

Flow morphology and patterns in porous media convection: A persistent homology analysis

Marco De Paoli^{1,2} , Sergio Pirozzoli³ , Catherin Neena Lalu⁴ and Lou Kondic⁵ 

¹Institute of Fluid Mechanics and Heat Transfer, TU Wien, 1060 Vienna, Austria

²Physics of Fluids Group and Max Planck Center for Complex Fluid Dynamics and J. M. Burgers Centre for Fluid Dynamics, University of Twente, PO Box 217, 7500AE Enschede, The Netherlands

³Dipartimento di Ingegneria Meccanica e Aerospaziale, Sapienza Università di Roma, Rome, Italy

⁴Department of Physics, New Jersey Institute of Technology, Newark, NJ 07102, USA

⁵Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: Marco De Paoli, marco.de.paoli@tuwien.ac.at

(Received 25 December 2025 UTC; revised 9 June 2026 UTC; accepted 14 June 2026 UTC)

Convective mixing in porous media is crucial in both geophysical and industrial fields, spanning applications ranging from carbon dioxide sequestration to geothermal energy extraction. Key processes are affected by convective heat transport or diffusion of chemical species in porous formations. Intense convection flow and mixing create complex, dynamic patterns that are difficult to predict and measure. The present work focuses on the use of topological data analysis, in particular, the measures emerging from the growing field of persistent homology (PH), to quantify these patterns. These measures are objective and quantify structures across all temperature or concentration values simultaneously. These techniques, when applied to classical porous media set-ups, such as one-sided and Rayleigh–Bénard flow configurations, provide new insights into a system's structure, flow patterns and macroscopic mixing properties. Using large datasets we make publicly available, comprising original simulations as well as those presented in previous works, we correlate the behaviour of the heat transport rate (quantified by the Nusselt number) with the evolution of the flow structures (quantified by the PH measures). Finally, we provide a detailed analysis of the flow evolution over a wide range of governing parameters, namely the Rayleigh–Darcy number and the domain size.

Key words: pattern formation, convection in porous media, geophysical and geological flows

1. Introduction

Transport of heat and chemical species in porous media is relevant to natural and industrial flows: from latent heat thermal energy storage systems (Trelles & Duflý 2003; Xu *et al.* 2017) to the formation of sea ice (Feltham *et al.* 2006; Wells, Hitchen & Parkinson 2019), several key processes are controlled by the redistribution of heat and solutes in confined domains. When the motion is driven by density gradients within the fluid layer, and the density field depends on the local distribution of the scalar (e.g. temperature or solute concentration), the flow is controlled by natural convection. In porous-media convection, the governing equations for thermally driven and solute-driven flows are equivalent once expressed in dimensionless form, with temperature and solute concentration playing analogous roles as buoyancy-driving scalars. As a result, insights obtained from one formulation can be directly transferred to the other. In the present work, we adopt a thermal formulation of the problem, which provides a convenient and widely used framework while retaining full relevance for solutal convection applications. In convective flows, local density differences are contrasted by the dissipative mechanisms of friction (due to narrow pore spaces) and diffusion (which reduce the gradients of the scalar field), which, in turn, affect the flow field. The relative importance of driving (convection) and dissipative (diffusion, friction) mechanisms is quantified by the Rayleigh–Darcy number Ra (hereinafter defined as Rayleigh number). A similar dynamics occurs in the presence of key geophysical systems, where buoyancy is induced by solute concentration gradients rather than temperature variations, e.g. geothermal flows in underground sites (Hu, Xu & Yang 2023), thawing of permafrost (Wang *et al.* 2025), dispersion of contaminants in groundwater flows (Simmons, Fenstermaker & Sharp 2001; De Paoli *et al.* 2025*b*) and storage of carbon dioxide (CO₂) in saline aquifers. The latter, in particular, has been extensively studied in recent decades, due to its enormous relevance in mitigating the effects of climate change (Metz *et al.* 2005). Geological sequestration of CO₂ involves injecting large amounts of CO₂ into underground geological porous formations for permanent storage (Emami-Meybodi *et al.* 2015; Jin 2024). A key question in assessing the suitability of potential sequestration sites is determining the CO₂ mixing rate in brine.

Although the physical motivation for this study is rooted in solutal convection relevant to CO₂ sequestration, the analysis presented here is carried out using a thermal formulation. Owing to the formal equivalence between thermal and solutal convection in porous media, all results can be directly reinterpreted in terms of mass transfer, with the dimensionless governing and response parameters retaining the same physical meaning. The archetypal system used to analyse this problem is the one-sided (or semi-infinite) configuration (Hidalgo *et al.* 2012): the flow is idealised as a rectangular box, initially filled with low-density fluid (representing brine). At the top boundary (representing the CO₂–brine interface) the fluid density is maximum (Jin 2024). At the bottom boundary, the domain is assumed to be closed due to the presence of an impermeable rock layer. The flow dynamics that takes place is marked by a transient behaviour, controlled by the convective flow structures that populate the system. Numerical simulations have been extensively deployed to analyse these flows (Hewitt, Neufeld & Lister 2013; Slim 2014; Wen *et al.* 2018). Despite these efforts, several research questions remain unanswered. For sufficiently large domains and high values of Rayleigh numbers, the flow evolution is still transient, but it is also independent of the domain size and Ra (De Paoli *et al.* 2025*d*). What is the minimal flow unit required to achieve such a self-similar behaviour? And what is the effect of the domain size on the mixing properties of the flow? In this work, we answer these questions by systematically analysing the effects of the domain size and of the Rayleigh number Ra of the emerging flow patterns.

The one-sided configuration described above presents a major drawback: it is intrinsically transient, which markedly complicates any theoretical analysis of the flow. In contrast, the ‘two-sided’ (Rayleigh–Bénard) configuration, in which the fluid density is fixed at the top and also at the bottom, attains a (statistically) steady state, which makes it more suitable for theoretical analysis. In addition, it has been shown (Hewitt *et al.* 2013) that the two-sided flow is directly related to the one-sided configuration during the late-stage dynamics, indicated as the ‘shutdown of convection’ and corresponding to a well-mixed bulk flow configuration. The two-sided system is well defined in terms of boundary conditions, and during the statistically steady state, it is possible to derive exact equations that describe the flow transport properties as a function of the governing parameters (Grossmann & Lohse 2000, 2001; Lohse & Shishkina 2024). For these reasons, a number of works (Hewitt, Neufeld & Lister 2012; Hewitt, Neufeld & Lister 2014; Wen, Corson & Chini 2015; De Paoli, Zonta & Soldati 2016; Pirozzoli *et al.* 2021; De Paoli *et al.* 2022; Hu *et al.* 2023; Zhu, Fu & De Paoli 2024) have investigated the two-sided configuration. However, several questions remain unanswered in this case as well. The near-wall flow structure organises into multiple hierarchical levels, from the small-scale (near-wall) plumes nesting into large-scale structures (labelled as ‘supercells’; De Paoli *et al.* 2022), which have been only partially characterised by analysing scalar fields after applying *ad hoc* threshold filtering. Can the supercells be identified more robustly and univocally? Can their formation be predicted based on the values of the domain size and Ra considered? Can a Ra -independent description of the near-wall small-scale structures be derived? What is the flow structure at small and large Ra in the presence of large domains?

As a result of the complex interplay between diffusion and buoyancy, the flow structures organise into tangled, dynamic patterns, making it extremely challenging to predict and quantify their behaviour. Initial efforts in this direction include using Fourier-based analysis (Hewitt *et al.* 2014; Pirozzoli *et al.* 2021; De Paoli *et al.* 2025*d*) as well as utilising cell sizes as a measure of emerging patterns (Fu, Cueto-Felgueroso & Juanes 2013; De Paoli *et al.* 2022). However, there is much more that can be done in this direction. One approach that appears appropriate is based on topological measures, which provide an efficient and detailed analysis of structures characterised by a value of a scalar field (in this case, solute concentration). While several approaches of varying complexity can be implemented for this purpose, measures based on the exploration of the connectivity of the considered structures appear to be the most appropriate. Such approaches, based on a computational topology discipline, persistent homology (PH), have been used to analyse a variety of systems emerging from materials science, such as granular matter (Kondic *et al.* 2012; Kramár *et al.* 2014*a,b*; Taghizadeh *et al.* 2024), Rayleigh–Bénard convection (Kramár *et al.* 2016), suspensions (Gameiro *et al.* 2020) and porous media (Suzuki *et al.* 2021), among others. One advantage of using PH-based measures is that they are objective in the sense that they do not require specifying a threshold value (such as average temperature or a similar quantity), but instead, they quantify the structure across all thresholds at once. We will see that PH-based approaches provide significant new information about the structures developing in the considered system, their connection to heat transport, and in particular, that they allow for answering some questions formulated so far.

The paper is organised as follows. In §2, we present the governing equations and flow configurations considered, and discuss the equivalence between thermal and solutal formulations in porous-media convection. In §3, we provide an overview of the PH methods used to quantify the emerging patterns. The results of the one-sided and two-sided

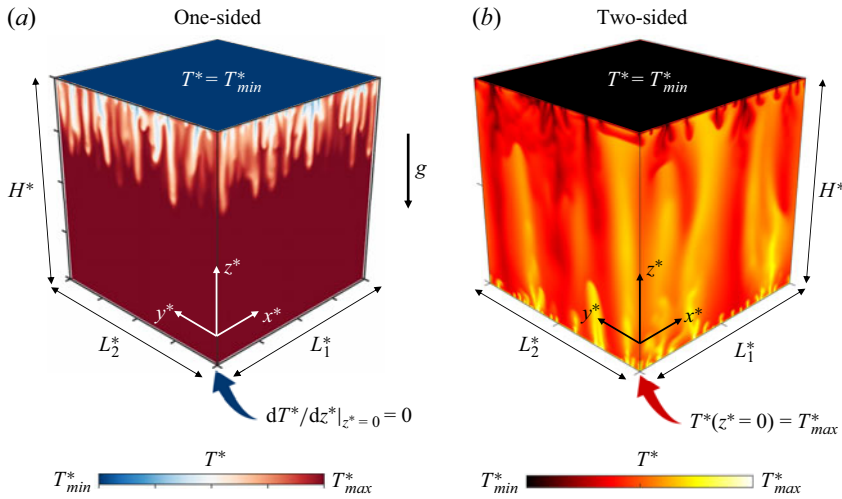


Figure 1. Temperature distribution over the surface of the domain, with indication of the dimensional extension in each direction (L_1^* , L_2^* , H^*) and of the boundary conditions. (a) One-sided flow configuration, discussed in § 4. (b) Two-sided flow configuration, discussed in § 5.

systems analysed are presented in §§ 4 and 5, respectively. Finally, an overview of the work and conclusions are presented in § 6.

2. Governing equations

We consider convective porous media flows at the Darcy scale, i.e. the flow properties averaged over a reference elementary volume (Whitaker 1998; Nield & Bejan 2017). In particular, we focus on density-driven flows in which the source of buoyancy is due to a temperature-induced density field, and with the solid locally in thermal equilibrium with the liquid phase. We consider thermally driven flows, as they are relevant to both the one-sided and two-sided configurations (Hu *et al.* 2023). However, the same dimensionless equations apply to solutal convection. The governing equations in dimensional and dimensionless form are presented in §§ 2.1 and 2.2, respectively, and the data analysed are discussed in § 2.3.

2.1. Problem formulation

We consider a three-dimensional (3-D) porous domain with uniform porosity ϕ and permeability K . The medium is fully saturated with fluid, whose density depends on the temperature field, T^* . The local gradients of fluid density within the system drive the flow. Figure 1 shows a sketch of the domain; here, two different flow configurations, discussed in detail in the following, are described.

We indicate by x^* and y^* the horizontal directions and by z^* the vertical direction (perpendicular to the walls) along which the gravitational acceleration g is aligned. The temperature field T^* , averaged over the reference elementary volume, varies between T_{min}^* and T_{max}^* , and it is controlled by the advection–diffusion equation (Nield & Bejan 2017):

$$[(1 - \phi)(\rho c)_s + \phi(\rho c)_f] \frac{\partial T^*}{\partial t^*} + (\rho c)_f \mathbf{u}^* \cdot \nabla^* T^* = \nabla^* \cdot [(1 - \phi)\lambda_s + \phi\lambda_f] \nabla^* T^*, \quad (2.1)$$

where s and f refer to the solid and fluid phases, respectively, c is the specific heat (considered at constant pressure for the fluid) and λ the thermal conductivity (the superscript $*$ indicates dimensional variables). In this case, we consider a system in local thermal equilibrium ($T_f = T_s = T$) and without heat production. In addition, when the heat capacity of the solid is negligible ($(\rho c)_s \rightarrow 0$), (2.1) reduces to (Hewitt 2020)

$$\phi \frac{\partial T^*}{\partial t^*} + \mathbf{u}^* \cdot \nabla^* T^* = \nabla^* \cdot (\kappa \nabla^* T^*), \quad (2.2)$$

where t^* is time, $\mathbf{u}^* = (u^*, v^*, w^*)$ is the volume-averaged velocity field, ϕ is the porosity of the medium and $\kappa = [(1 - \phi)\lambda_s + \phi\lambda_f]/(\rho c)_f$ is the thermal diffusivity of the medium, considered constant here. The fluid density, ρ^* , is assumed to depend linearly on temperature:

$$\rho^*(T^*) = \rho^*(T_{min}^*) - \Delta\rho^* \frac{T^* - T_{min}^*}{T_{max}^* - T_{min}^*}, \quad (2.3)$$

with $\Delta\rho^* = \rho^*(T_{min}^*) - \rho^*(T_{max}^*)$ being the maximum density contrast within the domain. Assuming the validity of the Boussinesq approximation (Landman & Schotting 2007), the flow field is described by the continuity and the Darcy equations,

$$\nabla^* \cdot \mathbf{u}^* = 0, \quad \mathbf{u}^* = -\frac{K}{\mu} (\nabla^* P^* + \rho^* g \mathbf{k}), \quad (2.4)$$

respectively, with μ the (constant) fluid viscosity, P^* the pressure and $\mathbf{k}$ the vertical unit vector. The walls are considered impermeable to the fluid, i.e. the following boundary conditions for the flow field apply:

$$w^*(z^* = 0) = 0, \quad w^*(z^* = H^*) = 0. \quad (2.5)$$

Note that a slip at the walls is possible, due to the Darcy formulation adopted. Periodicity is considered across the vertical boundaries (directions x , y) for all flow variables.

Regarding the temperature field, at the upper and lower walls, both Dirichlet (T^* fixed) or Neumann ($\partial T^*/\partial z^*$ fixed) boundary conditions are considered:

- (i) Semi-infinite domain: the temperature is fixed at the top and no-flux condition is assumed at the bottom (see figure 1a), namely

$$T^*(z^* = H^*) = T_{min}^*, \quad \left. \frac{\partial T^*}{\partial z^*} \right|_{z^*=0} = 0. \quad (2.6)$$

In addition, the domain is initially filled with still fluid and at temperature T_{max}^* . Thus, the velocity and temperature fields are initialised as $\mathbf{u}^* = 0$ and

$$T^*(x^*, y^*, z^* < H^*, t_0^*) = T_{min}^* + (T_{max}^* - T_{min}^*) \operatorname{erf} \left[\sqrt{\frac{\phi}{\kappa}} \frac{(H^* - z^*)}{2\sqrt{t_0^*}} \right], \quad (2.7)$$

respectively. The initial condition (2.7) is obtained evaluating at time t_0^* the self-similar solution of (2.2) in the absence of convection ($\mathbf{u}^* = 0$) in a semi-infinite domain ($H^* \rightarrow \infty$) (De Paoli, Zonta & Soldati 2017). The instant considered to initialise the simulation is $t_0^* = 250\phi\kappa(g\Delta\rho^*K/\mu)^{-2}$, corresponding to a dimensionless time of $\hat{t}_0 = 250$ or $t_0 = 250/Ra$, expressed in diffusive or convective units, respectively (see § 2.2 for further details). Finally, a random perturbation (white noise) is added. The amplitude of this noise controls the time at which the plumes form, i.e. the onset of convection occurs (De Paoli *et al.* 2025d). This flow

configuration, also labelled as ‘one-sided’, is representative of a domain cooled from above and insulated from below, and will be analysed in § 4.

- (ii) Two-sided configuration: the temperature is imposed at both boundaries (see figure 1b), such that the corresponding density profile is unstable:

$$T^*(z^* = H^*) = T_{min}^*, \quad T^*(z^* = 0) = T_{max}^*. \quad (2.8)$$

The velocity field is again initialised as $\mathbf{u}^* = 0$, while a linear profile is used for the temperature distribution,

$$T^*(x^*, y^*, z^*, t^* = 0) = T_{max}^* - \frac{z^*}{H^*}(T_{max}^* - T_{min}^*), \quad (2.9)$$

corresponding to the initial condition labelled as IC1 in Hewitt *et al.* (2014). This configuration, labelled as ‘Rayleigh–Bénard’ or ‘two-sided’, is analysed in § 5.

2.2. Dimensionless equations

Depending on whether the domain height H^* or an intrinsic diffusive length scale ℓ^* (defined later) is used to make the flow quantities dimensionless, two different formulations can be derived (convective or diffusive units, both used in this work), described in §§ 2.2.1 and 2.2.2, respectively.

2.2.1. Convective units

Natural flow scales relevant to the convective system considered are the buoyancy velocity, $\mathcal{U}^* = g\Delta\rho^*K/\mu$, and the domain height, H^* . Using the following set of dimensionless variables:

$$T = \frac{T^* - T_{min}^*}{T_{max}^* - T_{min}^*}, \quad x = \frac{x^*}{H^*}, \quad \mathbf{u} = \frac{\mathbf{u}^*}{\mathcal{U}^*}, \quad (2.10)$$

$$t = \frac{t^*}{\phi H^*/\mathcal{U}^*}, \quad p = \frac{p^*}{\Delta\rho^*gH^*}, \quad (2.11)$$

and introducing the reduced pressure $p^* = P^* - \rho^*(T_{min}^*)gz^*$, we obtain the dimensionless form of the governing equations (2.2)–(2.4):

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \frac{1}{Ra} \nabla^2 T, \quad (2.12)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.13)$$

$$\mathbf{u} = -(\nabla p - T\mathbf{k}), \quad (2.14)$$

where

$$Ra = \frac{g\Delta\rho^*KH^*}{\kappa\mu} = \frac{\mathcal{U}^*H^*}{\kappa}. \quad (2.15)$$

The flow is controlled by three dimensionless parameters: the Rayleigh number Ra and the domain width in horizontal directions, $L_1 = L_1^*/H^*$ and $L_2 = L_2^*/H^*$.

A distinction should be made depending on the behaviour of the solid phase concerning the transported scalar. When the solid phase is impermeable to the scalar field considered, e.g. in the case of mass transport problems, the definition (2.15) of the Rayleigh number changes as $Ra = \mathcal{U}^*H^*/(\kappa\phi)$. However, the dimensionless formulation of the problem is the same for both permeable and impermeable behaviour, provided that when the

permeable case is considered, the two phases are in local thermal equilibrium. Further details on this matter are provided by De Paoli (2023).

2.2.2. Diffusive units

An alternative non-dimensionalisation approach involves using $\ell^* = \kappa/\mathcal{U}^*$ as the reference length scale. This strategy was introduced by Fu *et al.* (2013) and is suitable for describing local flow dynamics that is not influenced by the largest scales of the flow. Defining the dimensionless variables as

$$T = \frac{T^* - T_{min}^*}{T_{max}^* - T_{min}^*}, \quad \hat{x} = \frac{x^*}{\ell^*}, \quad \mathbf{u} = \frac{\mathbf{u}^*}{\mathcal{U}^*}, \quad (2.16)$$

$$\hat{t} = \frac{t^*}{\phi \ell^* / \mathcal{U}^*}, \quad \hat{p} = \frac{p^*}{\Delta \rho^* g \ell^*}, \quad (2.17)$$

we obtain the following dimensionless form of the governing equations (2.2)–(2.4):

$$\frac{\partial T}{\partial \hat{t}} + \mathbf{u} \cdot \hat{\nabla} T = \hat{\nabla}^2 T, \quad (2.18)$$

$$\hat{\nabla} \cdot \mathbf{u} = 0, \quad (2.19)$$

$$\mathbf{u} = -(\hat{\nabla} \hat{p} - T \mathbf{k}), \quad (2.20)$$

where $\hat{\cdot}$ indicates variables made dimensionless with respect to diffusive units. Note that in this case Ra does not appear explicitly in the equations; instead, it represents the dimensionless height of the system, $Ra = \mathcal{U}^* H^* / \kappa = H^* / \ell^*$.

2.3. Data analysed

The data analysed in this work consist of a collection of results presented in previous works (Pirozzoli *et al.* 2021; De Paoli *et al.* 2022, 2025d) and original simulations presented here for the first time (see Appendix A for a detailed description of the database). The relevant part of the data analysed is made freely available via De Paoli *et al.* (2025a).

The original simulations presented here (simulations A in table 1) are obtained as follows. The governing equations (2.12)–(2.14) are solved numerically with the aid of an (open-source) second-order finite-difference solver, AFiD-Darcy (De Paoli 2025), which was previously employed to simulate convective porous media flows (De Paoli *et al.* 2025b,c). The code comprises an efficient parallel solver for simulating convective, incompressible and wall-bounded flows in porous media. This solver is based on the initial version of AFiD developed for turbulent flows (Van Der Poel *et al.* 2015). The algorithm utilises a pressure-correction scheme in conjunction with an efficient fast-Fourier-transform-based solver. Additional details of the algorithm are described by De Paoli *et al.* (2025c).

3. Overview of topological methods used for quantification of the emerging patterns: PH

For the present purposes, one can think of PH as a tool describing a complicated spatial pattern in the form of so-called persistence diagrams (PDs). These diagrams are obtained by filtration, i.e. thresholding the scalar field of interest (such as temperature), in such a manner that only the areas characterised by the value above the specified threshold appear.

Then, the simplest PD, which we refer to as β_0 PD, essentially traces how the regions of the high value of the considered scalar field appear as a filtration level is decreased, or disappear as two regions merge (β_0 stands here for the zeroth Betti number that essentially counts the number of regions that do not involve holes). In two spatial dimensions (the case that will be of interest to us in the present work, since we will consider temperature distributions on planes extracted from 3-D domains), there are two Betti numbers, β_0 and β_1 , corresponding to components and loops, and therefore also two PDs (in the interest of simplicity of notation, we use, e.g. β_0 to refer to both Betti number (the number of components at a given threshold value), and to the concept of a connected component). In three dimensions, there is an additional β_2 , counting enclosed 3-D structures, and therefore also an additional PD. In the present work we focus on two-dimensional (2-D) structures only; the consideration of 3-D aspects of the results is left for future work.

We note, and will discuss later in this section, that PDs contain information that is significantly more detailed than the Betti numbers, since they reveal the connectivity of the components over all thresholds, while Betti numbers are threshold-dependent. The reader is referred to Kramár *et al.* (2014a,b) for a more extensive discussion of PDs in the context of materials science problems. We note that Betti numbers were already used for morphology characterisation (Blunt 2017); an alternative approach based on Euclidean distance transform quantifying pore geometry was considered in Suzuki *et al.* (2021).

3.1. Persistent homology: an example

Here we discuss a 2-D example. The basic geometric structures of interest are connected components and loops (holes), denoted by β_0 and β_1 , respectively. To illustrate the topological quantities describing such structures, we use the simple 2-D example shown in figure 2. The reader is referred to Kramár *et al.* (2013, 2014b) for a more in-depth discussion, along with simpler one-dimensional examples; additional toy examples in the context of granular media analysis can be found in Kramár *et al.* (2014a) and Taghizadeh *et al.* (2024) and in the context of porous media flow in Illingworth *et al.* (2026). Specifics related to computations of PDs are discussed later in § 3.2.

The example shown in figure 2 is a two-sided convection simulation from § 5, discussed later in the text (settings: $Ra = 500$, case C5 in table 2, field taken at the boundary layer), with the colour map showing the temperature. Figure 2 shows two threshold values, namely 0.5 in figure 2(a) and 0.3 in figure 2(b); the parts of the domain characterised by the temperature values lower than the specified threshold are not shown and appear light blue. By comparing figures 2(a) and 2(b), we observe a change in connectivity and structure as the threshold value changes.

The change in connectivity, although visually apparent, needs to be quantified. Such quantification can be established by considering PDs. Let us consider first the PD corresponding to components, β_0 PD, shown in figure 2(c) (see also supplementary movie S1 available at <https://doi.org/10.1017/jfm.2026.11804>). As one descends from high values of temperature to lower ones, separated areas of high temperature appear (are born); each of these areas is called a component. Decreasing the threshold causes these components to merge; whenever two components merge, the ‘younger’ one (born at a lower temperature value) disappears (dies). The coordinates of each point (generator) in β_0 PD specify birth (b) and death (d) temperature for each component; the difference between birth and death values is called lifespan, $\mathcal{L}$, and shows the range of temperatures over which a component ‘lived’ before merging with an older one. Therefore, the generators near the diagonal are associated with minor spatial variations in temperature. All key trends were verified to

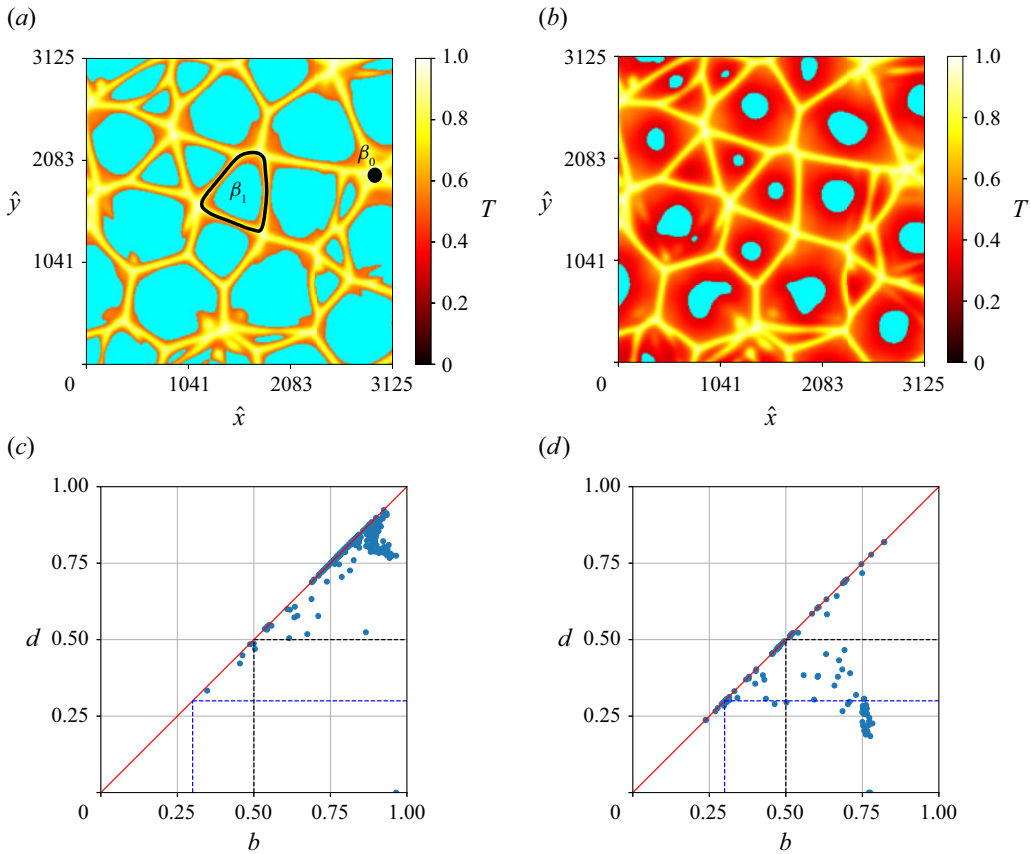


Figure 2. A 2-D example illustrating PDs for a function of two spatial coordinates, x and y (here we use the temperature results that will be discussed later in the text). The results for chosen thresholds of $T = 0.5$ (a) and $T = 0.3$ (b) – the light-blue areas are those where T is less than the specified threshold value. Persistence diagrams corresponding to (c) β_0 (components) and to (d) β_1 (holes), and schematic examples of which are reported in (a); here $T = 0.5$ (black dashed line) and $T = 0.3$ (blue dashed line) show the threshold levels. The fields shown in (a,b) are two snapshots from simulation C5 (see table 2 for additional details). We refer to supplementary movie S1 for a visual interpretation of the threshold-dependent temperature field and of the corresponding PDs.

be insensitive to variations of the persistence threshold and to the number of temporal samples (Appendix B.2), confirming that the observed behaviour is not an artefact of post-processing choices. In contrast, those far from the diagonal are related to components that remained separated from others over an extensive range of temperatures (one could use a landscape analogy and say that there is a deep ‘valley’ surrounding the local temperature maximum). One generator that is treated slightly differently is the one that corresponds to the component that appeared first (at the largest value of temperature); this component never dies and therefore has a death coordinate equal to zero. Inspecting carefully β_0 PD shows that this particular generator has the birth coordinate close to unity in the example shown in figure 2. Large ‘clump’ of generators in β_0 PD, with coordinates close to (0.8, 0.8), shows that there are multiple regions of the temperatures larger than 0.8 which merged at similar values of temperature threshold; checking figure 2(a), we observe that these generators are due to large temperatures of the narrow regions connecting the temperature ‘nodes’ (inspecting supplementary movie S1 shows that this is indeed the

case; the advantage of considering the PD instead is that this information is provided in a much simpler form).

Still considering β_0 PD, but shifting our focus to lower temperature values shows much less activity in the particular example shown in [figure 2](#), suggesting there are few topological changes at these temperatures. This diagram also shows that there are no topological changes (considering components only) below the temperature value of ≈ 0.3 ; the temperature field is fully connected. The reader familiar with the concept of Betti numbers (counting simply the number of components) will note that Betti number β_0 can be trivially obtained from the β_0 PD; the opposite is, however, not true, since a PD contains significantly more detailed information about connectivity of the considered scalar field than the corresponding Betti number. Most importantly, a PD includes the information over all thresholds (temperature values), while Betti numbers are threshold-dependent. Furthermore, PDs are stable with respect to noise (small perturbations), while the same cannot be claimed for Betti numbers (Kramár *et al.* 2014b). We also note in passing that some works present the information contained in PDs using different visual representations, such as bar-codes (Carlsson 2009), or sometimes consider sub-level thresholding instead of super-level as considered here; the connection between different graphical representations and thresholding approaches is straightforward.

In the 2-D example considered in [figure 2](#), we also have β_1 PD, corresponding to holes (loops) in the temperature field, shown in [figure 2\(d\)](#). A loop appears (is born) when the temperature threshold is decreased sufficiently that a closed structure, enclosing the area of lower temperatures, forms. A loop dies when the threshold is decreased to the degree that the whole interior of the loop is filled up. As expected, both birth and death coordinates of loops are typically lower than those of components; this is observed by inspecting the two PDs shown in [figure 2\(c,d\)](#). Supplementary movie [S1](#) helps in fully grasping the process of loop formation and disappearance.

We conclude the discussion of the presented example by noting that PDs provide significant data reduction, as they condense the information about the complicated two-variable function $T(x, y)$ into a point cloud (PD). This data reduction entails some information loss, primarily due to the loss of geometrical information (PD includes connectivity information but not the size (spatial extent) of each component). If such information is of importance, then one wants to combine the information available by considering PDs by a complementary measure. Furthermore, although PDs provide reasonably complete information about the connectivity of the considered field, they are still point clouds, and one can come up with measures summarising a PD. While many elaborate approaches are being considered to achieve this goal, in this work we resort to a simple approach and account for three measures only (separately for β_0 and β_1): the number of generators N_g , the average lifespan $\mathcal{L}$ (the average distance of the generators to the diagonal) and total persistence $TP = \mathcal{L}N_g$. One could think of TP as a simple measure describing (admittedly in a vague manner) the ‘intensity’ of a considered PD. We also note that to focus on the main features of the results, we exclude from consideration generators with $\mathcal{L} < 0.01$, as the fluctuations of the temperature field on such a scale are not of interest. Such exclusion does not influence the statistics of the (relevant) long-lived generators. In §§ 4.2 and 4.3 we comment in more detail on the influence of the specific value (0.01) on the results.

3.2. Persistent homology: specifics

Over the past decade, several research groups have developed and released open-source software tools for computing PDs and related topological measures. All computations and

results presented in this work were carried out using the GUDHI library (GUDHI 2025). In our calculations, we use the results obtained from the simulations directly, without any additional post-processing. Consistent with the simulation set-up, we assume periodic boundary conditions when computing topological measures (though using non-periodic boundaries yields only minor changes in the results). As illustrated by our toy example, figure 2, we use super-level filtration (meaning that the features that appear above the chosen threshold value are considered). For the considered data, the use of computing resources is very reasonable; e.g. for one data slice of simulation B7 (see table 1 in Appendix A), the computing time to build a cubical complex and compute persistence is ≈ 1.41 s (on M2 Mac Studio with 128 GB of RAM). An example Python script for interacting with the GUDHI library is included in the supplementary material.

4. Results: one-sided convection

The one-sided flow configuration is representative of a domain that is cooled from above, insulated at the bottom and initially filled with hot fluid, $T(x, y, z, t = 0) = 1$, corresponding to the boundary conditions (2.6) and the initial condition (2.7), respectively. A sketch of the flow configuration with explicit indication of the boundary conditions is shown in figure 1(a), with quantities expressed in dimensional form. The flow (and then the evolution of the flow morphology) is controlled by the following dimensionless parameters: the domain size in horizontal directions ($\widehat{L}_1, \widehat{L}_2$) and the Rayleigh number (Ra). In this work, we investigate the influence of these parameters independently: First we describe in §4.1 the flow evolution in the case of large Ra and large domain size ($\widehat{L}_1 = \widehat{L}_2 = 10^4$), and then we analyse in detail in §4.2 how the flow morphology is influenced by the governing parameters, considering the effect of size and Ra independently. The details of all the cases analysed are listed in table 1 (see Appendix A), where it is indicated which simulations have been specifically performed for this study (simulations A, data available via De Paoli *et al.* (2025a)) and which are obtained from previous works (simulations B, presented by De Paoli *et al.* (2025d)).

4.1. Flow dynamics

The evolution of the flow in one-sided systems has been previously studied in two dimensions by Slim (2014) and in three dimensions by De Paoli *et al.* (2025d), among others (Slim *et al.* 2013; Hewitt *et al.* 2013; Fu *et al.* 2013; De Paoli *et al.* 2017; Wen *et al.* 2018; Dhar *et al.* 2022). The global response parameter representative of the state of the flow is the Nusselt number evaluated at the top boundary:

$$Nu_T = -\frac{1}{L_1 L_2} \int_0^{L_1} \int_0^{L_2} \frac{\partial T}{\partial z} \bigg|_{z=1} dx dy, \quad (4.1)$$

which is a dimensionless measure of the importance of convective relative to diffusive heat transport (Nield & Bejan 2017). The transient nature of these systems has been described and modelled in detail (see Hewitt *et al.* (2013), Slim (2014) and De Paoli *et al.* (2025d) and references therein). In the following, we provide a brief description of the flow dynamics during each phase of the mixing process, with emphasis on the convective flux Nu_T (reported in figure 3a) and on the near-wall flow structures (figure 3b–i). We consider the simulation labelled as A1 in table 1 (see Appendix A) to discuss this dynamics.

The domain is initially filled with hot ($T = 1$) fluid at rest ($\mathbf{u} = 0$); see figure 3(b) for the temperature distribution. Under these conditions, mixing is controlled by diffusion at the boundary layer located at the upper wall; an analytical self-similar solution for the

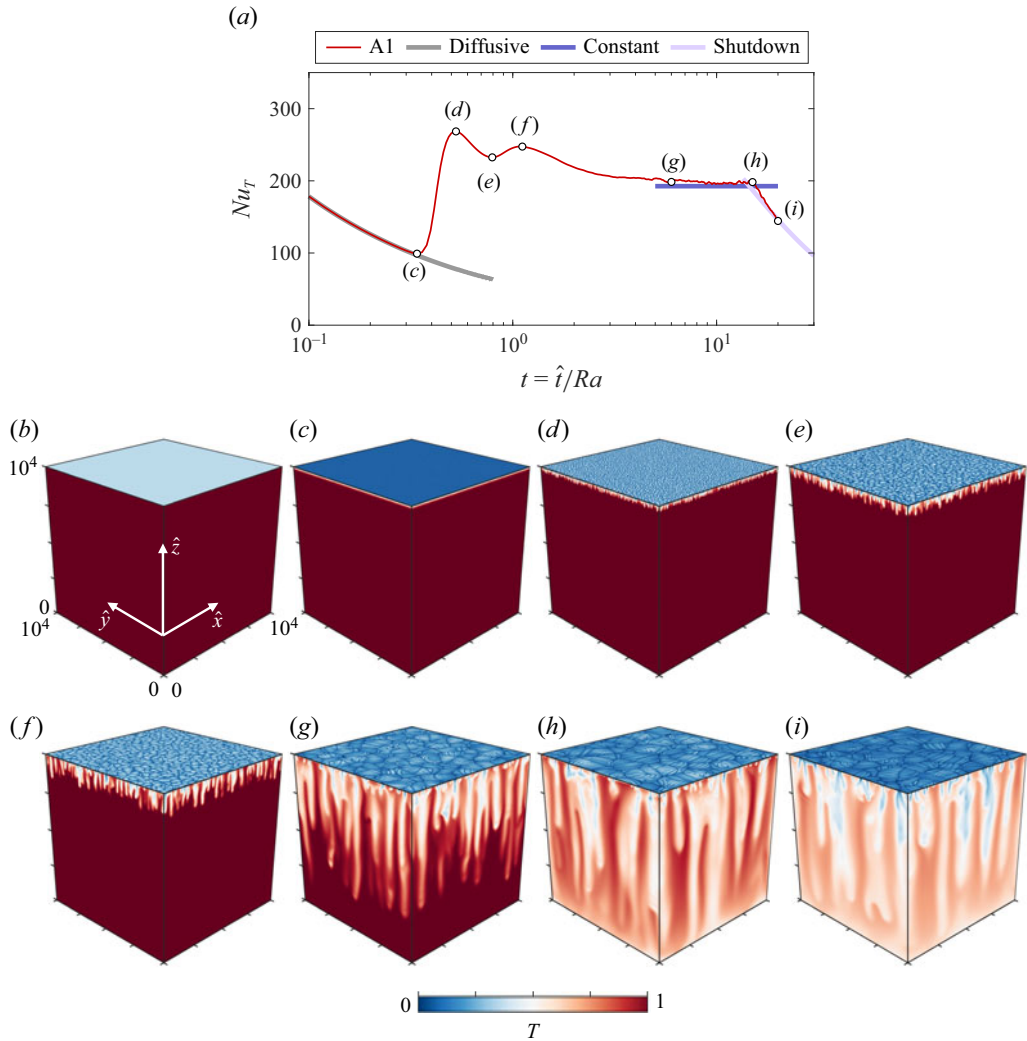


Figure 3. Evolution of a one-sided system (simulation A1; see [table 1](#) for details). (a) Flux at the top wall Nu_T , defined in (4.1). The flux is reported as a function of the convective time (note the logarithmic scale for the time variable). The analytical predictions of the flux during the initial (diffusive, (4.2)), intermediate (constant, (4.3)) and late (shutdown, (4.4)) stages are also reported. (b–i) Temperature distribution at different times, from the initial condition (b) to the shutdown of convection (i). The time instants corresponding to (c–i) are also indicated in (a). See § 4.1 for additional details.

evolution of the temperature distribution can be obtained (Slim 2014; De Paoli *et al.* 2025d), and in dimensionless convective units reads $T = \text{erf}[(1 - z)Ra/\sqrt{4tRa}]$. The corresponding evolution of Nu_T is inferred using (4.1), and it yields (De Paoli *et al.* 2017)

$$Nu_T = \left(\frac{\pi t}{Ra} \right)^{-1/2}. \quad (4.2)$$

The diffusive behaviour describes well that observed numerically and reported in [figure 3\(a\)](#) for early times ($t \leq 0.35$).

When a sufficiently thick layer of dense (cold) fluid forms beneath the top boundary, the initial temperature fluctuations amplify, leading to the formation of thermal plumes

(figure 3c) (Ennis-King, Preston & Paterson 2005; Riaz *et al.* 2006), which eventually grow and extend vertically bringing downward dense fluid, and enhancing transport due to convection (figure 3d) (Slim *et al.* 2013). Afterwards, the dynamics of the plumes stops being individual, and the interaction among neighbouring flow structures is such that small plumes merge into larger descending plumes (Riaz *et al.* 2006; Backhaus, Turitsyn & Ecke 2011). This phase occurs over several generations of plumes, and corresponds to a reduction of the flux (figures 3e and 3f). The space left by the plumes that merged is eventually filled with dense fluid (Slim 2014), and the process of plume formation and merging continues in a statistically steady fashion (figure 3g). Eventually, the plumes grow and reach the lower boundary (at $t \approx 7$), and the average fluid density in the domain increases progressively (i.e. the bulk flow temperature diminishes). During this regime, the Nusselt number is nearly constant and equal to (De Paoli *et al.* 2025d)

$$Nu_T = 0.01926 \times Ra. \quad (4.3)$$

At a later stage ($t \approx 14$ –15 in three dimensions; figure 3h), this process leads to an increase of the fluid density in the upper layer, corresponding to a reduction of the density contrast between the fluid in the upper layer and the upper boundary. This marks the start of the shutdown of convection (Hewitt *et al.* 2013): a sudden reduction of the driving force corresponds to a decrease of Nu_T , the evolution of which has been accurately predicted by physical models (Hewitt *et al.* 2013; Slim 2014; De Paoli *et al.* 2025d). We choose here to employ the formulation proposed by De Paoli *et al.* (2025d) (with the same parameters) to model the shutdown phase, given by

$$Nu_T = \frac{14.4 \times Ra}{[0.7303 \times (t - 14) + 27.0]^2}. \quad (4.4)$$

Also in this late-stage phase, the simulation and the model predictions are in excellent agreement (see figure 3a).

In what follows, we quantitatively analyse the morphology evolution of the flow structures near the upper boundary. First we consider the effect of the domain size for a fixed (and large) value of Ra , and then we investigate the role of Ra . In table 1 we report the details of the simulations analysed. Simulations indicated with ‘A’, are presented here for the first time, and are used to investigate the effect of the domain size. These simulations are performed at constant Rayleigh number, $Ra = 10^4$, and variable domain extension in horizontal directions, L_1 and L_2 . An overview of the system’s evolution as well as examples of the flow fields are shown in figures 3 and 4. For simulations indicated by ‘B’, initially presented in De Paoli *et al.* (2025d), the domains are of square cross-section. These data, obtained at different values of Ra between 10^2 and 8×10^4 , are used to investigate the effect of the driving parameter on the flow morphology. The domain extension is chosen to be sufficiently large, allowing for the neglect of any confinement- or periodicity-induced influence on the development of the flow structures. We refer to De Paoli *et al.* (2025d) for a detailed discussion on the minimal domain size to be employed at $Ra \geq 10^4$.

4.2. Connecting the flux to the flow structure next to the top wall

In this section, we focus on quantifying the morphology of the flow structure that develops next to the top and the flux throughout the domain, and on showing the connection between the measures describing this structure and the heat flux. We start by presenting an example of results obtained from PDs for one flow configuration (A1), and then proceed to illustrate the correlation between the computed topological measures and the heat flux across all considered configurations.

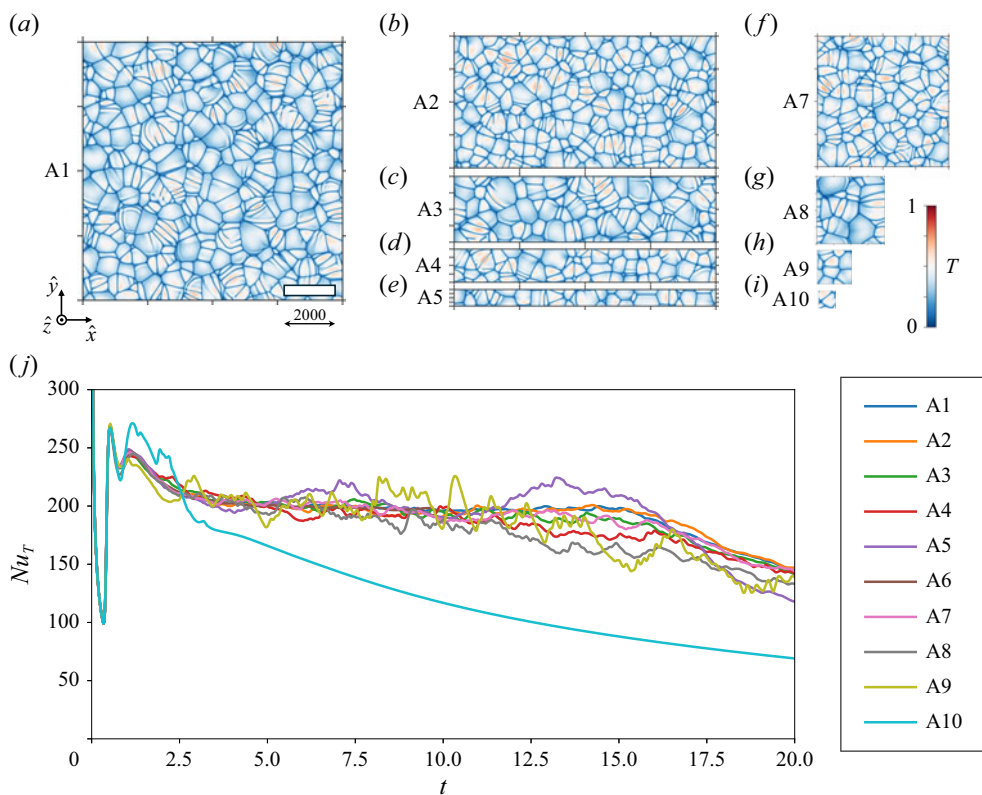


Figure 4. (a–i) Examples of temperature fields taken near the upper wall ($z = 0.998$) at time $t \approx 0.8$, for simulations A listed in table 1 with the exception of A6, the domain size of which is too large to be shown to scale. The full domain size is shown, and a scale bar (in diffusive units) is reported in (a) for reference. We refer to supplementary movie S2 for the time-dependent evolution of these patterns. (j) The time evolution of the flux at the top wall, Nu_T , defined by (4.1), for the domains considered and listed in table 1 (time is displayed in convective units). Note the different behaviour of the A10 configuration caused by the small domain size.

Figure 5 shows the evolution of pattern formation for simulation A1, focusing on early times. The related heat flux (Nusselt number) is also shown. We observe complex evolving structure, which is initially maze-like. As time passes, this structure gradually transitions into a well-defined cellular network that eventually stabilises. When comparing these pattern changes with the flux variations shown in figure 5(b), a clear relationship emerges: the evolution of the cellular pattern correlates closely with the changes in flux.

Our next goal is to quantify the structures shown in figure 5(a) and correlate them with the observed heat flux in figure 5(b). The PDs shown in figure 6 provide a detailed description of the structure at various time instances. These diagrams show the evolution of connected components, β_0 , and of loops (cycles), β_1 . Carefully comparing the temperature plots, figure 5(a), and corresponding diagrams, figure 6, illustrates the ability of PDs to quantify complex structures in a manner that captures the main topological features while also providing a clear visual and (as we will see) quantitative description of the results.

Focusing first on connected components, figure 6(a), we observe how the topological properties of the temperature patterns develop as time progresses. For very early times,

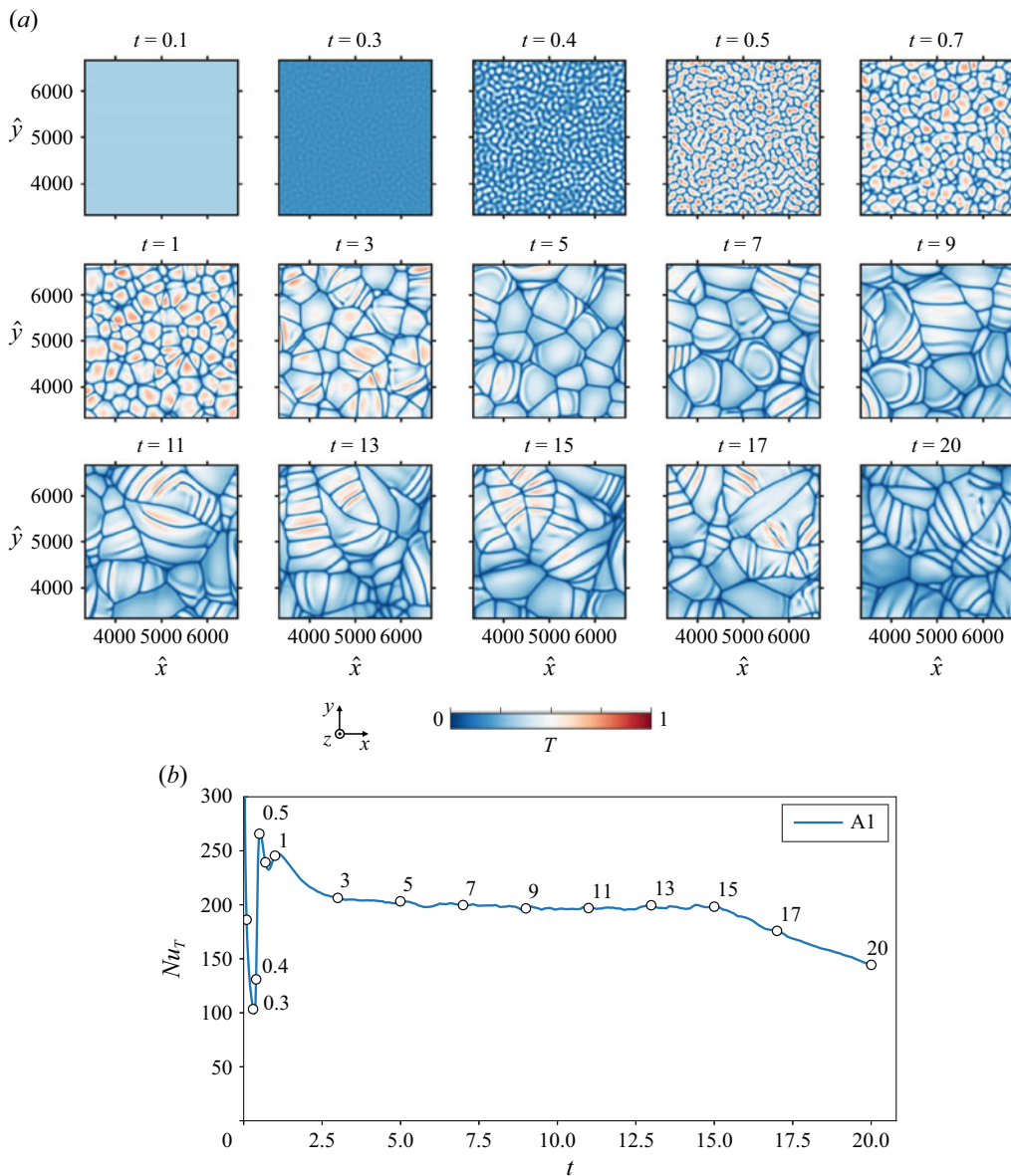


Figure 5. Pattern formation evolution for simulation A1. (a) Temperature distribution close to the top wall (a zoomed portion of the domain is shown, corresponding to $\hat{L}_1 \leq \hat{x}$, $\hat{y} \leq 2\hat{L}_1/3$); the plots' titles show the time instances at which the temperature field was recorded. (b) Heat flux measured by Nu_T . The time instances from (a) are shown as well.

$t \sim 0.1$, the temperature is nearly uniform and we note a single generator in the corresponding PD. More elaborate patterns start developing around $t = 0.3$ – 0.4 around the temperature $T \sim 0.3$, and then rapidly, by $t \sim 0.7$, the structures develop for the temperatures as high as $T \sim 0.7$. Around this time, the generators begin to gradually separate into upper and lower groups. The upper group (close to the diagonal) reflects

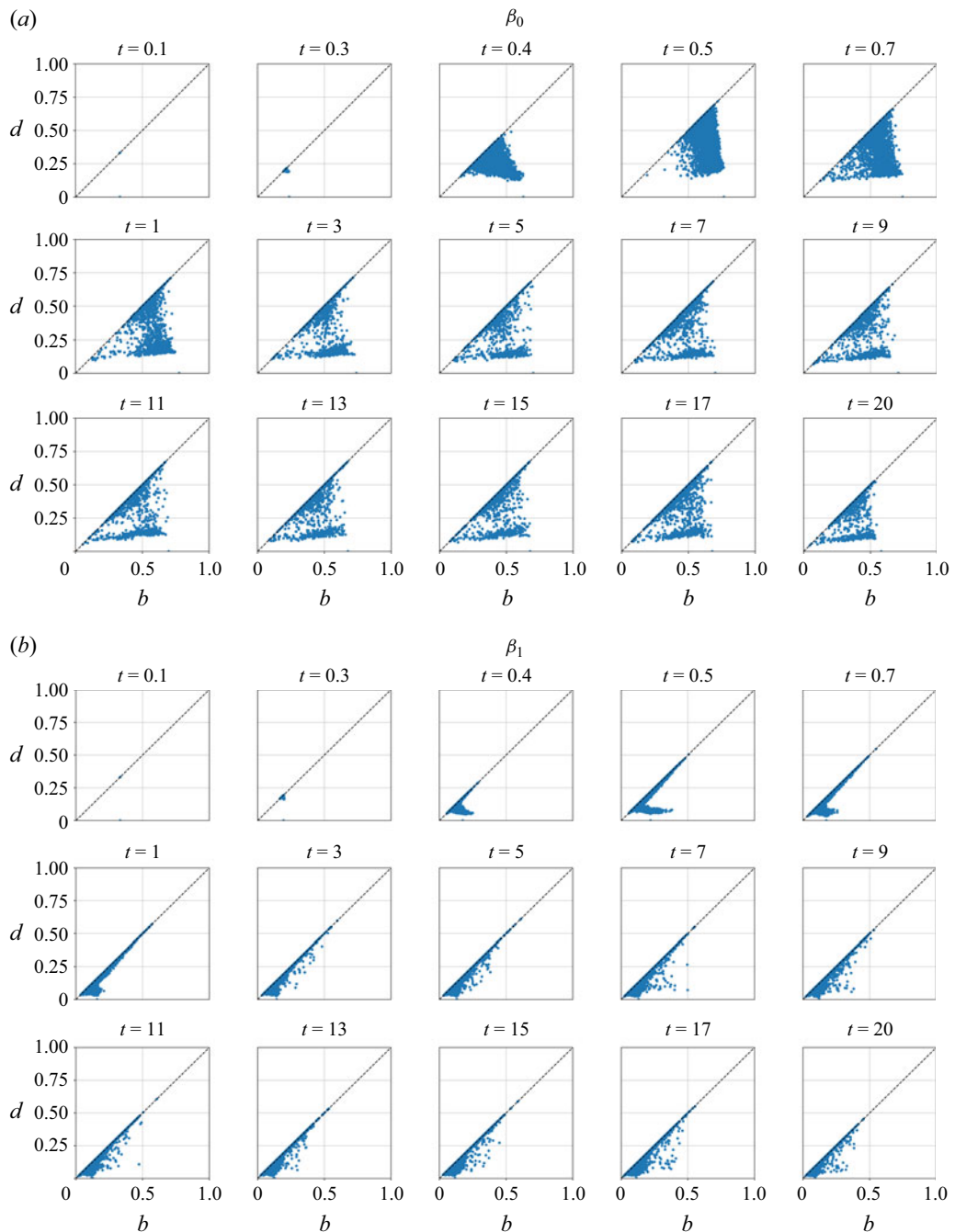


Figure 6. The PDFs corresponding to the temperature field shown in figure 5(a). (a) Connected components, β_0 , and (b) loops, β_1 . Here, ‘b’ stands for birth and ‘d’ for death of the considered features.

‘noise’ (due to small temperature variations, to be discussed further below), while the lower group corresponds to (hot) cell interiors, indicating that cells remain separated over a wide range of temperature thresholds (that is, the areas of high temperature appear already at $T \sim 0.7$, but they connect (merge) only at the temperatures $T \sim 0.2$).

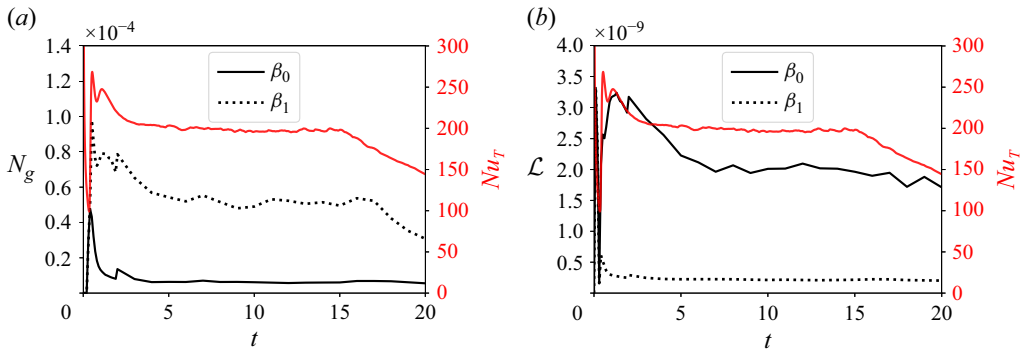


Figure 7. Flux and topological measures for simulation A1, normalised by the area ($\widehat{L}_1 \times \widehat{L}_2$). Flux is reported in red (right axis). Number of generators N_g (a) and average lifespan $\mathcal{L}$ (b) are indicated for components (β_0 , solid lines) and loops (β_1 , dotted lines).

Further insight is reached by considering loops (see figure 6b). Here we note the formation of loops with a significant birth temperature ($T \sim 0.3$) only at relatively early times ($t \lesssim 0.7$), as illustrated by the elongated tails in the corresponding PDs. As the pattern stabilises, these tails disappear, indicating that loops form only at very low temperatures. Overall, the diagrams capture the topological transition from the maze phase to an ordered cellular network.

Figure 7 demonstrates a strong similarity in the time-dependent behaviour of heat flux (figure 7a), total persistence (figure 7b), number of generators (figure 7c) and average lifespans (figure 7d). All considered quantities exhibit consistent trends, suggesting that the evolution of heat flux and topological measures is closely correlated. One peculiarity of the particular set of patterns considered here is illustrated by the number of generators, N_g , and lifespans, $\mathcal{L}$, for components (β_0) and loops (β_1): N_g is much larger for the loops, and $\mathcal{L}$ for the components. Taken together, these results suggest the existence of small variations of temperature leading to loop formation (very close to the diagonal, as suggested by small $\mathcal{L}$): these flow structures correspond to the footprint of the small plumes near the upper wall (see e.g. figure 3d–f). While Fourier analysis provides a dominant wavelength, it cannot distinguish between disconnected plumes and interconnected cellular networks with similar spectral content. Instead, PH resolves these differences through topological measures such as generator count and lifespan. On the other hand, $\mathcal{L}$ is significantly larger for components, showing the existence of isolated areas of elevated temperature which persist over a significant temperature range. These areas of elevated temperature are the feature of the results that is of real physical interest; regarding small variations of temperature, further inspection shows that these variations occur on the computational grid scale, consistently with the tail of the Fourier spectrum for the large wavenumbers for the results discussed in previous works (De Paoli *et al.* 2025d). They are therefore not relevant for our purposes and we remove them from future discussion, as mentioned in § 3.1. The noise band width used for this removal is a post-processing choice tied to the temperature-field scale and numerical resolution, and not a universal value. We note that the numerical resolution employed is appropriate and does not affect the main PH trends: extensive validation has been performed in previous works (De Paoli *et al.* 2022, 2025d), showing that numerical noise is well separated from physically meaningful flow features.

We now offer a more quantitative interpretation of the correlation between heat flux and topological metrics. Initially, when heat transfer is purely conductive, no generators are detected, neither components nor loops, and Nu_T diminishes as in (4.2). This process leads to a thickening of the thermal boundary layer beneath the interface. When the temperature fluctuations contained within grow sufficiently (leading to the formation of topological structures), the boundary layer becomes unstable and plumes form. These plumes remove heat from the top wall much more efficiently than conduction, and as a result, a sudden increase in the flux is observed. This process leads to the sudden emergence of topological generators with significant lifespans, linked to the formed plumes, followed by a subsequent reduction ($t > 1$). The increase in average lifespan reflects the progressive merging of plumes into larger coherent structures, which reduces interfacial area and modifies the effective transport pathways, thereby explaining the observed variation in Nusselt number. This dynamics suggests a rearrangement of the flow structures, i.e. a process of plumes merging, which also leads to a reduction of Nu_T : they cannot keep growing independently, as they interact with neighbouring plumes and with the hot rising fluid in the interplume spacing (Slim 2014). This dynamic constrains plume displacement and reduces the amount of heat that can be removed compared with an undisturbed plume. Using similar reasoning, one can conclude that when Nu_T is constant, there is no major difference in the flow morphology, i.e. the number of plumes and their organisation (as characterised by N_g and $\mathcal{L}$) remain statistically unchanged. Finally, for longer times ($t > 15$) during the shutdown phase, plumes reduce in number, and so does Nu_T due to the progressively reducing density contrast resulting from the saturation of the domain.

Figure 8 shows the results for all simulations A from table 1. While there is some noise in the results, we find that they are essentially independent of domain size for most domains considered. The main exception is simulation A10, which involves a very small domain.

4.3. Influence of the Rayleigh number on the pattern morphology

Here, we focus on the influence of the value of Ra , and therefore on simulations B. These data, obtained at different values of Ra ranging from 10^2 to 8×10^4 , are used to investigate the effect of the driving parameter on the flow morphology. The domain size is chosen to be sufficiently large to neglect any influence of confinement and periodicity on the development of the flow structures. We refer to De Paoli *et al.* (2025d) for a detailed discussion on the minimal domain size to be employed at $Ra \geq 10^4$.

Figure 9 shows the emerging patterns at time $\hat{t} = 5 \times 10^3$ ($t = 5 \times 10^3 / Ra$). We observe that, while the patterns for the simulations obtained with smaller values of Ra ($\lesssim 2000$, simulations B2–B5), there are significant differences between the patterns, for larger values of Ra , simulations B6–B9, the emerging patterns become more similar, at least visually.

Figure 10 puts the statements above on a firmer footing. This figure shows the topological measures obtained from the corresponding persistence diagrams for components and loops. The main observation is that for the considered measures (number of generators, N_g , and average lifespan, $\mathcal{L}$), the results for simulations B6–B10 are essentially identical, while for smaller values of Ra we see that the results depend on the value of Ra . This dependence is not obvious from the snapshot of the patterns, such

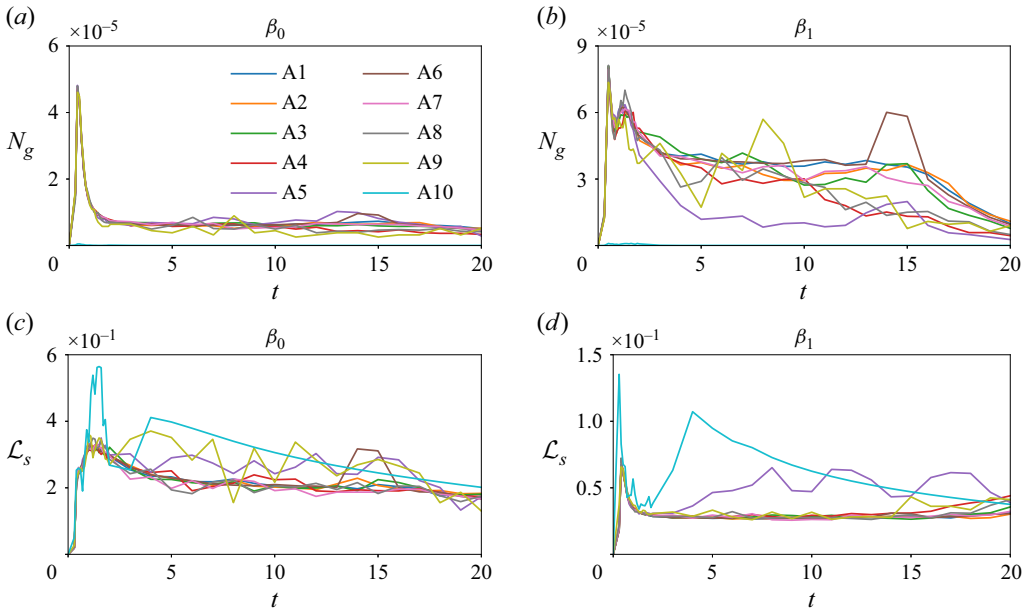


Figure 8. Number of generators N_g (a,b) (per area) and average lifespan $\mathcal{L}$ (c,d) (with band of width 0.01 removed), for simulations A1–A10. (a,c) Components (β_0) and (b,d) loops (β_1). Note that for simulation A10, N_g reaches only very small values (barely distinguishable from 0 on the scale plotted).

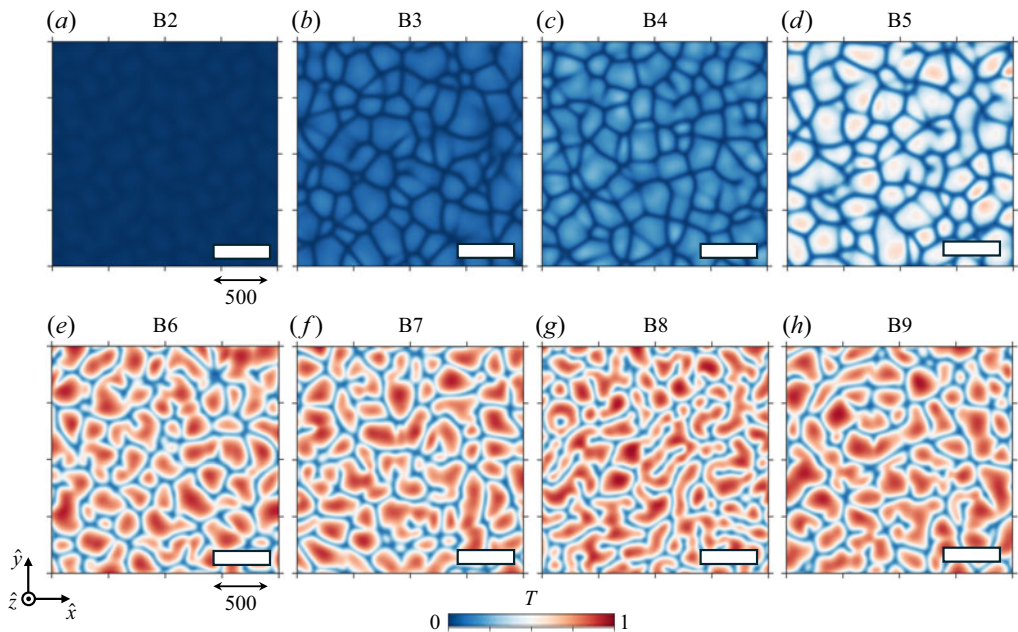


Figure 9. Temperature distribution over a horizontal ($\hat{x}$, $\hat{y}$) plane near the upper wall at time $\hat{t} = 5 \times 10^3$ for (a–h) simulations B2–B9 (see table 1). Only a squared portion of the domain having side 2×10^3 is reported (note that the temperature field shown for B2 is repeated in a periodic manner, because the original domain, of side 10^3 diffusive units, is smaller than the domain shown here, 2×10^3). A scale bar of length 500 is shown as a reference. The morphology of the flow appears to be dependent on the Rayleigh number considered for B2–B5, while no macroscopic change is observed for B6–B9.

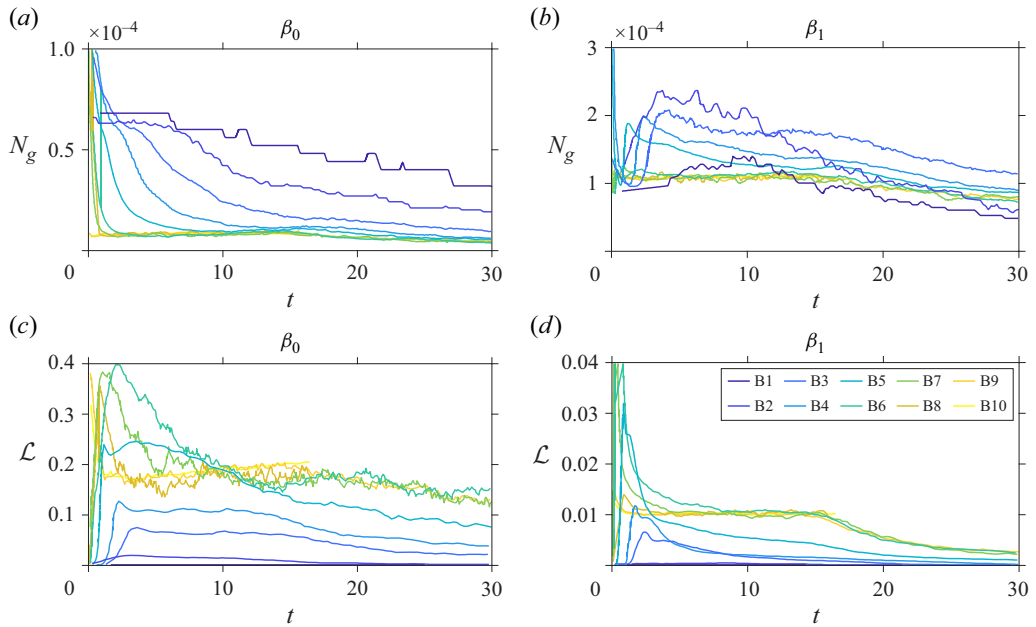


Figure 10. Number of generators N_g (*a,b*) (per area) and average lifespan $\mathcal{L}$ (*c,d*) (with noise band 0.01 removed), for simulations B1–B9. Note that simulation B10 is shorter than the others due to the limited amount of data available.

as shown in [figure 9](#), and it is much easier to quantify by considering PDs shown in [figure 10](#). For example, considering components, shown in [figures 10\(a\)](#) and [10\(c\)](#), one can observe that the number of generators (therefore, features in the temperature field) per area is larger for smaller values of Ra ; however, these features are much less prominent (on average) since their lifespan, $\mathcal{L}$, becomes progressively smaller as Ra decreases. Regarding the loops, we observe a clear correlation between their lifespans, shown in [figure 10\(d\)](#), and the onset of shutdown stage (see [figures 3a](#) and [5](#)).

Persistent diagrams also allow addressing how emerging length scales depend on the value of Ra . This question was discussed in previous works by [Fu *et al.* \(2013\)](#) and [De Paoli *et al.* \(2025d\)](#), and here we find these values using PDs; as we will see, additional information can be extracted since PDs encode detailed information about the topology of the temperature field. In particular, it is straightforward to extract the number of ‘features’, measured by Betti numbers, from PDs – essentially, for a given threshold value, one counts how many generators were born, but have not died yet – this is the number of features or objects of interest. Such objects could be components, measured by β_0 , or loops, measured by β_1 . As we now show, our findings illustrate the fact that the emerging length scales are strongly influenced both by the temperature threshold and by the type of features considered. Before proceeding, we note that the results that follow are essentially independent of the width of the band of generators next to the diagonal that we remove (as long as the width is small, corresponding to typical lifespans); those generators (close to the diagonal) do not influence the number of components or loops, except for very small thresholds as we discuss further below.

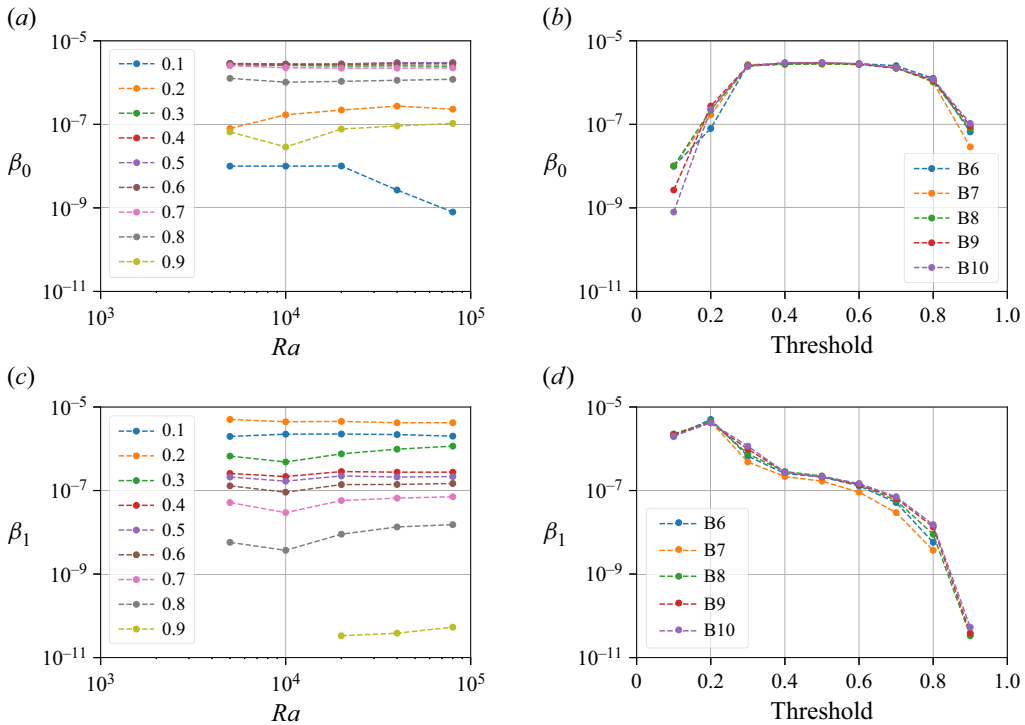


Figure 11. Betti numbers (per area) for simulations B6–B10: β_0 (a,b) and β_1 (c,d). (a,c) How Betti numbers depend on Ra for different thresholds. (b,d) How Betti numbers depend on the threshold value for all values of Ra considered.

Figure 11 shows β_0 and β_1 per area for different values of Ra (corresponding to simulations B6–B10), and for a set of thresholds; to help understanding of the results, we also plot how Betti numbers depend on the threshold for each value of Ra .

Focusing first on β_0 (figure 11a) we note that the results are non-monotonic when considering different threshold; the reason for the non-monotonicity is clear from figure 11(b) which shows that the values of β_0 are small either for very small or for very large thresholds. For small threshold values, β_0 is small since the areas of low temperature are all connected, so their number is small. For large threshold values, the number of areas with high temperatures is small, leading again to small values of β_0 . For intermediate thresholds, the values of β_0 reach their maximum. We note that the decrease in β_0 values for the two largest considered values of Ra at the threshold of 0.1 suggests increased connectivity of areas with temperature values of at least 0.1. We also note that figure 11(b) illustrates that the dependence of β_0 on the threshold value is robust across all considered values of Ra .

Considering β_1 (figure 11c,d), we note a different trend of the results as the considered temperature threshold is modified. Here (see figure 11d), there is an (almost) monotonic trend in β_1 , since the number of loops (per area) decreases as the threshold increases. Recall that the loops are born when the weakest link appears (as the threshold is decreased), and therefore for lower thresholds, there are more loops. For very large threshold values, there may not even be any loops, as can be seen in figure 11(d) for the two smallest values of Ra considered. The observed trend (decrease of β_1) persists for all but

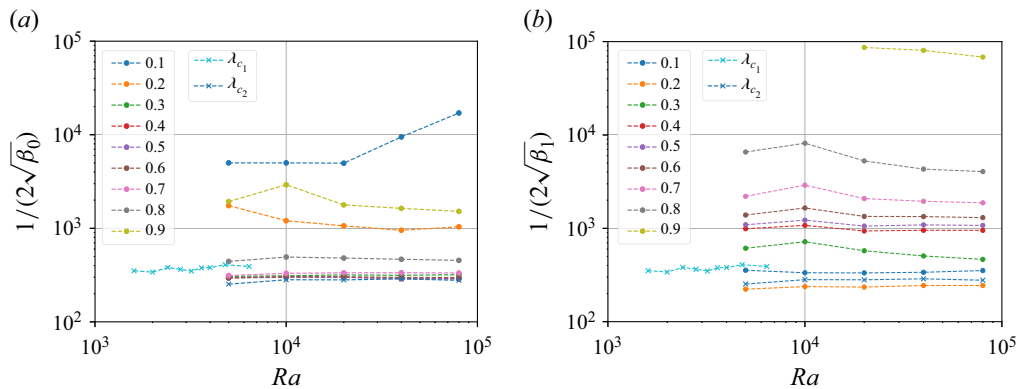


Figure 12. The length scales for simulations B6–B10 obtained using Betti numbers β_0 (a) and β_1 (b) as the considered temperature threshold is varied. We also show (by crosses) the values obtained by Fu *et al.* (2013) (λ_{c1}) and by De Paoli *et al.* (2025d) (λ_{c2}).

the smallest threshold; the reason for this change of trend is that for very small thresholds, the loops become filled up (recall the toy example shown in figure 2) and therefore their number decreases.

A brief comment is in order regarding the band of generators next to the diagonal, which is excluded from the consideration. As mentioned previously, if the bandwidth is small compared with a typical lifespan, these generators can be safely ignored. This requirement is satisfied for all but the smallest threshold (0.1) considered; additional results (not shown for brevity) show that indeed the Betti numbers are larger for very small thresholds if such a band is included; however, the trend of the results remains the same. We also note that minor deviations in the trends for β_0 and β_1 are expected due to the finite domain size considered. We also note in passing that similar trends in Betti number behaviour were observed when considering interaction networks in particulate-based systems (Kondic *et al.* 2012).

To obtain a typical length scale corresponding to Betti numbers shown in figure 11, we consider the following question: assuming that N features (components or loops) are randomly distributed in a rectangular domain, what is the typical minimal distance between such features? In two spatial dimensions, it is easy to see that for a domain of linear dimensions $\widehat{L}_1$ and $\widehat{L}_2$, such distance is given by $\sqrt{\widehat{L}_1 \widehat{L}_2 / N} / 2$ (this is an asymptotic expression obtained in the limit of large N and ignoring boundary effects). Figure 12 shows the corresponding results obtained when substituting β_0 or β_1 for N (per area) in the above expression. For β_0 , we find non-monotonic behaviour for different thresholds, as expected since β_0 are non-monotonic as well. For intermediate values of thresholds (0.3–0.7), the computed length scale is close to that found in De Paoli *et al.* (2025d) and slightly lower than that reported by Fu *et al.* (2013). When considering β_1 (figure 12b), we find agreement with Fu *et al.* (2013) and De Paoli *et al.* (2025d) for low thresholds (0.1–0.2), for which almost all loops present in the considered dataset are formed. In Appendix C we show that the results are essentially independent of the width of the excluded noise band (see figure 21).

To summarise, the topological analysis based on PH enhances our understanding of pattern formation in the system under consideration, illustrating, in particular, that results for emerging length scales depend on the approach used to compute them.

5. Two-sided convection

5.1. Convection regimes

We consider here a system heated from below and cooled from above, where a high (low) temperature is fixed at the bottom (top) wall, corresponding to the boundary conditions specified in (2.8). The domain is periodic in the horizontal directions, and the flow is initialised with a linear temperature distribution as in (2.9). After an initial transient phase, the flow attains a statistically steady state. The boundary conditions and an example of temperature distribution on the lateral boundaries of the domain are illustrated in figure 1(b).

In free fluids, i.e. in the absence of a porous medium, the two-sided flow is also called Rayleigh–Bénard flow, and it is particularly suitable to be analysed theoretically, due to its well-defined boundary conditions and its statistically steady nature: it represents a perfect candidate to develop new concepts on instabilities and dynamical systems (Lohse & Shishkina 2024). For the same reasons, the porous counterpart (also called Rayleigh–Bénard–Darcy or Rayleigh flow, where the free fluid is replaced by a fluid-saturated porous medium) has been widely investigated. In particular, extensive studies focused on the onset of the flow instabilities ($0 < Ra < O(10^2)$; see Horton & Rogers Jr 1945; Lapwood 1948) and on the dynamics at moderate to high Ra ($O(10^2) < Ra < O(10^4)$; see Graham & Steen 1994; Otero *et al.* 2004). In these systems, the global response parameter is the Nusselt number Nu defined as in (4.1) which, for $Ra > 10^3$, scales as $Nu \sim Ra$ plus sublinear corrections (Zhu *et al.* 2024). As a result, an increase of Ra by a factor 10 would require substantially increasing the number of degrees of freedom required to resolve the flow, namely by a factor of $O(10^3)$ in three dimensions, corresponding to an even larger increase of the computational cost. In recent decades, however, numerical advancements have allowed one to explore in detail the flow dynamics in three dimensions and at large Ra (Hewitt *et al.* 2014; Pirozzoli *et al.* 2021; De Paoli *et al.* 2022; Hu *et al.* 2023), and a rich flow morphology has been observed. A review for $Ra \leq 740$ is provided by Hewitt *et al.* (2014), which we summarise and extend in the following.

Figure 13 shows a few examples of the temperature distributions for three values of Ra , illustrating the changes of the flow morphology. To highlight the complex near-wall flow pattern, the temperature field is shown on a horizontal slice located near the top boundary and located at $z = z_t$ ($0.95 \leq z_t \leq 0.999$, depending on Ra), focusing on $Ra \geq 100$. For smaller values (specifically, for $Ra < 4\pi^2$) any instability is suppressed, and the flow is purely conductive. When Ra is slightly increased, a 2-D unstable mode appears ($4\pi^2 \leq Ra \lesssim 4.5\pi^2$), and dominates the flow also at larger Ra , when multiple modes are present ($4.5\pi^2 \lesssim Ra \lesssim 97$). In these regimes, temperature structures spanning the entire system height, from one boundary layer to the other, populate the domain. The flow eventually develops a 3-D steady structure ($97 \lesssim Ra \lesssim 300$; see figure 13a), and multiple possible states exist. Finally, the flow is unsteady for $Ra \gtrsim 300$: small flow instabilities develop and grow from the thermal boundary layers, move parallel to the horizontal walls and eventually merge into larger plumes, still covering the whole domain height, but now being unsteady (figure 13b). For sufficiently large driving, namely $Ra \geq 1750$, the system enters the so-called high- Ra regime and the flow morphology does not exhibit any regular background structure (Hewitt *et al.* 2014) (see e.g. figure 13c). In this regime, the flow can be separated into three different parts: (i) a thin and time-dependent thermal boundary layer at the walls; (ii) an intermediate region populated by

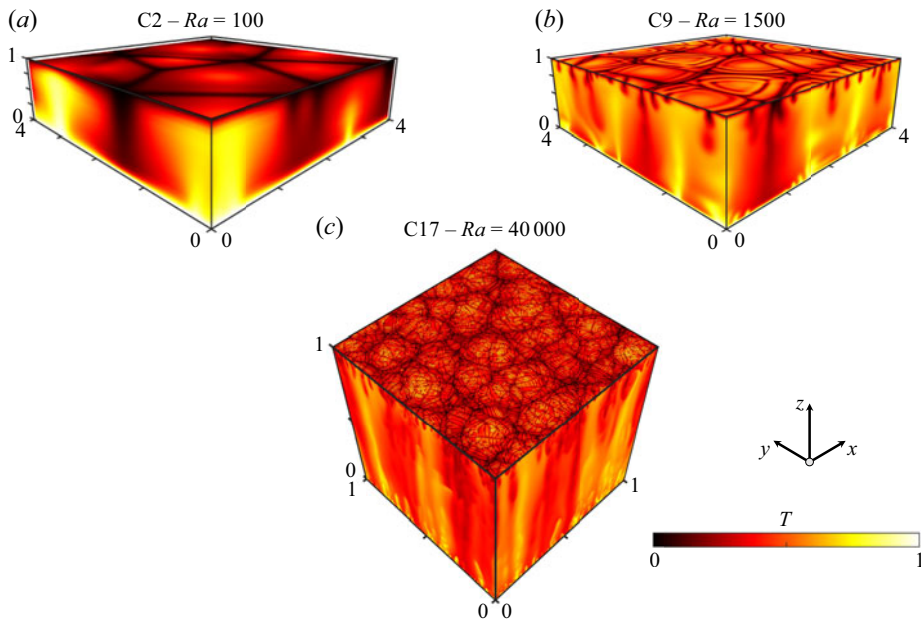


Figure 13. Temperature distribution over a portion of the domains considered, namely $0 \leq x \leq 4$, $0 \leq y \leq 4$ and $0 \leq z \leq z_t$ (with $0.95 \leq z_t \leq 0.999$, depending on Ra). Distributions shown are relative to simulations C2 (a), C9 (b) and C17 (c) (see table 2 for additional details).

very dynamical sheet-like plumes, originated from the time-dependent wall layer; and (iii) the bulk, controlled by large columnar structures of hot (cold) rising (sinking) fluid, labelled megaplumes.

The dynamics at larger Ra ($Ra \lesssim 10^5$) has been recently explored in detail by Pirozzoli *et al.* (2021) and De Paoli *et al.* (2022). Those authors observed that when $Ra \gtrsim 10^4$, the temperature distribution in the intermediate region is characterised not only by the presence of the small plumes originated from the boundary layer, but also by the existence of large structures, entraining many plumes and representing the footprint of the megaplumes populating the bulk of the flow (De Paoli *et al.* 2022). These structures, labelled as supercells, have been identified visually as the large loops gathering many smaller cells (see e.g. the dark loops on the horizontal cut in figure 13c), and have been observed to be: (i) persistent in time and (ii) correlated in space with the megaplumes (De Paoli *et al.* 2022). Time persistence has been identified by averaging the temperature distribution in the near-wall region and observing that a pattern matching that of the supercells emerges. Spatial correlation with the megaplumes was demonstrated by filtering the near-wall temperature field to remove flow structures with wavelengths smaller than the dominant wavelength at the mid-plane.

Despite these efforts, a precise definition of the supercells and an accurate determination of the conditions required for their appearance remain elusive. The following questions remain unanswered: Can we provide a more detailed description of the supercells? Can we describe their presence/formation as a function of Ra ? Do we have to rely on the bulk temperature field (i.e. the temperature field far from the top and bottom boundaries) to determine the existence and the morphology of the supercells? In this work, we aim at

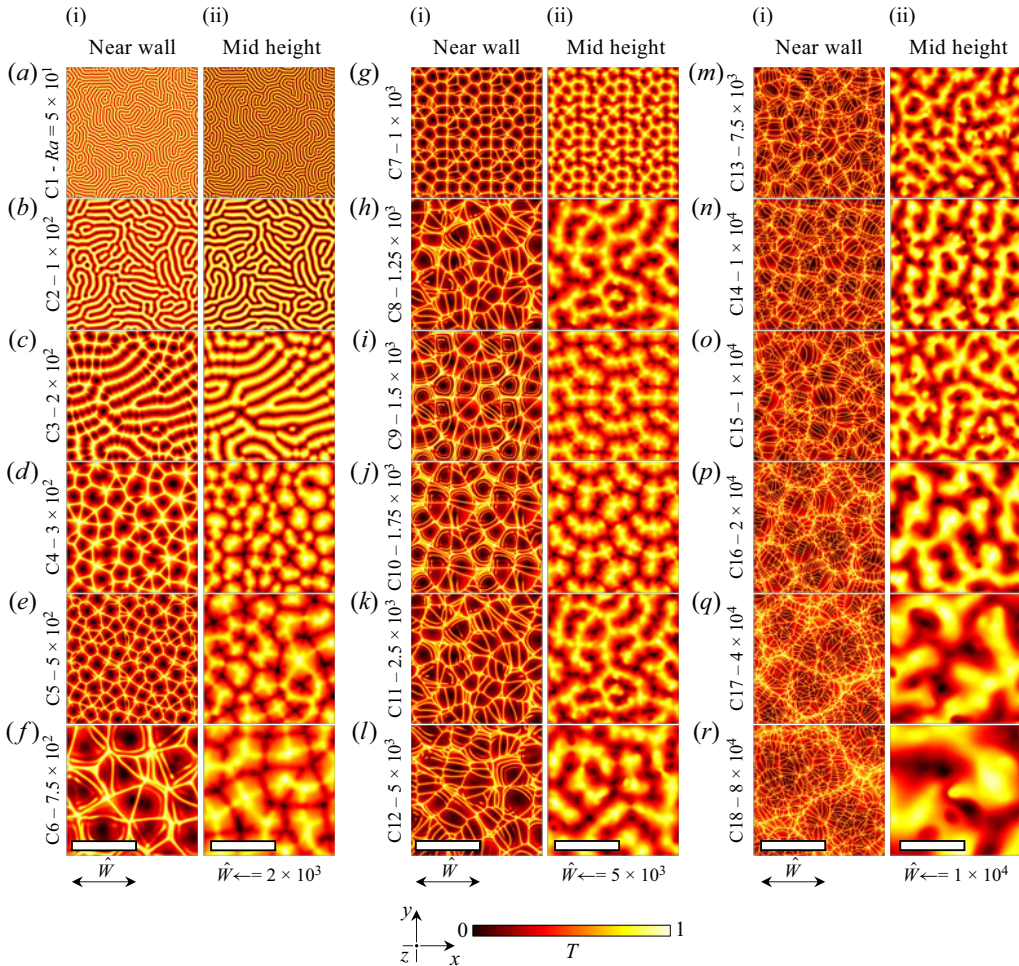


Figure 14. Visualisation of the temperature fields on horizontal (x, y) planes near the bottom wall (i) and at the mid-height ($z = 1/2$) (ii). The ‘C’ simulations and the Rayleigh numbers Ra are indicated; for all other details we refer to [table 2](#) (see [Appendix A](#)). The size of the domain shown is constant within each column, but varies from left to right for visualisation purposes. A scale bar (width $\hat{W}$ expressed in diffusive units) is reported for reference. In some cases (e.g. C18), only a small portion of the domain is shown, while in others (e.g. C7), the field is replicated to match the desired size. Note the difference between the cell structure in the present results and in those of Hewitt *et al.* (2014) (see the results labelled as IC1 in Hewitt *et al.* (2014), corresponding to the same initial condition employed here), suggesting the strong effect of the domain size on the determination of the flow structure at low Rayleigh numbers ($Ra \leq 500$). The data analysed relative to the ‘C’ simulations are available via De Paoli *et al.* (2025a).

precisely answering these questions: we provide a robust criterion to identify supercells and describe their formation across a wide range of Ra .

5.2. Qualitative analysis of the flow morphology

We analyse the flow morphology for $5 \times 10^1 \leq Ra \leq 8 \times 10^4$, by considering the horizontal temperature distributions in the near-wall region (panels (i) in [figure 14](#)) and at the midplane (panels (ii) in [figure 14](#)). For ease of comparison, the size of the portion of domains shown in [figure 14](#) is constant (in diffusive units) within each column (scale

bar is reported as a reference). The simulations considered are either those by Pirozzoli *et al.* (2021) and De Paoli *et al.* (2022), or presented here for the first time, and the data are made available via De Paoli *et al.* (2025a).

At $Ra \leq 1750$, the flow is also dependent on the initial condition, and it exhibits hysteresis effects (Otero *et al.* 2004). Hewitt *et al.* (2014) explored the dynamics following two initial conditions: IC1, corresponding to a perturbed linear temperature profile, and IC2, obtained starting from a steady-state temperature field obtained from a simulation at a lower Ra . Therefore, the initial condition adds here to the flow governing parameters ($Ra, \widehat{L}_1, \widehat{L}_2$). To reduce the parameter space, here we consider only one initial condition (corresponding to IC1 of Hewitt *et al.* (2014)) characterised by a perturbed form of the stable state, (2.9), and we consider very large domains ($\widehat{L}_1 = \widehat{L}_2 \geq 4 \times 10^3$), such that the flow is independent of the values of $\widehat{L}_1$ and $\widehat{L}_2$. As a result, the only remaining governing parameter is Ra . Simulations performed in this work have been initialised with a linear temperature distribution, (2.9), which in dimensionless terms reads $T = 1 - z$, and have been run for $t \geq 1000$, to make sure the steady state (determined by keeping track of the time-averaged value of the Nusselt number) is achieved. Then, the temperature fields used to analyse the flow pattern are saved approximately every 10 or 20 convective time units. All the relevant simulation details are indicated in table 2 (see Appendix A).

Before quantifying the flow pattern using PH techniques, we notice that a visual inspection reveals some interesting features previously unobserved. In particular, at low Rayleigh numbers the flow organisation described in the literature (see figure 2b of Hewitt *et al.* (2014)) consists of cells forming a very regular pattern that for $Ra \leq 2 \times 10^2$ is one-dimensional (rolls spanning across the entire domain width) and for $2 \times 10^2 \leq Ra \leq 3 \times 10^2$ is 2-D and still very regular (equally squared cells). The domains considered in figure 14 are about 100 times larger than those in Hewitt *et al.* (2014), and the cells are organised in a more chaotic manner. For $Ra \leq 2 \times 10^2$, corresponding to figure 14(a–c), we still observe the formation of rolls, which in contrast to Hewitt *et al.* (2014) do not span across the entire domain in the horizontal directions. Also for $Ra = 3 \times 10^2$ (figure 14d) there are differences compared with previous works, where the cells were all equal: we do observe the transition from rolls to cells, but such cells have an irregular shape and are, in general, different from each other (polygons with 3–6 sides). These differences suggest that the previously observed regular flow pattern emerging at small Ra is a genuine domain size effect, which is induced by the periodic boundary conditions applied in the horizontal directions (additional details are provided in Appendix B.1). These findings highlight the fact that at small Ra the influence of the domain size on the flow morphology is considerable. Finally, we observe that at very large Rayleigh numbers ($Ra \geq 2 \times 10^4$; figure 14p–r), where the supercells are clearly visible, their size increases with Ra . This is not surprising, since the temperature structures in the bulk (megaplumes of size $\sim Ra^{1/2}$ in diffusive units; see Hewitt *et al.* 2014; Pirozzoli *et al.* 2021) are responsible for the formation of the supercells (De Paoli *et al.* 2022). However, this highlights once again the importance of domain size in capturing the large-scale structures in the flow.

5.3. Quantification of pattern formation

We focus the presentation that follows on the identification of cells and supercells and, more generally, on the influence of Ra on the pattern-formation process. We discuss loop-related measures only, as they provide a clear description of the patterns of interest.

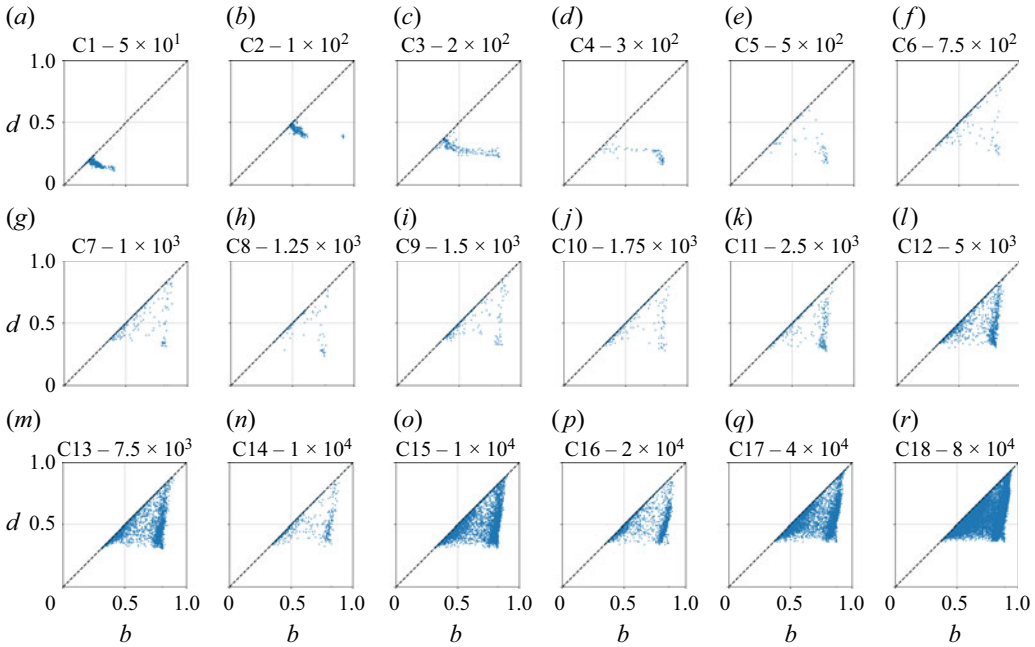


Figure 15. Loop PDs, β_1 PDs, in two-sided convection simulation near the wall (simulations C in table 2). As the Rayleigh number (Ra) increases, the morphology of the boundary develops significant changes, leading to variations in the corresponding β_1 PD. Between $Ra = 300$ and 500 , we observe a morphological pattern change, with a clear transformation from a maze to a cellular network (see also figure 14). Proceeding from $Ra = 1000$, a division emerges, forming two distinct clusters of points in the PDs. This indicates an increasing number of loops within the boundary layer, appearing at high temperatures ($T \approx 0.75$). For $Ra \geq 2500$, the cluster at high temperatures becomes denser, suggesting an increased number of loops in the patterns shown in figure 14.

Figure 15 shows the loop (β_1) PDs for the simulations listed in table 2. The emerging behaviour, as Ra increases, is an increase in the number of generators, and, in particular, the gradual appearance of a band of generators with the birth coordinates of $T \gtrsim 0.75$ starting at $Ra \gtrsim 1000$. For larger Ra , the number of points in the PDs increases substantially, and their structure becomes difficult to visualise. For this reason, we consider cumulative measures, which simplify the interpretation of the results, as discussed next.

Figure 16 shows the normalised distribution (probability density function, p.d.f.) of the birth coordinate of the loops, using the data from PDs shown in figure 15, but averaged over 10 time outputs (saved approximately every 10 or 20 convective time units) in order to improve statistics. We recall that the birth coordinate represents the temperature at which the loops close and, therefore, at which the cells (or supercells) form. At low values of Ra , we note a dramatic change in the p.d.f. as Ra increases from 200 (figure 16c) to 300 (figure 16d), illustrating the topological change between rolls (figure 16c) and cells (figure 16d). This topological change can be visualised in figure 14: the rolls initially present in the system (figure 14a–i) start to connect, forming cells (figures 14b–i and 14c–i). These cells fully develop and are clearly visible at larger Ra (starting with figure 14d–i). Further increase of Ra leads to development of a clear and well-defined structure of the p.d.f.s shown in figure 16 with a dominant peak at $T \approx 0.8$

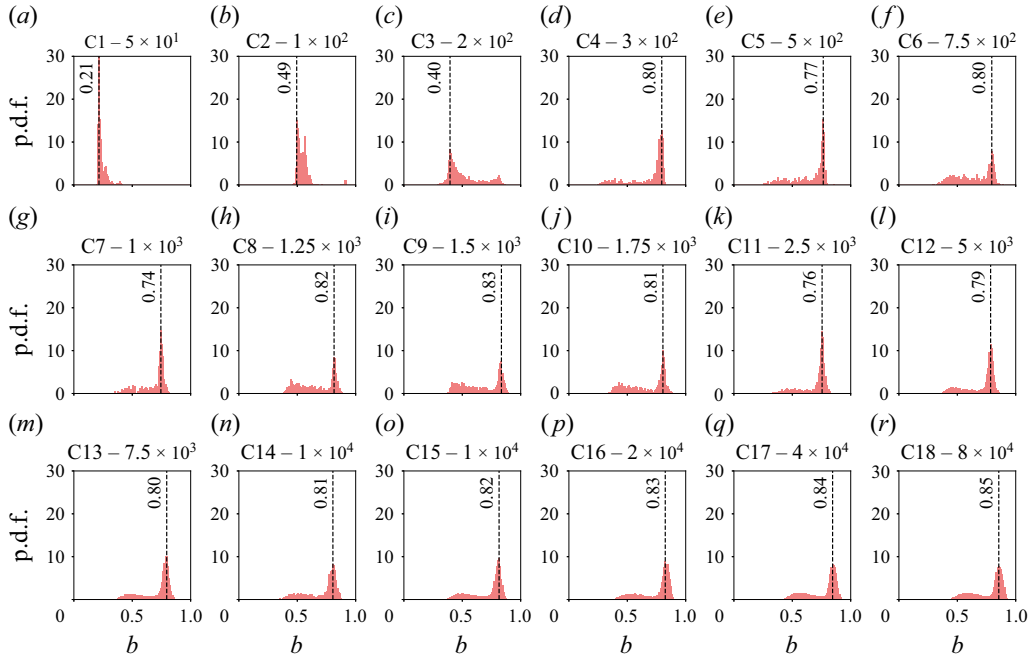


Figure 16. The p.d.f. of the birth values of β_1 topological generators, accumulated over 10 time outputs for each simulation (noise band 0.01 is removed). The birth coordinate corresponding to the maximum value of the p.d.f. is indicated by the dashed line and is reported in each panel. A transition from rolls (*a*–*c*) to cells/supercells (*d*–*r*) is clearly observed.

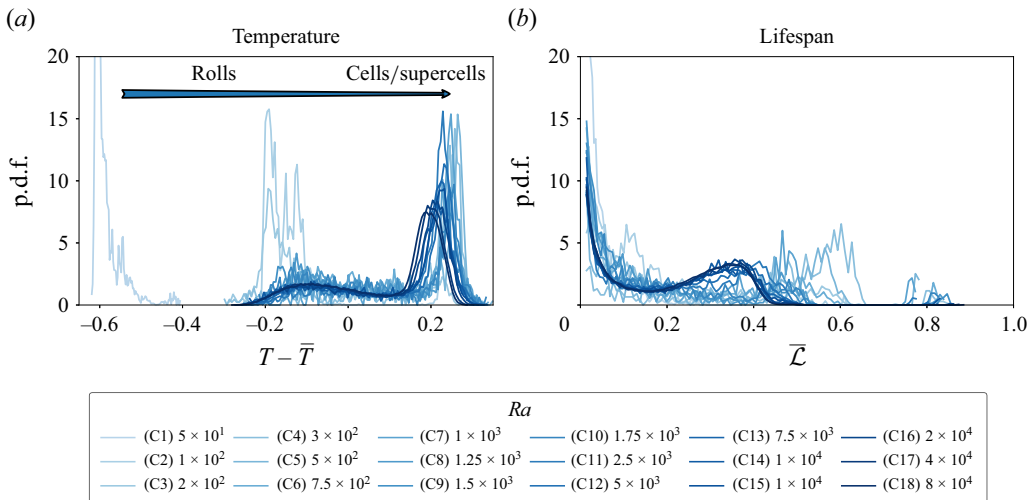


Figure 17. (*a*) The p.d.f. of β_1 topological generators as a function of the shifted temperature, with $\bar{T}$ corresponding to the space- and time-averaged temperature in the considered horizontal slice; (*b*) p.d.f. of the lifespan, $\bar{\mathcal{L}}$ (noise band 0.01 is removed). Note that the lifespan is independent of the temperature shift. We observe approximate self-similarity of the results for large values of Ra .

corresponding to supercells (see e.g. figures 14*o–i* to 14*r–i*), and broad distribution of the birth values for $0.5 \lesssim T \lesssim 0.7$, corresponding to cells. We note that the portions of the flow in which $T \approx 0.75$ roughly correspond to the regions in which the horizontal temperature gradient is maximum (De Paoli *et al.* 2022). We also comment that the results for p.d.f.s of connected components do not provide any useful information; therefore, considering loops is crucial for describing cells and supercells.

Further insight can be obtained by considering the loops' birth coordinate as a function of $T - \bar{T}$, where $\bar{T}$ is the temperature averaged in space (over the horizontal plane considered) and in time (over 10 time outputs as also implemented for the plots shown in figure 16). Figure 17 plots combined results, showing the β_1 birth coordinate distribution and the β_1 lifespan. Most importantly, the approximate collapse of the curves for $Ra \gtrsim 10^4$ for both birth values in figure 17(*a*) and lifespans in figure 17(*b*) illustrates the validity of the topological properties describing the temperature field across the wide range of parameters considered. More precisely, the collapse in figure 17(*a*) shows that the supercells appear at similar values of the shifted temperature $T - \bar{T}$, while the collapse visible in figure 17(*b*) shows that the cells also die (merge with other cells) at similar values of the shifted temperature (since lifespan measures difference between the birth and death coordinates). For smaller values of Ra , while the lifespans shown in figure 17 are too noisy to reach any useful insight, the birth coordinate distribution shows a clear shift in the peak of p.d.f.s from $T - \bar{T} \sim -0.6$ for $Ra = 50$ to $T - \bar{T} \sim -0.2$ for $100 \leq Ra \leq 200$, and then to $T - \bar{T} \sim 0.2$ for larger values of Ra . Additional details are provided in Appendix B.2, where we prove the robustness of our findings, with respect to the number of samples and the different aspect ratios considered.

6. Summary and conclusions

We examine the morphology of flow patterns arising in convection in porous media by combining large-scale numerical simulations with tools from topological data analysis, in particular, PH. The PH results we present are based on planar near-wall temperature fields and describe the topology of those slices, not the full 3-D topology of the flow. By analysing both one-sided and two-sided flow configurations over a broad range of Rayleigh–Darcy numbers and domain sizes, we demonstrate that PH provides an objective, quantitative framework for characterising flow structures. In particular, given the temperature distributions, we observe the emergence of complex temperature structures, which are described in the present work. Unlike classical approaches previously employed, based on threshold selection (De Paoli *et al.* 2022), Fourier analysis (Pirozzoli *et al.* 2021) or cell-size measurements (Fu *et al.* 2013), PH quantifies structures across all temperature levels simultaneously and captures their number, connectivity and persistence. It is important to note that PH analysis shows that the emerging length scales depend on the threshold selection, with only a specific threshold range reproducing the existing results. This analysis allows us to relate flow morphology directly to the system's transport properties, such as the Nusselt number Nu , and to identify features that are common across different regimes.

For the one-sided configuration, PH measures reveal clear signatures associated with the canonical stages of convective dissolution. During the diffusive stage, the number and prominence of topological features remain small, reflecting the smooth gradients observed across the temperature field. As convection initiates and thermal plumes form, both the

number of topological generators and their lifespans increase, capturing the onset of multi-scale structures. Subsequent plume merging and coarsening are reflected by a reduction in the number of generators but an increase in their lifespans, consistent with the emergence of larger coherent plumes. These trends correlate strongly with the temporal evolution of the Nusselt number, highlighting the tight coupling between topology and heat flux. In particular, the rapid decay of PH lifespans correlates with the onset of shutdown and may provide a useful post-processing signature of this transition.

For the two-sided configuration, where the flow morphology is richer than in the one-sided case, PH successfully identifies the hierarchical organisation of structures near the walls. We show that the transition from near-wall maze-like structures to cellular supercells corresponds to systematic variations in the distributions of birth and death coordinates of loop generators. At sufficiently high Rayleigh numbers, the p.d.f.s of both birth values and lifespans collapse onto each other, demonstrating a self-similar behaviour in the high- Ra regime. The topological measures, therefore, show generic, self-similar characteristics of the emerging patterns in this regime. This finding suggests that, despite the apparent visual complexity of the patterns, their near-wall topological organisation becomes Ra -independent once convection is sufficiently vigorous, the domain is sufficiently large and relative to the initial condition considered. In addition, we also find that large-aspect-ratio simulations at low Ra (namely at $Ra \leq 10^2$) lead to a maze-like flow organisation that differs from the more regular cell pattern previously observed in smaller domains (Hewitt *et al.* 2014).

Overall, our results show that PH is a powerful tool for analysing convective patterns in porous media, enabling us to uncover structural transitions, quantify multi-scale behaviour and establish links between morphology and macroscopic transport. Unlike Fourier-based or cell-size analyses, PH captures the connectivity and merging of plume structures, allowing identification of topological transitions (e.g. plume coalescence into supercells) that are not directly accessible from spectral measures. The measures introduced here are fully objective, computationally efficient and well suited for large datasets. They represent a possible direction for developing predictive reduced-order models or data-driven modelling of heat and solute transport, and can be extended to pore-scale or heterogeneous media (Blunt 2017). The availability of a large open database of simulations further enables the community to build upon this work. Future research directions include extending PH analyses to 3-D structures and integrating topological features into machine-learning frameworks for real-time prediction of heat-transport properties in convective flows.

The topological measures, therefore, show generic, self-similar characteristics of the emerging patterns for large values of Ra .

Supplementary material and movies. Supplementary material and movies are available at <https://doi.org/10.1017/jfm.2026.11804>.

Acknowledgements. We acknowledge the contribution by Z. Cao, who (while at NJIT) carried out some of the topological data analyses performed during the initial stage of this work.

Funding. Funded by the European Union (ERC, MORPHOS, 101163625). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. We acknowledge the EuroHPC Joint Undertaking for awarding the projects EHPC-EXT-2024E02-122 and EHPC-BEN-2024B08-060 to access the EuroHPC supercomputer MareNostrum5 hosted at the Barcelona Supercomputing Center (Spain). L.K. acknowledge partial funding by NSF grants DMR-2410985 and DMS-2201627. M.D. and L.K. acknowledge the Global Fellowship Program of TU Wien. The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

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

Case	Ra	Convective units			Diffusive units			Grid resolution $N_x \times N_y \times N_z$	References
		L_1	L_2	H	$\hat{L}_1$	$\hat{L}_2$	$\hat{H}$		
A1	1.0×10^4	1	1	1	1.0×10^4	1.0×10^4	1.0×10^4	$768 \times 768 \times 256$	This work
A2	1.0×10^4	1	1/2	1	1.0×10^4	5.0×10^3	1.0×10^4	$768 \times 384 \times 256$	This work
A3	1.0×10^4	1	1/4	1	1.0×10^4	2.5×10^3	1.0×10^4	$768 \times 192 \times 256$	This work
A4	1.0×10^4	1	1/8	1	1.0×10^4	1.25×10^3	1.0×10^4	$768 \times 96 \times 256$	This work
A5	1.0×10^4	1	1/16	1	1.0×10^4	6.25×10^2	1.0×10^4	$768 \times 48 \times 256$	This work
A6	1.0×10^4	2	2	1	2.0×10^4	2.0×10^4	1.0×10^4	$1536 \times 1536 \times 256$	This work
A7	1.0×10^4	1/2	1/2	1	5.0×10^3	5.0×10^3	1.0×10^4	$384 \times 384 \times 256$	This work
A8	1.0×10^4	1/4	1/4	1	2.5×10^3	2.5×10^3	1.0×10^4	$192 \times 192 \times 256$	This work
A9	1.0×10^4	1/8	1/8	1	1.25×10^3	1.25×10^3	1.0×10^4	$96 \times 96 \times 256$	This work
A10	1.0×10^4	1/16	1/16	1	6.25×10^3	6.25×10^3	1.0×10^4	$48 \times 48 \times 256$	This work
B1	1.0×10^2	5	5	1	5.0×10^2	5.0×10^2	1.0×10^2	$64 \times 64 \times 16$	De Paoli <i>et al.</i> (2025d)
B2	2.0×10^2	5	5	1	1.0×10^3	1.0×10^3	2.0×10^2	$128 \times 128 \times 24$	De Paoli <i>et al.</i> (2025d)
B3	5.0×10^2	5	5	1	2.5×10^3	2.5×10^3	5.0×10^2	$256 \times 256 \times 32$	De Paoli <i>et al.</i> (2025d)
B4	1.0×10^3	5	5	1	5.0×10^3	5.0×10^3	1.0×10^3	$512 \times 512 \times 48$	De Paoli <i>et al.</i> (2025d)
B5	2.0×10^3	5	5	1	1.0×10^4	1.0×10^4	2.0×10^3	$1024 \times 1024 \times 64$	De Paoli <i>et al.</i> (2025d)
B6	5.0×10^3	2	2	1	1.0×10^4	1.0×10^4	5.0×10^3	$1024 \times 1024 \times 128$	De Paoli <i>et al.</i> (2025d)
B7	1.0×10^4	1	1	1	1.0×10^4	1.0×10^4	1.0×10^4	$1024 \times 1024 \times 256$	De Paoli <i>et al.</i> (2025d)
B8	2.0×10^4	1/2	1/2	1	1.0×10^4	1.0×10^4	2.0×10^4	$1024 \times 1024 \times 512$	De Paoli <i>et al.</i> (2025d)
B9	4.0×10^4	1/2	1/2	1	2.0×10^4	2.0×10^4	4.0×10^4	$2048 \times 2048 \times 1024$	De Paoli <i>et al.</i> (2025d)
B10	8.0×10^4	1/2	1/2	1	4.0×10^4	4.0×10^4	8.0×10^4	$4096 \times 4096 \times 2048$	De Paoli <i>et al.</i> (2025d)

Table 1. Summary of the simulations considered in the one-sided case. For each simulation, the Rayleigh number Ra , the domain extension in convective (L_1 , L_2 , H) and diffusive ($\hat{L}_1$, $\hat{L}_2$, $\hat{H}$) units and the grid resolution $N_x \times N_y \times N_z$ are indicated. Simulations A are carried out in domains of variable sizes, assuming a constant Rayleigh number ($Ra = 1 \times 10^4$). Simulations B (presented by De Paoli *et al.* (2025d)) are carried out in square domains ($\hat{L}_1/\hat{L}_2 = L_1/L_2 = 1$) while Ra varies. The data analysed relative to simulations A are available via De Paoli *et al.* (2025a).

Appendix A. Additional details on numerical simulations

A summary of the simulations considered is presented in tables 1 and 2, corresponding to the one-sided and two-sided cases, respectively. For each simulation, the Rayleigh number Ra , the domain extension in convective (L_1 , L_2 , H) and diffusive ($\hat{L}_1$, $\hat{L}_2$, $\hat{H}$) units and the grid resolution $N_x \times N_y \times N_z$ are indicated. The data analysed relative to simulations A and C are available via De Paoli *et al.* (2025a).

Appendix B. Effect of the domain size

B.1. Low Rayleigh numbers

To illustrate the combined effect that, at low Ra , domain size and periodic boundary conditions in the horizontal directions have on the flow morphology, we consider the case at $Ra = 10^2$, corresponding to simulation C2 in table 2. In addition to C2 ($L = 40$), we perform additional simulations at smaller values of $L_1 = L_2 = L$. To ensure that steady state is achieved, the simulations are run for $t > 5000$. Figure 18 then shows the temperature field in the centre of the domain at the final time instant. A gradual increase of the domain size L allows us to observe the change from structures spanning the

Case	Ra	Convective units			Diffusive units			Grid resolution $N_x \times N_y \times N_z$	References
		L_1	L_2	H	$\widehat{L}_1$	$\widehat{L}_2$	$\widehat{H}$		
C1	5.00×10^1	80	80	1	4.0×10^3	4.0×10^3	5.0×10^1	$384 \times 384 \times 16$	This work
C2	1.00×10^2	40	40	1	4.0×10^3	4.0×10^3	1.0×10^2	$384 \times 384 \times 16$	This work
C3	2.00×10^2	20	20	1	4.0×10^3	4.0×10^3	2.0×10^2	$384 \times 384 \times 16$	This work
C4	3.00×10^2	40/3	40/3	1	4.0×10^3	4.0×10^3	3.0×10^2	$384 \times 384 \times 16$	This work
C5	5.00×10^2	8	8	1	4.0×10^3	4.0×10^3	5.0×10^2	$384 \times 384 \times 32$	This work
C6	7.50×10^2	16/3	16/3	1	4.0×10^3	4.0×10^3	7.5×10^2	$384 \times 384 \times 32$	This work
C7	1.00×10^3	4	4	1	4.0×10^3	4.0×10^3	1.0×10^3	$384 \times 384 \times 32$	De Paoli <i>et al.</i> (2022)
C8	1.25×10^3	4	4	1	5.0×10^3	5.0×10^3	1.25×10^3	$512 \times 512 \times 64$	This work
C9	1.50×10^3	10/3	10/3	1	5.0×10^3	5.0×10^3	1.50×10^3	$512 \times 512 \times 64$	This work
C10	1.75×10^3	20/7	20/7	1	5.0×10^3	5.0×10^3	1.75×10^3	$512 \times 512 \times 64$	This work
C11	2.50×10^3	4	4	1	1.0×10^4	1.0×10^4	2.5×10^3	$768 \times 768 \times 64$	Pirozzoli <i>et al.</i> (2021)
C12	5.00×10^3	4	4	1	2.0×10^4	2.0×10^4	5.0×10^3	$1536 \times 1536 \times 128$	Pirozzoli <i>et al.</i> (2021)
C13	7.50×10^3	4	4	1	3.0×10^4	3.0×10^4	7.5×10^3	$2304 \times 2304 \times 192$	De Paoli <i>et al.</i> (2022)
C14	1.00×10^4	1	1	1	1.0×10^4	1.0×10^4	1.0×10^4	$768 \times 768 \times 256$	Pirozzoli <i>et al.</i> (2021)
C15	1.00×10^4	4	4	1	4.0×10^4	4.0×10^4	1.0×10^4	$3072 \times 3072 \times 256$	De Paoli <i>et al.</i> (2022)
C16	2.00×10^4	1	1	1	2.0×10^4	2.0×10^4	2.0×10^4	$1536 \times 1536 \times 512$	Pirozzoli <i>et al.</i> (2021)
C17	4.00×10^4	1	1	1	4.0×10^4	4.0×10^4	4.0×10^4	$3072 \times 3072 \times 1024$	Pirozzoli <i>et al.</i> (2021)
C18	8.00×10^4	1	1	1	8.0×10^4	8.0×10^4	8.0×10^4	$6144 \times 6144 \times 2048$	Pirozzoli <i>et al.</i> (2021)

Table 2. Summary of two-sided flow configuration at different Rayleigh numbers Ra (data are taken from Pirozzoli *et al.* (2021) and De Paoli *et al.* (2022)). The domain dimensions (H , L_1 and L_2) are indicated. Quantities are expressed in dimensionless convective units ($H = 1$, $L_1 = L_1^*/H^*$, $L_2 = L_2^*/H^*$, defined as in § 2.2.1) and dimensionless diffusive units ($\widehat{H} = Ra$, $\widehat{L}_1 = L_1^*/\ell^*$, $\widehat{L}_2 = L_2^*/\ell^*$, defined as in § 2.2.2). The data analysed relative to simulations C are available via De Paoli *et al.* (2025a).

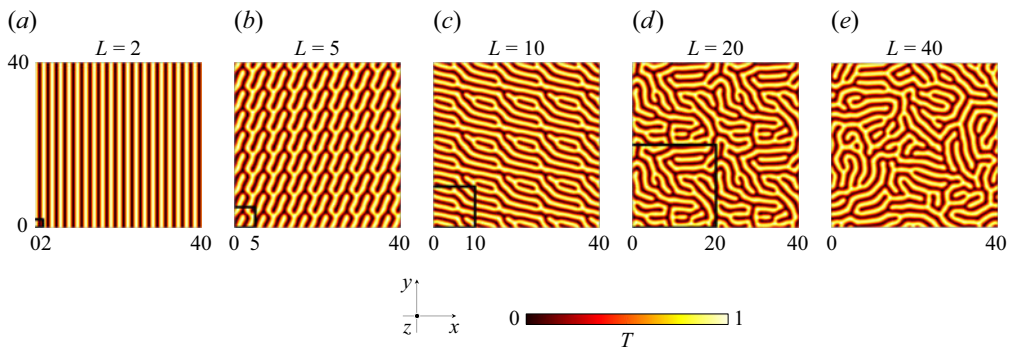


Figure 18. Simulations performed at $Ra = 10^2$ and (a–e) different values of domain width ($L_1 = L_2 = L$). The black, square frame indicates the domain simulated, which is repeated for visualisation purposes to match the size of the largest simulation considered, C2, corresponding to $L = 40$ (e).

entire domain in the horizontal direction (figure 18a) to a more chaotic flow organisation (figure 18d,e). Such transition, however, only mildly affects the Nusselt number, which varies between 2.54 ($L = 5$; figure 18b) and 2.61 ($L = 2$; figure 18a).

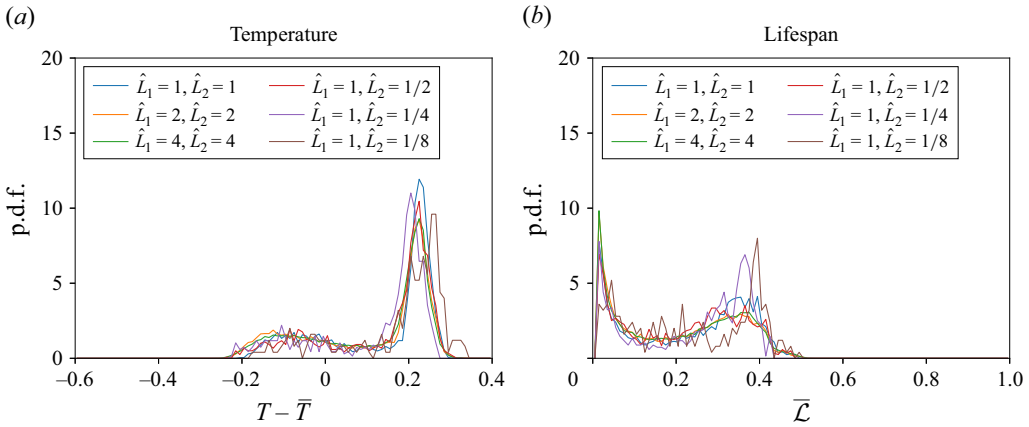


Figure 19. Influence of the domain size aspect ratio for $Ra = 10^4$ (compare figure 17). (a) The p.d.f. of β_1 topological generators as a function of the shifted temperature, with $\bar{T}$ corresponding to the space- and time-averaged temperature in the considered horizontal slice; (b) p.d.f. of the lifespan, $\bar{\mathcal{L}}$ (noise band 0.01 is removed).

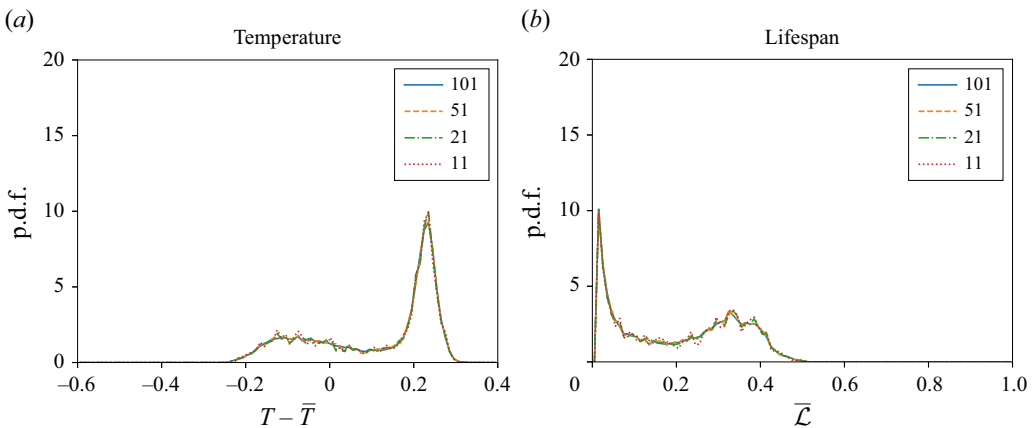


Figure 20. Influence of the number of considered samples for $Ra = 10^4$ (compare figure 17). (a) The p.d.f. of β_1 topological generators as a function of the shifted temperature, with $\bar{T}$ corresponding to the space- and time-averaged temperature in the considered horizontal slice; (b) p.d.f. of the lifespan, $\bar{\mathcal{L}}$ (noise band 0.01 is removed).

B.2. High Rayleigh numbers

To illustrate the robustness of the results, here we include additional results for $Ra = 10^4$ as the aspect ratio is modified (see figure 19). We have also confirmed that the results are robust with respect to the number of samples considered (see figure 20). These results confirm the robustness of our findings regarding the collapse of the PH measures for large values of Ra .

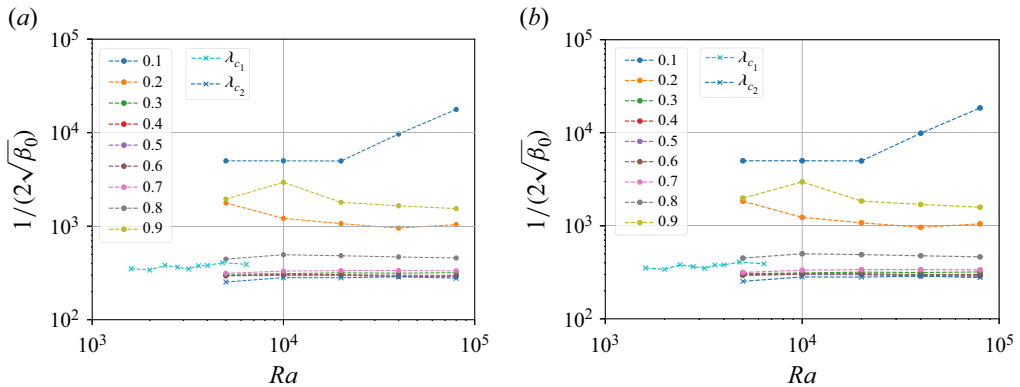


Figure 21. Betti numbers (per area) for simulations B6–B10 for different values of the band width; compare with figure 12(a). (a) Removing the 0.005 band and (b) removing the 0.02 band.

Appendix C. Influence of the removed band width

Here we briefly comment regarding the influence of the width of the excluded band used when computing the quantities of interest, such as emerging length scales. As an illustrative example, figure 21 reproduces figure 12(a) from § 4.3 obtained by changing this width; direct comparison shows that the results are essentially independent of the band width.

REFERENCES

- BACKHAUS, S., TURITSYN, K. & ECKE, R.E. 2011 Convective instability and mass transport of diffusion layers in a Hele–Shaw geometry. *Phys. Rev. Lett.* **106** (10), 104501.
- BLUNT, M.J. 2017 *Multiphase Flow in Permeable Media: A Pore-Scale Perspective*. Cambridge University Press.
- CARLSSON, G. 2009 Topology and data. *Bull. Amer. Math. Soc. (N.S.)* **46** (2), 255–308.
- DE PAOLI, M. 2023 Convective mixing in porous media: a review of Darcy, pore-scale and Hele–Shaw studies. *Eur. Phys. J. E* **46** (12), 129.
- DE PAOLI, M. 2025 AFiD-Darcy. Available at: <https://github.com/depaolimarco/AFiD-Darcy>.
- DE PAOLI, M., PIROZZOLI, S. & KONDIC, L. 2025a Data supporting: flow morphology and patterns in porous media convection: a persistent homology analysis. Available at: <https://doi.org/10.5281/zenodo.17529345>.
- DE PAOLI, M., PIROZZOLI, S., ZONTA, F. & SOLDATI, A. 2022 Strong Rayleigh–Darcy convection regime in three-dimensional porous media. *J. Fluid Mech.* **943**, A51.
- DE PAOLI, M., YERRAGOLAM, G.S., LOHSE, D. & VERZICCO, R. 2025c AFiD-Darcy: a finite difference solver for numerical simulations of convective porous media flows. *Comput. Phys. Commun.* **312**, 109579.
- DE PAOLI, M., YERRAGOLAM, G.S., VERZICCO, R. & LOHSE, D. 2025b Solute mixing in porous media with dispersion and buoyancy. *J. Fluid Mech.* **1020**, A24.
- DE PAOLI, M., ZONTA, F., ENZENBERGER, L., COLIBAN, E. & PIROZZOLI, S. 2025d Simulation and modeling of convective mixing of carbon dioxide in geological formations. *Geophys. Res. Lett.* **52** (7), e2025GL114804.
- DE PAOLI, M., ZONTA, F. & SOLDATI, A. 2016 Influence of anisotropic permeability on convection in porous media: implications for geological CO₂ sequestration. *Phys. Fluids (1994–Present)* **28** (5), 056601.
- DE PAOLI, M., ZONTA, F. & SOLDATI, A. 2017 Dissolution in anisotropic porous media: modelling convection regimes from onset to shutdown. *Phys. Fluids* **29** (2), 026601.
- DHAR, J., MEUNIER, P., NADAL, F. & MEHEUST, Y. 2022 Convective dissolution of carbon dioxide in two- and three-dimensional porous media: the impact of hydrodynamic dispersion. *Phys. Fluids* **34** (6), 064114.
- EMAMI-MEYBODI, H., HASSANZADEH, H., GREEN, C.P. & ENNIS-KING, J. 2015 Convective dissolution of CO₂ in saline aquifers: progress in modeling and experiments. *Int. J. Greenhouse Gas Control* **40**, 238–266.
- ENNIS-KING, J., PRESTON, I. & PATERSON, L. 2005 Onset of convection in anisotropic porous media subject to a rapid change in boundary conditions. *Phys. Fluids (1994–Present)* **17** (8), 084107.

- FELTHAM, D.L., UNTERSTEINER, N., WETTLAUER, J.S. & WORSTER, M.G. 2006 Sea ice is a mushy layer. *Geophys. Res. Lett.* **33** (14), L14501.
- FU, X., CUETO-FELGUEROSO, L. & JUANES, R. 2013 Pattern formation and coarsening dynamics in three-dimensional convective mixing in porous media. *Phil. Trans. R. Soc. A* **371** (2004), 20120355.
- GAMEIRO, M., SINGH, A., KONDIC, L., MISCHAIKOW, K. & MORRIS, J.F. 2020 Interaction network analysis in shear thickening suspensions. *Phys. Rev. Fluids* **5**, 034307.
- GRAHAM, M.D. & STEEN, P.H. 1994 Plume formation and resonant bifurcations in porous-media convection. *J. Fluid Mech.* **272**, 67–90.
- GROSSMANN, S. & LOHSE, D. 2000 Scaling in thermal convection: a unifying theory. *J. Fluid Mech.* **407**, 27–56.
- GROSSMANN, S. & LOHSE, D. 2001 Thermal convection for large Prandtl numbers. *Phys. Rev. Lett.* **86** (15), 3316.
- GUDHI 2025 GUDHI: geometry understanding in higher dimensions. Available at: <http://gudhi.gforge.inria.fr>.
- HEWITT, D.R. 2020 Vigorous convection in porous media. *Proc. Math. Phys. Engng Sci.* **476** (2239), 20200111.
- HEWITT, D.R., NEUFELD, J.A. & LISTER, J.R. 2012 Ultimate regime of high Rayleigh number convection in a porous medium. *Phys. Rev. Lett.* **108** (22), 224503.
- HEWITT, D.R., NEUFELD, J.A. & LISTER, J.R. 2013 Convective shutdown in a porous medium at high Rayleigh number. *J. Fluid Mech.* **719**, 551–586.
- HEWITT, D.R., NEUFELD, J.A. & LISTER, J.R. 2014 High Rayleigh number convection in a three-dimensional porous medium. *J. Fluid Mech.* **748**, 879–895.
- HIDALGO, J.J., FE, J., CUETO-FELGUEROSO, L. & JUANES, R. 2012 Scaling of convective mixing in porous media. *Phys. Rev. Lett.* **109** (26), 264503.
- HORTON, C.W., ROGERS, F.T.Jr. 1945 Convection currents in a porous medium. *J. Appl. Phys.* **16** (6), 367–370.
- HU, C., XU, K. & YANG, Y. 2023 Effects of the geothermal gradient on the convective dissolution in co₂ sequestration. *J. Fluid Mech.* **963**, A23.
- ILLINGWORTH, M., GU, B., CUMMINGS, L. & KONDIC, L. 2026 Influence of topology on performance of pore networks in membrane filters. *Phys. Rev. E* **113**, 024301.
- JIN, Y. 2024 A pore-scale resolved direct numerical simulation study for scaling analysis of the solutal convection in porous media. *J. Fluid Mech.* **1000**, A21.
- KONDIC, L., GOULLET, A., O’HERN, C.S., KRAMAR, M., MISCHAIKOW, K. & BEHRINGER, R.P. 2012 Topology of force networks in compressed granular media. *Europhys. Lett.* **97**, 54001.
- KRAMAR, M., GOULLET, A., KONDIC, L. & MISCHAIKOW, K. 2013 Persistence properties of compressed granular matter. *Phys. Rev. E* **87**, 042207.
- KRAMAR, M., GOULLET, A., KONDIC, L. & MISCHAIKOW, K. 2014a Evolution of force networks in dense particulate media. *Phys. Rev. E* **90**, 052203.
- KRAMAR, M., GOULLET, A., KONDIC, L. & MISCHAIKOW, K. 2014b Quantifying force networks in particulate systems. *Physica D* **283**, 37–55.
- KRAMAR, M., LEVANGER, R., TITHOF, J., SURI, B., XU, M., PAUL, M., SCHATZ, M.F. & MISCHAIKOW, K. 2016 Analysis of Kolmogorov flow and Rayleigh–Bénard convection using persistent homology. *Physica D* **334**, 82–98.
- LANDMAN, A.J. & SCHOTTING, R.J. 2007 Heat and brine transp. porous media: the Oberbeck–Boussinesq approximation revisited. *Transp. Porous Media* **70** (3), 355–373.
- LAPWOOD, E.R. 1948 Convection of a fluid in a porous medium. *Math. Proc. Camb. Phil. Soc.* **44**, 508–521.
- LOHSE, D. & SHISHKINA, O. 2024 Ultimate Rayleigh–Bénard turbulence. *Rev. Mod. Phys.* **96** (3), 035001.
- METZ, B., *et al.* 2005 *Carbon Dioxide Capture and Storage: Special Report of the IPCC*. IPCC.
- NIELD, D. & BEJAN, A. 2017 *Convection in Porous Media*. Springer.
- OTERO, J., DONTCHEVA, L.A., JOHNSTON, H., WORTHING, R.A., KURGANOV, A., PETROVA, G. & DOERING, C.R. 2004 High-Rayleigh number convection in a fluid-saturated porous layer. *J. Fluid Mech.* **500**, 263–281.
- PIROZZOLI, S., DE PAOLI, M., ZONTA, F. & SOLDATI, A. 2021 Towards the ultimate regime in Rayleigh–Darcy convection. *J. Fluid Mech.* **911**, R4.
- RIAZ, A., HESSE, M., TCHELEPI, H.A. & ORR, F.M. 2006 Onset of convection in a gravitationally unstable diffusive boundary layer in porous media. *J. Fluid Mech.* **548**, 87–111.
- SIMMONS, C.T., FENSTEMAKER, T.R., SHARP, J.M.Jr. 2001 Variable-density groundwater flow and solute transport in heterogeneous porous media: approaches, resolutions and future challenges. *J. Contam. Hydrol.* **52** (1–4), 245–275.

- SLIM, A.C. 2014 Solutal-convection regimes in a two-dimensional porous medium. *J. Fluid Mech.* **741**, 461–491.
- SLIM, A.C., BANDI, M.M., MILLER, J.C. & MAHADEVAN, L. 2013 Dissolution-driven convection in a Hele-Shaw cell. *Phys. Fluids (1994-Present)* **25** (2), 024101.
- SUZUKI, A., MIYAZAWA, M., MINTO, J.M., TSUJI, T., OBAYASHI, I., HIRAOKA, Y. & ITO, T. 2021 Flow estimation solely from image data through persistent homology analysis. *Sci. Reps.* **11**, 17948.
- TAGHIZADEH, K., LUDING, S., BASAK, R. & KONDIC, L. 2024 Understanding slow compression of frictional granular particles by network analysis. *Soft Matt.* **20**, 6440–6457.
- TRELLES, J.P. & DUFLY, J.J. 2003 Numerical simulation of porous latent heat thermal energy storage for thermoelectric cooling. *Appl. Therm. Engng* **23** (13), 1647–1664.
- VAN DER POEL, E.P., OSTILLA-MÓNICO, R., DONNERS, J. & VERZICCO, R. 2015 A pencil distributed finite difference code for strongly turbulent wall-bounded flows. *Comput. Fluids* **116**, 10–16.
- WANG, Y., XIE, J.H., YANG, W., LI, X., ABULAITI, Z., ZHENG, S., ZHU, J. & XU, K. 2025 Permafrost thawing under overlaying salt water. *Sci. Adv.* **11** (13), eadp2808.
- WELLS, A.J., HITCHEN, J.R. & PARKINSON, J.R.G. 2019 Mushy-layer growth and convection, with application to sea ice. *Phil. Trans. R. Soc. A* **377** (2146), 20180165.
- WEN, B., AKHBARI, D., ZHANG, L. & HESSE, M.A. 2018 Convective carbon dioxide dissolution in a closed porous medium at low pressure. *J. Fluid Mech.* **854**, 56–87.
- WEN, B., CORSON, L.T. & CHINI, G.P. 2015 Structure and stability of steady porous medium convection at large Rayleigh number. *J. Fluid Mech.* **772**, 197–224.
- WHITAKER, S. 1998 *The Method of Volume Averaging*, vol. 13. Springer Science & Business Media.
- XU, Y., REN, Q., ZHENG, Z.-J. & HE, Y.-L. 2017 Evaluation and optimization of melting performance for a latent heat thermal energy storage unit partially filled with porous media. *Appl. Energy* **193**, 84–95.
- ZHU, X., FU, Y. & DE PAOLI, M. 2024 Transport scaling in porous media convection. *J. Fluid Mech.* **991**, A4.