	1.02
$G_5^{cr}$	2.4	0.36	0.82	1.06	$G_5^{ec}$	18.7	0.18	0.78	1.14
	100.0	1.11	1.03	0.99		100.0	0.57	0.89	1.03

TABLE 4. Conditional averages over each of the groups extracted from the (a) CR and (b) EC regimes for  $(Ra = 5 \cdot 10^7, Pr = 3)$ .  $p_{cr}$  (resp.  $p_{ec}$ ) denotes the probability for the system to be inside  $G_k^{cr}$  (resp.  $G_k^{ec}$ ) while  $E[a]$  denotes the conditional average of  $a(t)$  over a cluster normalised by the average  $\bar{a}$  over the complete series.

$G_1^{cr}$  and  $G_2^{cr}$ , while it spends 14% of time inside  $G_3^{cr}$  and  $G_4^{cr}$  each. This is in good agreement with the duration of each phase of the generic reversal cycle with an energetical viewpoint (Castillo-Castellanos *et al.* 2016). The sequences described above are also consistent with the evolution of  $A_{2D}$ ,  $E_{kin}$ , and  $E_{aprot}$  during a generic reversal. For instance, the sequence  $(\mu_1^+, \mu_2^+, \mu_3^+)$  displays a gradual increase of  $E[E_{aprot}]$  as in the accumulation phase. By contrast, the group  $G_5^{cr}$  is only visited whenever deviations from standard reversals become important, like at the beginning and at the end of a series of consecutive reversals (see, for instance  $t \sim 16200$  and  $t \sim 16550$  in figure 11). For such events the residence time is always of a few per cent (table 4a).

## 7. Exploring the EC regime using the cluster-based analysis

In the previous section, we have shown that cluster analysis recovers the features of the CR regime obtained using the statistical method of section 5.3 and it is even able

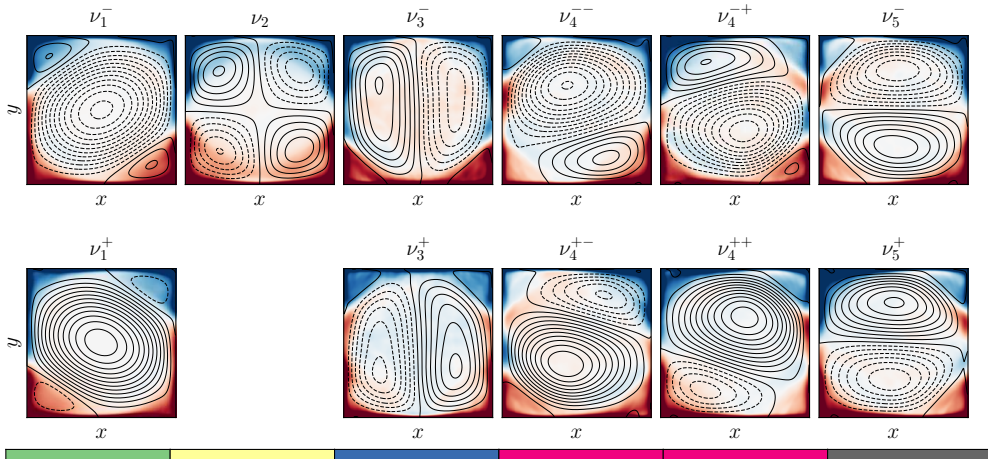


FIGURE 13. Flow patterns corresponding to the 11 cluster centroids extracted from the EC subset for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . Colour indicates the temperature field, while lines correspond to streamlines (negative values of streamfunction in dashed). The 5 groups of centroids are identified by a different colour as shown at the bottom ruler of the figure.

to obtain more detailed aspects. Here, the same cluster analysis explores the regime of extended cessations (EC), which is far less well understood. Note that the statistical method used in section 5.3 would be ineffective for the EC regime, since this method relies on the presence of a dominant timescale (the waiting time between consecutive reversals) to rescale time and extract a dynamical pattern from many realisations. Such dominant timescale has not been identified for the EC regime.

### 7.1. Flow patterns associated to the cluster centroids in the EC regime

For  $(Ra = 5 \cdot 10^7, Pr = 3)$ , partitioning the EC subset using  $K = 11$  clusters is found to be satisfactory (see appendix A). For clarity, the centroids extracted from the EC subset are noted  $\nu_k$  in order to differentiate the CR and EC subsets. These clusters are organised in 5 groups ( $G_k^{ec}, k = 1, \dots, 5$ ), each associated to a different colour in figures 14 to 16:

- A single group  $G_1^{ec}$  with two clusters which are  $\mathbb{S}_x$  and  $\mathbb{S}_y$  antisymmetric. Centroids are denoted  $\{\nu_1^+, \nu_1^-\}$ , superscripts indicating the sign of mode L. Colour is green.
- A single group  $G_2^{ec}$  with a single cluster  $\{\nu_2\}$  which satisfies all the symmetries. Colour is yellow.
- A single group  $G_3^{ec}$  with two clusters which are  $\mathbb{S}_x$ -symmetric. The cluster centroids are denoted  $\{\nu_3^+, \nu_3^-\}$ , superscripts indicating the sign of mode S\*. Colour is blue.
- A single group  $G_4^{ec}$  with four clusters. Centroids are denoted  $\{\nu_4^{++}, \nu_4^{+-}, \nu_4^{-+}, \nu_4^{--}\}$ , superscripts indicating quadrant in the (L,S) plane. Colour is red.
- A single group  $G_5^{ec}$  with two clusters which are  $\mathbb{S}_y$ -symmetric. The cluster centroids are denoted  $\{\nu_5^+, \nu_5^-\}$ , superscripts indicating the sign of mode S. Colour is grey.

The flow patterns associated to each centroid are shown in figure 13. The large diagonal rolls are associated to  $\nu_1^+$  and  $\nu_1^-$  while a quadrupolar mode, commonly observed during standard reversals in the CR regime, is associated to  $\nu_2$ . A horizontally stacked double-roll pattern with small corner-rolls along the top or bottom plates is associated to  $\nu_3^+$  and  $\nu_3^-$ . For  $\{\nu_4^{++}, \nu_4^{+-}, \nu_4^{-+}, \nu_4^{--}\}$  the flow pattern corresponds to a large vortex squeezed between one small and one big corner-roll, clearly reminiscent to that of  $\{\mu_4^{++}, \mu_4^{+-}, \mu_4^{-+}, \mu_4^{--}\}$ . Finally, a vertically stacked double-roll pattern with corner-rolls

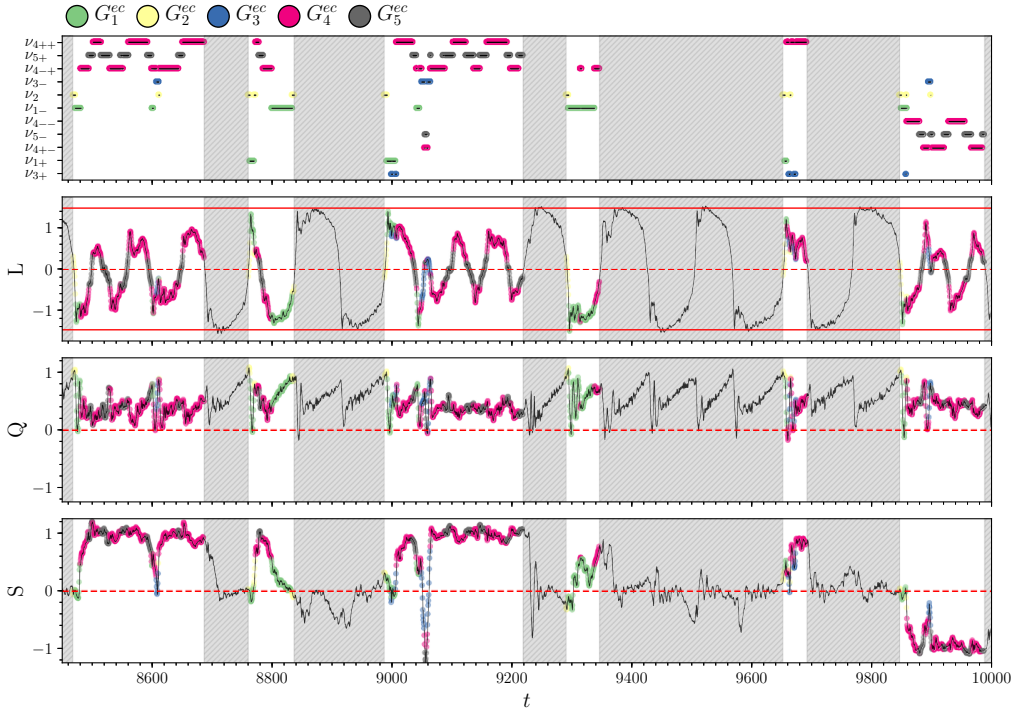


FIGURE 14. Time series indicating the cluster  $c_k^{ec}$  visited at time  $t$  and the corresponding amplitudes of modes  $L$ ,  $Q$ , and  $S$ , during a series of extended cessations interspersed by several periods of consecutive flow reversals for  $(Ra = 5 \cdot 10^7, Pr = 3)$ . The EC subset corresponds to white areas. Data-points from the EC subset are coloured to match the groups  $G_k^{ec}$ , see colour code at the top part of the figure.

along the left or right side-walls, nearly identical to that of  $\mu_5^+$  and  $\mu_5^-$ , is associated to  $\nu_5^+$  and  $\nu_5^-$ .

## 7.2. A cluster space approach of cessations

For the EC regime, figure 14 displays the clusters visited as a function of time alongside a time series of modes  $L$ ,  $Q$ , and  $S$ , as done previously in figure 11 for the CR regime. Similarly, figure 15 displays the phase diagrams in the  $(L, S)$  and  $(S, S^*)$  planes, being the most relevant planes for the cessation dynamics.

First, note that data-points in figure 15 are evenly distributed between the clusters inside groups  $G_1^{ec}$ ,  $G_4^{ec}$ , and  $G_5^{ec}$ . This is confirmed by computing the residence time for each *single* cluster, which is at most 12% and is consistent with table 4b since the residence time  $\%p_{ec}$  in each group is nearly proportional to the number of elements.

The cluster-based analysis discriminates between different kinds of dynamical events within EC regime. Based on the sequence of clusters visited, one identifies two dynamics: weak reversals (windows in figure 14 containing  $t \sim 8800$  and  $t \sim 9300$ ) and actual cessations of the LSC (the remaining windows). Weak reversals deviate from standard reversals in CR mainly because of a larger amplitude of mode  $S$ . This is likely due to the asymmetric growth of corner-rolls. During these events, data-points are found to visit prominently the cluster group  $G_1^{ec}$  and to a smaller degree the group  $G_4^{ec}$ . At the beginning of each weak reversal, the system visits  $G_2^{ec}$ , then  $G_1^{ec}$ . It may then visit  $G_4^{ec}$  and  $G_1^{ec}$  through different paths.

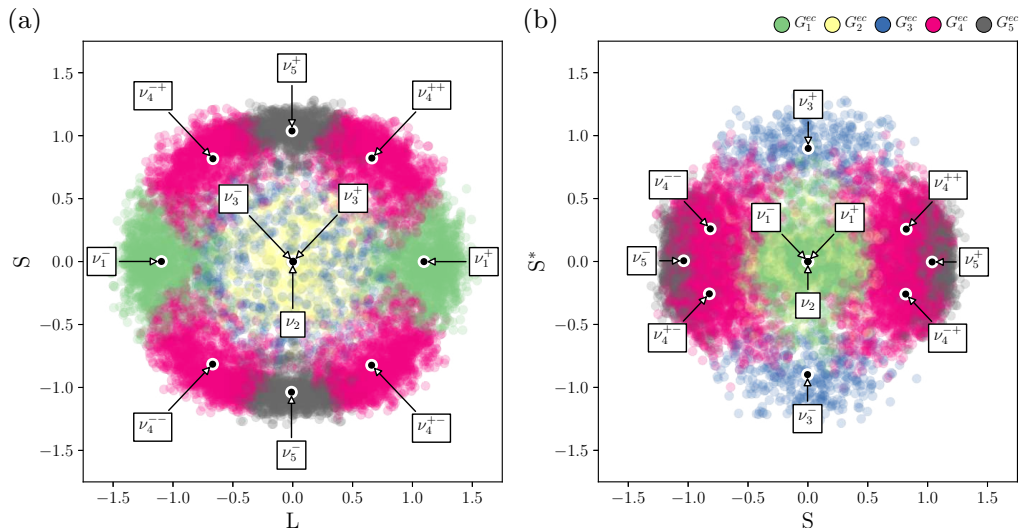


FIGURE 15. Phase diagram in the  $(L,S)$  and  $(L,S^*)$  planes for  $(Ra = 5 \cdot 10^7, Pr = 3)$  inside the EC subset. Colour is used to identify each cluster group, see colour code at the top of the figure:  $G_1^{ec}$ , and  $G_2^{ec}$  predominantly associated to weak reversals, and  $G_3^{ec}$ ,  $G_4^{ec}$ , and  $G_5^{ec}$  predominantly associated to extended cessations. Solid marks indicate the corresponding centroids.

At the beginning of every actual cessation, the system visits  $G_2^{ec}$ , then  $G_1^{ec}$  as for weak reversals. However the situation is then dynamically quite different: mode  $L$  oscillates around zero,  $Q$  fluctuates around a nearly constant value, while  $S$  displays a plateau sporadically interrupted by short excursions like the ones observed around  $t \sim 8600$ ,  $t \sim 9100$ , and  $t \sim 9900$  in figure 14. On the  $(L,S)$  and  $(S,S^*)$  planes, clusters  $\{\nu_4^{++}, \nu_5^+, \nu_4^{++}\}$  and  $\{\nu_4^{--}, \nu_5^-, \nu_4^{--}\}$  are found to be sufficiently close to one another to be considered as part of two attractive basins. This qualitative assessment can be confirmed by measuring the distance between centroids in the 6-D subspace. These basins precisely correspond to the actual cessations (plateau in mode  $S$ , oscillation in  $L$ ). In trajectory terms, the system moves back and forth between three clusters inside the groups  $G_4^{ec}$  and  $G_5^{ec}$ , e.g.  $\{\nu_4^{++}, \nu_5^+, \nu_4^{++}\}$  and the short excursion of mode  $S$  are through  $G_1^{ec}$ ,  $G_2^{ec}$ , and  $G_3^{ec}$  before returning to one of the attractive basins. From the centroid patterns of figures 13, it can be inferred that the back and forth between  $G_4^{ec}$  and  $G_5^{ec}$  is related to the competition between two counter-rotating vertically stacked vortices, where the relative size is reflected on the sign of  $L$ . This is observed in the snapshots presented in figures 16a to 16f. Oscillations in mode  $L$  are physically related to a continuous push and shove motion between a large ascending plume and one descending along one of the side-walls from which an horizontal jet is formed. Both plumes remain locked along the same side-wall until the plateau in mode  $S$  is disrupted during a short excursion. A short excursion is presented in figures 16g to 16l. This particular event is triggered by the weakening of the horizontal jet (figure 16g) which favours the emergence of a quadrupolar pattern (figure 16h). A new unstable pattern forms from the merging of vortices coming from opposite corners (figure 16i). This new flow structure rotates until it eventually re-organises into a more stable vertically stacked double-roll pattern. The most commonly observed feature during the rotation of the unstable pattern is the brief passage of an intense plume through the centre of the cavity, which corresponds to the passage through  $G_3^{ec}$  (figure 16j). This rapid event is characterised by large fluctuations in mode  $Q$  and a substantial increase (of around 40%) in the global heat-transfer. For such rapid events, the residence

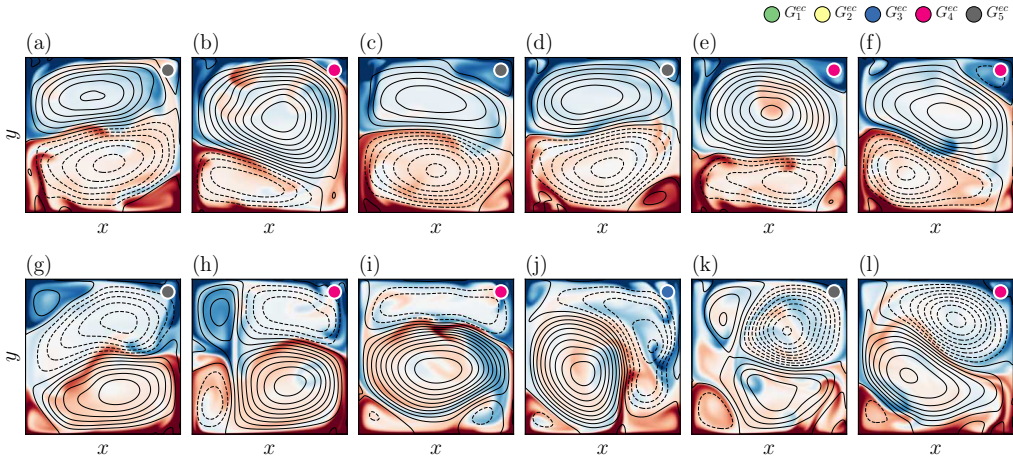


FIGURE 16. Sequence of instantaneous fields observed during actual cessations for ( $Ra = 5 \cdot 10^7, Pr = 3$ ). Colour indicates the temperature field  $\theta(\mathbf{x}, t)$ , lines are streamlines (negative streamfunction values in dashed). The coloured symbol of the top-right corner indicates the current cluster group (colour code at the top of the figure). Snapshots (a) to (f) are taken every 20 t.u. from  $t = 9090$  to  $t = 9190$ , whereas snapshots (g) to (l) are taken every 4 t.u. from  $t = 9882$  to  $t = 9902$  illustrating a short excursion.

time is 4.6% of the total EC subset, that is less than 1% of the total simulation length (table 4b).

It is remarkable how the attractive basins  $\{\nu_4^{-+}, \nu_5^+, \nu_4^{++}\}$  and  $\{\nu_4^{--}, \nu_5^-, \nu_4^{-+}\}$  are reminiscent of the double-roll modes (noted DMR1, DRM2, and DRM3 in Xi & Xia (2008b)) observed inside cylindrical cells. For certain aspect ratios, the asymmetric configurations are more frequently observed than the symmetric one. This behaviour could be explained in part by a 3-D analogue of the push and shove motion described above.

## 8. Evolution of dynamical regimes as function of Rayleigh number

For a given  $Pr$  (in practice  $Pr = 4.3$  or  $Pr = 3$ ), we take advantage of the POD and cluster-based analyses to make evidence of the dynamical flow regimes observed for different values of  $Ra$  which covers the transition from a steady-state to a turbulent flow regime. More precisely, a first study is based on the energetic content of the leading POD modes, a second on the residence time inside the different groups of clusters.

### 8.1. Energy contents of leading POD modes

To allow for a direct comparison between different  $Ra$ , we use the modes  $\phi_k$  extracted from the CR subset for  $Ra = 5 \cdot 10^7$  for the  $Pr$  considered (see §4). For the different  $Ra$ , we project the instantaneous velocity and temperature fields onto the six leading POD modes  $\phi_k$  to obtain the set of amplitudes  $(\alpha_k, k = 1, \dots, 6)$ . As in §5, these modal amplitudes are then scaled by a common factor which is the value  $1/|L|$  for  $Ra = 5 \cdot 10^7$ . Afterwards we compute  $\bar{\alpha}_k \alpha_k$ , which is proportional to the part of the combined energy  $E_{comb}$  contained inside mode  $\phi_k$  (see section 2.3). We check the statistical convergence by splitting each series in half and comparing their mean values. For the values of  $(Ra, Pr)$  considered, energy contained inside the six POD modes is somewhere between 80% and 95% of the total energy  $E_{comb}$ , while the maximum relative difference is about 2%.

Let us focus on the energetic contents of the leading POD modes  $L$ ,  $Q$ , and  $S$  (figure



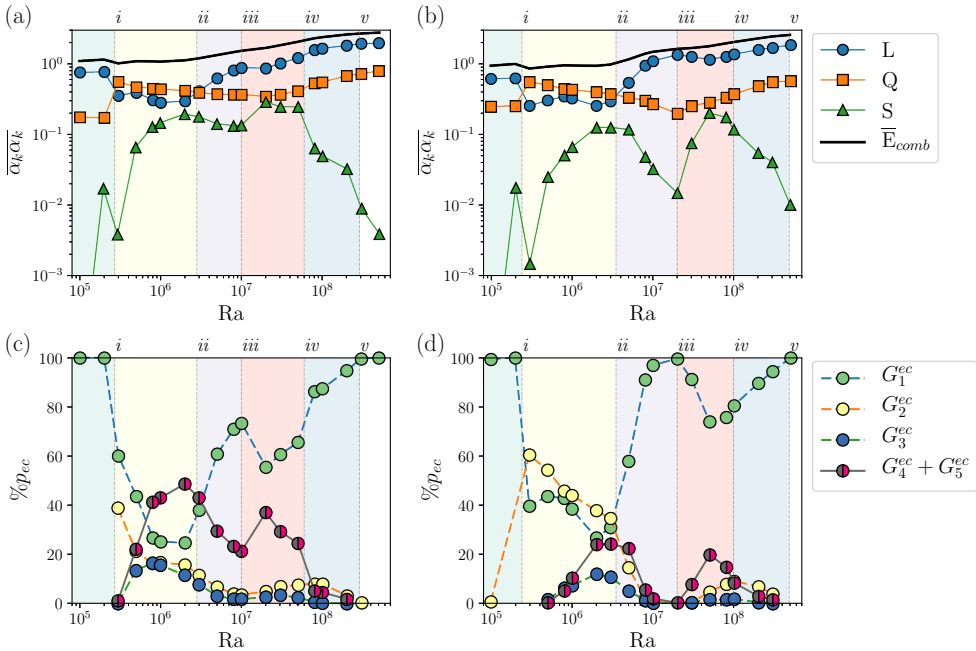


FIGURE 17. Dynamical regimes as a function of  $Ra$  for (a,c)  $Pr = 3$  and (b,d)  $Pr = 4.3$ . Marks  $i$  to  $v$  separate the different dynamical regimes (see text). Figs. (a,b) display the rescaled values of  $\overline{\alpha_k \alpha_k}$ , where  $\alpha_k$  is obtained by projection into CR modes for  $Ra = 5 \cdot 10^7$ . Combined energy  $\overline{E}_{comb}$  is shown for reference. Figs. (c,d) display the probability  $\%p_{ec}$  of the system being inside the cluster groups:  $G_1^{ec}$  is predominantly associated to plateaus;  $G_2^{ec}$  to transitions between plateaus;  $G_3^{ec}$  to short excursions; and  $G_4^{ec}$  and  $G_5^{ec}$  to extended cessations.

17), which contain the main features of both the CR and EC regimes. As function of  $Ra$ , we may identify different dynamical regimes: below mark  $i$  mode  $L$  is dominant; between marks  $i$  and  $ii$  mode  $Q$  is dominant; above  $ii$  mode  $L$  becomes again dominant and remains so for larger values of  $Ra$ ; between  $iii$  and  $iv$  mode  $S$  becomes comparable to  $Q$ ; above  $iv$  mode  $S$  decreases with respect to modes  $L$  and  $Q$ . Note that, on the figure, we add a mark  $v$  which will be used in the next section. In the following, since both values of  $Pr$  display similar trends, we focus on the case for  $Pr = 3$ .

In order to highlight the relation between the LSC dynamics and the energetic contents of POD modes, figure 18 displays a time series of  $A_{2D}$  for different values of  $Ra$  representing each dynamical regime. For values of  $Ra$  between the onset convection and mark  $i$ , a steady-state roll is observed. The jump of mode  $Q$  around mark  $i$  coincides with the transition to unsteady flows. Region  $i$ - $ii$  corresponds to continuous reversals of the LSC but without the accumulation phase, i.e. modes  $Q$  and  $L$  have the same dominant frequency and plateaus in  $A_{2D}$  are completely absent. Examples of this zone are a periodic solution like at  $Ra = 3 \cdot 10^5$  (figure 18a) or when mode  $S$  becomes more energetic, chaotic reversals like at  $Ra = 2 \cdot 10^6$  (figure 18b). Above mark  $ii$ , mode  $L$  becomes dominant once again. This coincides with the development of metastable plateaus in  $A_{2D}$  (figure 18c). Inside region  $iii$ - $iv$ , as mode  $S$  becomes more energetic again, plateaus become shorter while the CR and EC regimes are observed intermittently (figure 18d). Finally above mark  $iv$ , as mode  $S$  weakens again, plateaus size gradually increase and reversals become more regular (figure 18e). Once mode  $S$  becomes negligible, stable plateaus are observed for as long as 2000 convective time units (figure 18f).

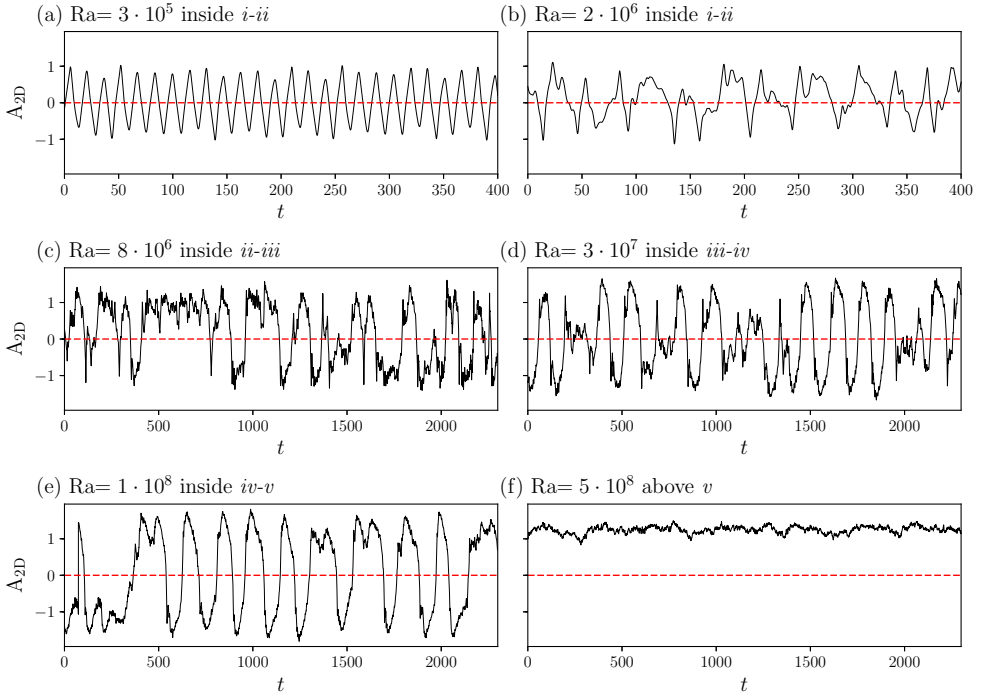


FIGURE 18. Time series of the angular impulse  $A_{2D}$  for  $Pr = 3$  and different  $Ra$  illustrating the different flow regimes: (a) periodic reversals for  $Ra = 3 \cdot 10^5$ , (b) chaotic reversals for  $Ra = 2 \cdot 10^6$ , (c) appearance of long plateaus for  $Ra = 8 \cdot 10^6$ , (d) flow reversals and extended cessations for  $Ra = 3 \cdot 10^7$ , (e) flow reversals for  $Ra = 1 \cdot 10^8$ , and (f) long plateaus for  $Ra = 5 \cdot 10^8$ .

This dependency on  $(Ra, Pr)$  of the flow dynamics is consistent with observations from Sugiyama *et al.* (2010) where flow reversals are observed over two discontinuous ranges of  $Ra$  for  $Pr \geq 4.3$ . The plateau length seems to be influenced by the energetic contents of mode  $S$ , with longer plateaus observed as mode  $S$  weakens, which is more pronounced for  $Pr = 4.3$  than for  $Pr = 3$  (see number of reversals in table 1). For instance, for  $(Ra = 10^7, Pr = 4.3)$  only 3 reversals were observed over 9600 t.u. For this case, the waiting time between reversals can be more than 300 times larger than for the region  $i-ii$ , and up to 50 times shorter than for the region  $iii-iv$ .

### 8.2. Residence time inside different cluster groups

The second approach uses the cluster partitioning for each value of  $Pr$ . To allow for a direct comparison between different  $Ra$ , we use the clusters extracted from the EC subset at  $Ra = 5 \cdot 10^7$ . This partitioning recovers the main features of the regime of extended cessations, but also of flow reversals. For a given  $Ra$ , we assign each data-point  $(\alpha_k, k = 1, \dots, 6)$  computed in §8.1 to one of the 5 groups  $G_k^{ec}$  presented in §7.1. Figure 19 displays the  $(L, S)$  phase-diagrams for the Rayleigh numbers corresponding to cases of figure 18. From this procedure, the residence time  $\%p_{ec}$  of each group  $G_k^{ec}$  can be evaluated (see figure 17c for  $Pr = 3$  and figure 17d for  $Pr = 4.3$ ).

First, consider the region  $i-ii$ . For  $Ra = 3 \cdot 10^5$ , the periodic solution corresponds to a closed path along the  $S$ -axis which passes through the cluster groups  $G_1^{ec}$  and  $G_2^{ec}$  (figure 19a). Data-points inside  $G_1^{ec}$  correspond to the diagonal rolls observed during plateaus, while  $G_2^{ec}$  corresponds to the quadrupolar mode observed during the transition between

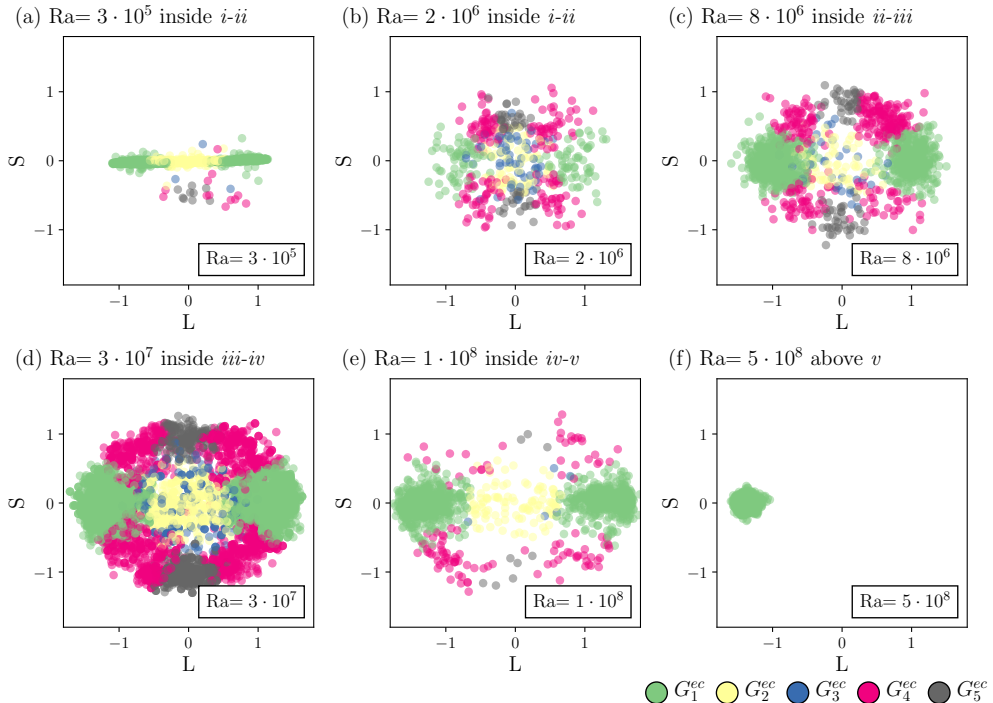


FIGURE 19. Phase diagram in the  $(L, S)$  plane for  $Pr = 3$  and different  $Ra$  illustrating the different flow regimes presented in figure 18. Cluster groups are obtained from the EC regime for  $Ra = 5 \cdot 10^7$ , see colour code at the bottom of the figure:  $G_1^{ec}$  is predominantly associated to plateaus;  $G_2^{ec}$  to transitions between plateaus;  $G_3^{ec}$  to short excursions; and  $G_4^{ec}$  and  $G_5^{ec}$  to extended cessations.

positive and negative plateaus. By contrast, the chaotic reversals observed inside the same region, do not display any particular orbit and explore all of the groups (figure 19b). In this region, we observe the emergence of cluster groups  $G_4^{ec}$  and  $G_5^{ec}$ , associated to cessation dynamics, which become dominant for the higher  $Ra$ . For all groups the residence time  $\%p_{ec}$  has the same order between 20% and 40% (figure 17d).

Inside region  $ii-iii$ ,  $G_1^{ec}$  becomes dominant at the expense of all the other groups, in particular of  $G_2^{ec}$ , illustrating the relative scarcity of flow reversals (figures 17d and 19c). Inside region  $iii-iv$ , the group  $G_1^{ec}$  becomes less prominent in favour of  $G_2^{ec}$  (indicating shorter plateaus), and in favour of  $G_4^{ec}$  and  $G_5^{ec}$  (associated to cessation dynamics) (figure 17d). The sum of residence times of  $G_4^{ec}$  and  $G_5^{ec}$  is as large as 40%, which illustrates the strong competition between the CR and EC regimes (figure 19d). Inside the region  $iv-v$ ,  $G_1^{ec}$  and  $G_2^{ec}$  are reinforced while cessation dynamics become rare or non-existing (figures 17d and 19e). Finally, as  $Ra$  increases above  $v$ , groups other than  $G_1^{ec}$  essentially disappear and only a long stable plateau remains (figures 17d and 19f).

A comparison of the residence times between the two  $Pr$ , suggests that a higher  $Pr$  reduces the occurrence of cessation dynamics. This effect is more pronounced inside the regions inside  $i-iv$ . This is consistent with results from Chandra & Verma (2013) for  $Pr = 1$ , where the Fourier mode  $[1,2]$  (roughly equivalent to mode  $S$ ) is dominant below  $Ra = 10^7$ . Additionally, the ratio between the Fourier mode  $[2,2]$  (roughly equivalent to  $Q$ ) and the  $[1,1]$  mode (equivalent to  $L$ ) of the vertical velocity was found to decrease monotonically for  $Ra$  between  $2 \cdot 10^7$  and  $10^9$ . A similar decrease is observed inside the

region *ii-iii*. However, above *iii* the ratio between modes  $Q$  and  $L$  reaches a near constant value.

## 9. Conclusion

In this paper we used long-term data from two-dimensional DNS of a square RB cell to extract and study the dynamics of the large-scale structures. In the range considered, intermittency between the regime of consecutive reversals (CR) and the regime of extended cessations (EC) is observed. Initially, we split, based on a criterion depending of the global angular momentum, the overall data into two subsets, one for the CR and one for the EC regimes. For each subset, a POD analysis is performed identifying the six most energetic large-scale flow structures. Both regimes however may be accurately represented by the same type of large-scale flow structures, suggesting that differences between the CR and EC regimes concerns, for the most part, the energetic distribution of these flow structures and more generally their time evolution. This suggests that both regimes are driven by different physical processes.

As for the energetic content, the leading modes in the CR regime are centrosymmetric monopole  $L$  and quadrupole  $Q$ , while the leading mode for the EC regime is the symmetry-breaking vertical dipole  $S$ . For both regimes, the leading mode amplitude exhibits alternating positive and negative plateaus, the duration of plateaus, known as the waiting time, being several times longer than the large-eddy turnover time.

As for the time evolution, we performed a separate analysis of the probability distribution function of waiting times between consecutive reversals. During the EC regime, the PDFs of waiting times for all reversing modes  $\{L, L^*, S, S^*\}$  correspond well to exponential distribution: this indicates that the system loses its memory during the EC regime. During the CR regime, this is also true for the symmetry-breaking modes  $S$  and  $S^*$ . By contrast, the waiting times for the reversing mode  $L$  (and to a lesser extent  $L^*$ ) strongly deviates from a Poissonian process. In this view, modes  $\{S, S^*\}$  become a source of noise for an otherwise predominantly coherent dynamics, as identified in (Podvin & Sergent 2017). This feature is confirmed by applying a conditional average over hundreds of realisations of particular reversals in the CR. This leads to the identification of the most common dynamical features of the POD amplitudes during a generic reversal. Centrosymmetric modes  $\{L, L^*, Q, Q^*\}$  are found to follow a generic cycle, whereas the symmetry-breaking modes  $\{S, S^*\}$  fluctuate around zero. The above observations are consistent with results obtained experimentally in a cylinder (Xi & Xia 2008a), where the strength of the rebound and the time period to the next reversal have been demonstrated to be correlated, despite a Poissonian process of occurrence of the reorientations.

A simplified representation based on a cluster partitioning reduces the problem to a finite network dynamics, which enables the study of the long-term evolution of large-scale structures. Such an approach not only recovers the main features of the CR regime previously obtained through a simple ensemble averaging, but also reveals that flow reversals may take one of several possible paths with equal probability during the beginning of the release phase leading to the rebound event.

The clustering method was then used to study the EC regime which has been much less explored. It discriminates between different coherent dynamical events regardless of their relative timescales: weak reversals and actual cessations. On the one hand, weak reversals deviate from standard reversals due to the intensity in mode  $S$ . On the other hand, the actual cessations are predominantly associated to a continuous push and shove motion between two vertically stacked vortices. This metastable state may be sustained for long

periods of time. Moreover, short excursions are also identified despite their shortness and scarcity, corresponding to the brief appearance of an intense plume crossing the cavity.

As a final analysis, we combined POD and cluster analysis to provide a description of the different spatio-temporal dynamics observed over a wide range of Rayleigh numbers. By considering the energetic content of the leading POD modes  $\{L, Q, S\}$  and the residence time in each group of clusters, we identified the successive dynamical regimes observed from the steady state to the turbulent regime.

## Acknowledgements

This work was granted access to the HPC resources of GENCI-IDRIS under allocation 2a0326 made by GENCI.

## REFERENCES

- ANTONIA, R. A. 1981 Conditional sampling in turbulence measurement. *Annual review of fluid mechanics* **13** (1), 131–156.
- ARTHUR, D. & VASSILVITSKII, S. 2007 k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035. Society for Industrial and Applied Mathematics.
- BAI, K., JI, D. & BROWN, E. 2016 Ability of a low-dimensional model to predict geometry-dependent dynamics of large-scale coherent structures in turbulence. *Physical Review E* **93** (2), 023117.
- BAILON-CUBA, J., EMRAN, M. S. & SCHUMACHER, J. 2010 Aspect ratio dependence of heat transfer and large-scale flow in turbulent convection. *Journal of Fluid Mechanics* **655**, 152–173.
- BELL, J.B., COLELLA, P. & GLAZ, H.M. 1989 A second-order projection method for the incompressible Navier-Stokes equations. *Journal of Computational Physics* **283**, 257–283.
- BROWN, E. & AHLERS, G. 2006 Rotations and cessations of the large-scale circulation in turbulent rayleigh-bénard convection. *Journal of Fluid Mechanics* **568**, 351–386.
- BROWN, E., NIKOLAENKO, A. & AHLERS, G. 2005 Reorientation of the large-scale circulation in turbulent Rayleigh-Bénard convection. *Physical Review Letters* **95**, 084503.
- CASTILLO-CASTELLANOS, A., SERGENT, A. & ROSSI, M. 2016 Reversal cycle in square Rayleigh-Bénard cells in turbulent regime. *Journal of Fluid Mechanics* **808**, 614–640.
- CHANDRA, M. & VERMA, M. K. 2011 Dynamics and symmetries of flow reversals in turbulent convection. *Physical Review E* **83** (6), 067303.
- CHANDRA, M. & VERMA, M. K. 2013 Flow Reversals in Turbulent Convection via Vortex Reconnections. *Physical Review Letters* **110** (11), 114503.
- DAS, A., GHOSAL, U. & KUMAR, K. 2000 Asymmetric squares as standing waves in Rayleigh-Bénard convection. *Physical Review E* **62** (3), R3051.
- FARANDA, D., PODVIN, B. & SERGENT, A. 2019 On reversals in 2D turbulent Rayleigh-Bénard convection: Insights from embedding theory and comparison with proper orthogonal decomposition analysis. *Chaos* **29**, 033110.
- FAUVE, S., HERAULT, J., MICHEL, G. & PÉTRÉLIS, F. 2017 Instabilities on a turbulent background. *Journal of Statistical Mechanics: Theory and Experiment* p. 064001.
- FOROOZANI, N., NIEMELA, J. J., ARMENIO, V. & SREENIVASAN, K. R. 2017 Reorientations of the large-scale flow in turbulent convection in a cube. *Physical Review E* **95** (3), 033107.
- GIANNAKIS, D., KOLCHINSKAYA, A., KRASNOV, D. & SCHUMACHER, J. 2018 Koopman analysis of the long-term evolution in a turbulent convection cell. *Journal of Fluid Mechanics* **847**, 735–767.
- GROSSMANN, S. & LOHSE, D. 2003 On geometry effects in Rayleigh-Bénard convection. *Journal of Fluid Mechanics* **486**, 105–114.
- HOLMES, P., LUMLEY, J. L., BERKOOZ, G. & ROWLEY, C. W. 2012 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, 2nd edn. Cambridge University Press.

- HORN, S. & SCHMID, P. J. 2017 Prograde, retrograde, and oscillatory modes in rotating Rayleigh–Bénard convection. *Journal of Fluid Mechanics* **831**, 182–211.
- HUGHES, G. O., GAYEN, B. & GRIFFITHS, R. W. 2013 Available potential energy in Rayleigh–Bénard convection. *Journal of Fluid Mechanics* **729**, R3.
- HUSSAIN, A.K.M. F. 1986 Coherent structures and turbulence. *Journal of Fluid Mechanics* **173**, 303–356.
- JAIN, A. K. 2010 Data clustering: 50 years beyond k-means. *Pattern Recognition Letters* **31** (8), 651 – 666.
- KADANOFF, L. P. 2001 Turbulent heat flow: Structures and scaling. *Physics today* **54** (8), 34–39.
- KAISER, E., NOACK, B. R., CORDIER, L., SPOHN, A., SEGOND, M., ABEL, M., DAVILLER, G., STH, J., KRAJNOVI, S. & NIVEN, R. K. 2014 Cluster-based reduced-order modelling of a mixing layer. *Journal of Fluid Mechanics* **754**, 365–414.
- LHULLIER, F., HULOT, G. & GALLET, Y. 2013 Statistical properties of reversals and chrons in numerical dynamos and implications for the geodynamo. *Physics of the Earth and Planetary Interiors* **220**, 19–36.
- LUMLEY, J. L. 1967 The structure of inhomogeneous turbulent flows. *Atmospheric turbulence and radio wave propagation*.
- McFADDEN, P.L. & MERRILL, R.T. 1986 Geodynamo energy source constraints from palaeomagnetic data. *Physics of the Earth and Planetary Interiors* **43** (1), 22 – 33.
- MERRILL, R., McELHINNY, M. & McFADDEN, P. 1998 Chapter five reversals of the earth’s magnetic field. In *The Magnetic Field of the Earth: Paleomagnetism, the Core, and the Deep Mantle, International Geophysics*, vol. 63, pp. 163 – 215. Academic Press.
- MOLENAAR, D., CLERCX, H. J. H. & VAN HEIJST, G. J. F. 2004 Angular momentum of forced 2D turbulence in a square no-slip domain. *Physica D: Nonlinear Phenomena* **196**, 329–340.
- NI, R., HUANG, S.-D. & XIA, K.-Q. 2015 Reversals of the large-scale circulation in quasi-2D Rayleigh–Bénard convection. *Journal of Fluid Mechanics* **778**, R5.
- NIEMELA, J.J., SKRBEK, L., SREENIVASAN, K.R. & DONNELLY, R.J. 2001 The wind in confined thermal convection. *Journal of Fluid Mechanics* **449**, 169–178.
- OKABE, A., BOOTS, B., SUGIHARA, K. & CHIU, S. N. 2009 *Spatial tessellations: concepts and applications of Voronoi diagrams*, , vol. 501. John Wiley & Sons.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. 2011 Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.
- PETSCHER, K., WILCZEK, M., BREUER, M., FRIEDRICH, R. & HANSEN, U. 2011 Statistical analysis of global wind dynamics in vigorous Rayleigh–Bénard convection. *Physical Review E* **84** (2), 026309.
- PODVIN, B. & SERGENT, A. 2015 A large-scale investigation of wind reversal in a square Rayleigh–Bénard cell. *Journal of Fluid Mechanics* **766**, 172–201.
- PODVIN, B. & SERGENT, A. 2017 Precursor for wind reversal in a square Rayleigh–Bénard cell. *Physical Review E* **95** (1), 013112.
- VAN DER POEL, E. P., STEVENS, R.J.A.M. & LOHSE, D. 2011 Connecting flow structures and heat flux in turbulent Rayleigh–Bénard convection. *Physical Review E* **84** (4), 045303.
- POPINET, S. 2003 Gerris: a tree-based adaptive solver for the incompressible euler equations in complex geometries. *Journal of Computational Physics* **190** (2), 572–600.
- POPINET, S. 2009 An accurate adaptive solver for surface-tension-driven interfacial flows. *Journal of Computational Physics* **228** (16), 5838–5866.
- POPINET, S. 2015 A quadtree-adaptive multigrid solver for the Serre–Green–Naghdi equations. *Journal of Computational Physics* **302**, 336–358.
- SCHMID, P. J. 2010 Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics* **656**, 5–28.
- SHRAIMAN, B. I. & SIGGIA, E. D. 1990 Heat transport in high-Rayleigh-number convection. *Physical Review A* **42** (6), 3650–3653.
- SIROVICH, L. 1987 Turbulence and the dynamics of coherent structures part I: coherent structures. *Quarterly of applied mathematics* **45** (3), 561–571.

- SREENIVASAN, K. R., BERSHADSKII, A. & NIEMELA, J.J. 2002 Mean wind and its reversal in thermal convection. *Physical Review E* **65** (5), 056306.
- SUGIYAMA, K., NI, R., STEVENS, R.J.A.M., CHAN, T. S., ZHOU, S.-Q., XI, H.-D., SUN, C., GROSSMANN, S., XIA, K.-Q. & LOHSE, D. 2010 Flow reversals in thermally driven turbulence. *Physical Review Letters* **105** (3), 034503.
- VALET, J.-P., FOURNIER, A., COURTILOT, V. & HERRERO-BERVERA, E. 2012 Dynamical similarity of geomagnetic field reversals. *Nature* **490** (7418), 89–93.
- VAN HEIJST, G. J. F., CLERCX, H. J. H. & MOLENAAR, D. 2006 The effects of solid boundaries on confined two-dimensional turbulence. *Journal of Fluid Mechanics* **554**, 411–431.
- VASILEV, A. Y. & FRICK, P. G. 2011 Reversals of large-scale circulation in turbulent convection in rectangular cavities. *JETP letters* **93** (6), 330–334.
- WICHT, J., STELLMACH, S. & HARDER, H. 2009 *Numerical Models of the Geodynamo: From Fundamental Cartesian Models to 3D Simulations of Field Reversals*, pp. 107–158. Berlin, Heidelberg: Springer Berlin Heidelberg.
- WINTERS, K. B., LOMBARD, P. N., RILEY, J. J. & D’ASARO, E. A. 1995 Available potential energy and mixing in density-stratified fluids. *Journal of Fluid Mechanics* **289**, 115–128.
- XI, H.-D. & XIA, K.-Q. 2007 Cessations and reversals of the large-scale circulation in turbulent thermal convection. *Physical Review E* **75** (6), 066307.
- XI, H.-D. & XIA, K.-Q. 2008a Azimuthal motion, reorientation, cessation, and reversal of the large-scale circulation in turbulent thermal convection: A comparative study in aspect ratio one and one-half geometries. *Physical Review E* **78** (3), 036326.
- XI, H.-D. & XIA, K.-Q. 2008b Flow mode transitions in turbulent thermal convection. *Physics of Fluids* **20** (5), 055104.
- XI, H.-D., ZHOU, Q. & XIA, K.-Q. 2006 Azimuthal motion of the mean wind in turbulent thermal convection. *Physical Review E* **73** (5), 056312.
- ZHANG, X. & ZIKANOV, O. 2015 Twodconvection turbulent convection in a toroidal duct of a liquid metal blanket of a fusion reactor. *Journal of Fluid Mechanics* **779**, 36–52.

## Appendix A. Cluster partitioning method

This method consists in partitioning a set of six-dimensional data points (in our case, a set of six POD amplitudes taken every  $\delta t$ ), into a finite set of  $K$  clusters. Each cluster  $c_k$  of points, is characterised by the average position  $\boldsymbol{\mu}_k$  of the ensemble of points in  $c_k$ . This point  $\boldsymbol{\mu}_k$ , called the centroid, corresponds to a typical flow pattern seen as a representative state of cluster  $c_k$ .

For this work, the cluster partitioning is performed *via* the K-means algorithm (see, for instance Jain (2010)) as implemented by the *scikit-learn* Python package (Pedregosa *et al.* 2011). This algorithm can be described as an optimisation method: given a number  $K$  of clusters and a set of data points, it is tasked to find a Voronoi partition (see, for instance Okabe *et al.* (2009)) of these data points so that the sum of the squared error

$$\hat{J}(c_1, \dots, c_K) = \sum_{k=1}^K \sum_{\boldsymbol{\alpha}' \in c_k} \|\boldsymbol{\alpha}' - \boldsymbol{\mu}_k\|^2 \quad \text{with } \boldsymbol{\alpha}' = (\alpha'_1, \alpha'_2, \dots, \alpha'_6) \quad (\text{A } 1)$$

between the centroid of a cluster, and its associated points is minimized over all sets of  $K$  clusters. The minimum squared error is denoted by  $J(K)$ . Starting from an initial guess of the cluster centroids  $\boldsymbol{\mu}_k^{i=0}$  obtained using *k-means++* (Arthur & Vassilvitskii 2007), one applies an iterative procedure. At step  $n$  of this procedure, one obtains a given value  $\boldsymbol{\mu}_k^{i=n}$  and performs two actions: first one builds a new Voronoi partition based on  $\boldsymbol{\mu}_k^{i=n}$ , second the centroids of this new partition  $\boldsymbol{\mu}_k^{i=n+1}$  are computed using the mean values of all the points of the new partition clusters. Steps are repeated until the current and subsequent centroids converge.

Because of the discrete flow symmetries  $\{\mathbb{S}_x, \mathbb{S}_y, \mathbb{R}_\pi\}$ , each cluster is *a priori* bound to

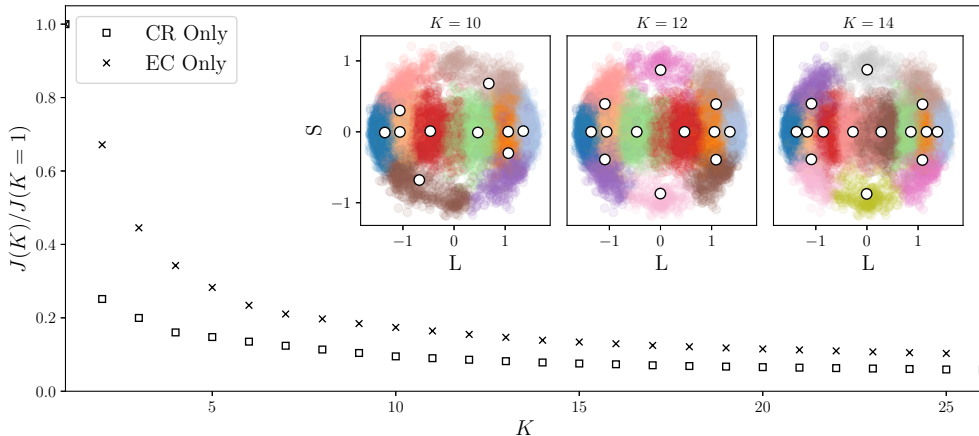


FIGURE 20. Normalised squared error  $J(K)/J(K = 1)$  as function of the number of clusters  $K$  for  $(Ra = 5 \cdot 10^7, Pr = 3)$  inside the CR and EC regimes. The insets illustrate the clusters (in colour) and the cluster centroids (in white marks) of the partitioned POD subspace for the CR regime using different values of  $K$ .

be included in one of three possible groups: *(i)* groups of 4 clusters, where none of the elements of the group satisfy a particular symmetry but each element is transformed into another member of the group by the action of  $\mathbb{S}_x$ ,  $\mathbb{S}_y$ , or  $\mathbb{R}_\pi$ ; *(ii)* groups of 2 clusters, where each element of the group is invariant by the action of one of the symmetries among the set  $\{\mathbb{S}_x, \mathbb{S}_y, \mathbb{R}_\pi\}$  and is transformed into another member of the group by the remaining symmetries; and *(iii)* groups of a single cluster, which is invariant by the action of  $\{\mathbb{S}_x, \mathbb{S}_y, \mathbb{R}_\pi\}$ . However, since the K-means algorithm is not constrained by symmetry, it sometimes leads to an optimum set of  $K$  clusters which does not verify the symmetry rules. If this is the case, we disregard this result.

Since  $J(K)$  always decreases as we increase  $K$ , one is required to fix the number  $K$  of clusters. This choice can be done based on the decrease of the minimum squared error  $J(K)$  with respect to  $K$  (like in figure 20 for instance). In our case, we also take advantage of the symmetric nature of the POD subspace to guide the choice of  $K$ . An adequate number of clusters  $K$  results from a good compromise between a small number of clusters, the decrease in the squared error  $J(K)$ , and also on the condition that the computed clusters verify the required symmetry rules.

For  $(Ra = 5 \cdot 10^7, Pr = 3)$  the partitioning the CR subset using  $K = 12$  is deemed satisfactory: this choice results in a decrease in the squared error by more than 90% with respect to the case with  $K = 1$  (figure 20). Smaller values of  $K$  fail to satisfy the required symmetries (see for instance,  $K = 10$  in the left inset in figure 20) or to recover important details such as the double-roll modes associated to cessations, while a larger value provides a marginal gain of information about the phases of the generic cycle (see for instance,  $K = 14$  in the right inset in figure 20). Similarly, partitioning the EC subset using  $K = 11$  clusters results in a decrease in the squared error by more than 80% with respect to the case  $K = 1$  (figure 20). Smaller values of  $K$  fail to recover the short excursions through mode  $S^*$  while larger values provide redundant information on the cessation dynamics.