

Abstract

3D images of porous materials often exhibit topologically homogeneous structure, meaning any sufficiently large sub-volume statistically resembles any other. We investigate this homogeneity hypothesis using persistent homology (PH) on a very large 3D binary image of a porous medium, determining how topological features (connected components and loops) scale with sample size and whether their normalized measures converge, indicating a representative volume. The motivation is both scientific and computational: confirming homogeneity justifies analyzing smaller subdomains instead of the entire huge dataset, significantly reducing computational cost.

Motivation

Porous media are ubiquitous in nature, industry, and biomedical processes — groundwater, oil recovery, ice/permafrost, CO₂ sequestration, filtration, and tissue engineering. Their internal structure governs how fluids flow through them. Modern X-ray computed tomography (μ CT) produces high-resolution 3D information, but the datasets are often enormous: a single sample may comprise billions of voxels, presenting significant computational challenges.

Topological data analysis (TDA), particularly persistent homology, quantifies the complex pore networks within such materials, tracking connected components (pores), loops (flow pathways), and cavities across scales. However, computing PH on billion-voxel datasets is essentially impossible due to the extensive computational cost. A potential solution is to analyze representative subdomains. The ultimate goal — the “Holy Grail” — is to predict permeability from structure alone, motivating TDA on subdomains that are both representative and computationally manageable.

Data

A very large 3D binary image of a porous material (solid vs. void), from YTEC, Argentina (original dimensions 1288 × 2240 × 2240 voxels) [1].

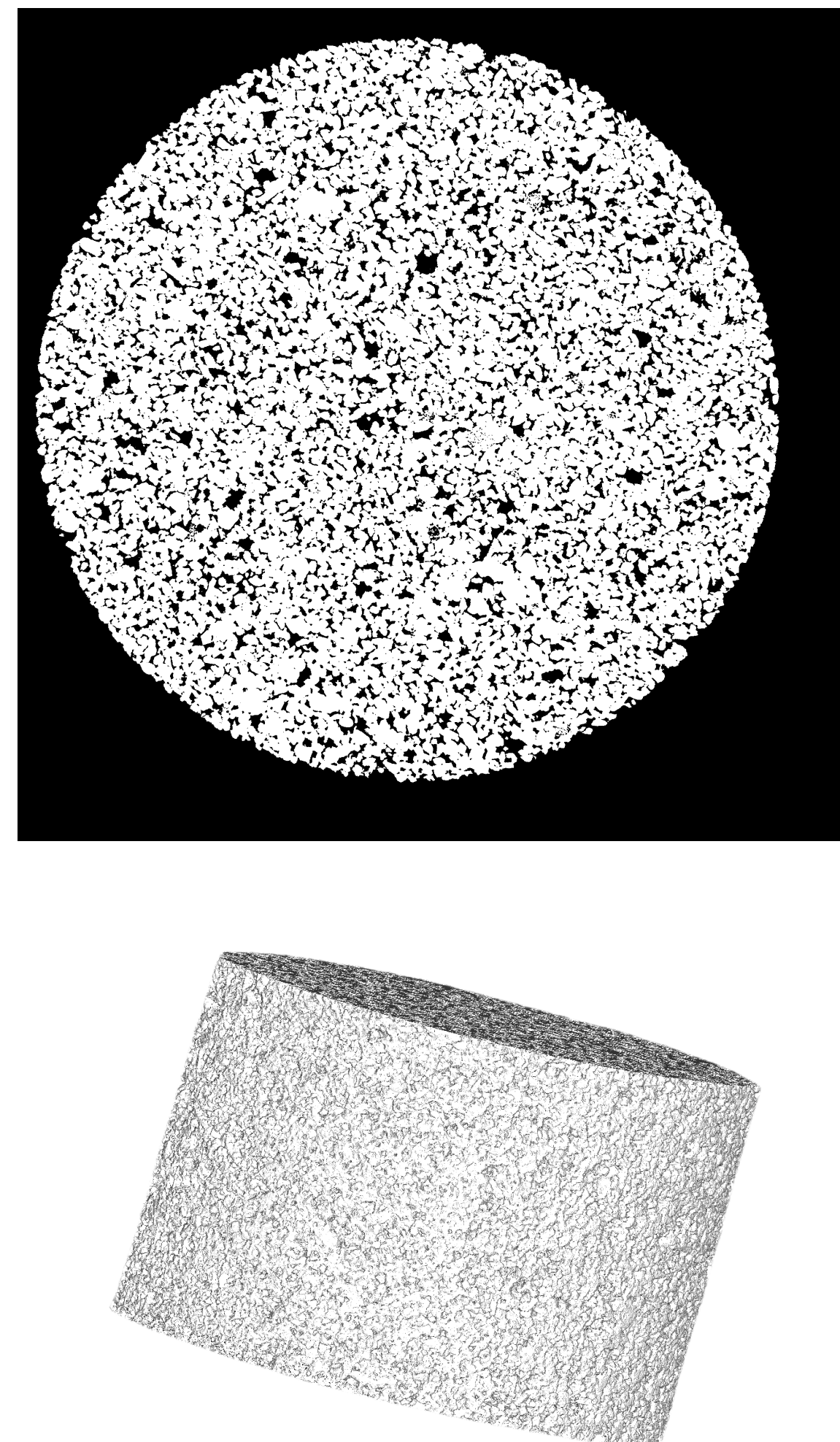


Figure 1. Top: 2D cross-section, 2240 × 2240 px. Bottom: 3D binary image, 1288 × 2240 × 2240 voxels.

Methodology

We extract cubic subdomains of increasing side length L (volume L^3), from 50^3 to 600^3 voxels. Overlapping (nested) cubes simulate an expanding window; non-overlapping cubes compare different physical locations. We use two complementary persistent homology approaches. Both produce persistence diagrams, which encode the birth and death of topological features across filtration scales. We summarize these features using Betti numbers: β_0 counts connected components, β_1 counts loops or tunnels, and β_2 counts enclosed cavities.

Alpha Complexes (via GUDHI [2])

The voxel coordinates of one phase (here the solid) form a point cloud; an alpha-complex filtration grows balls of increasing radius around them and records when components merge and loops form.

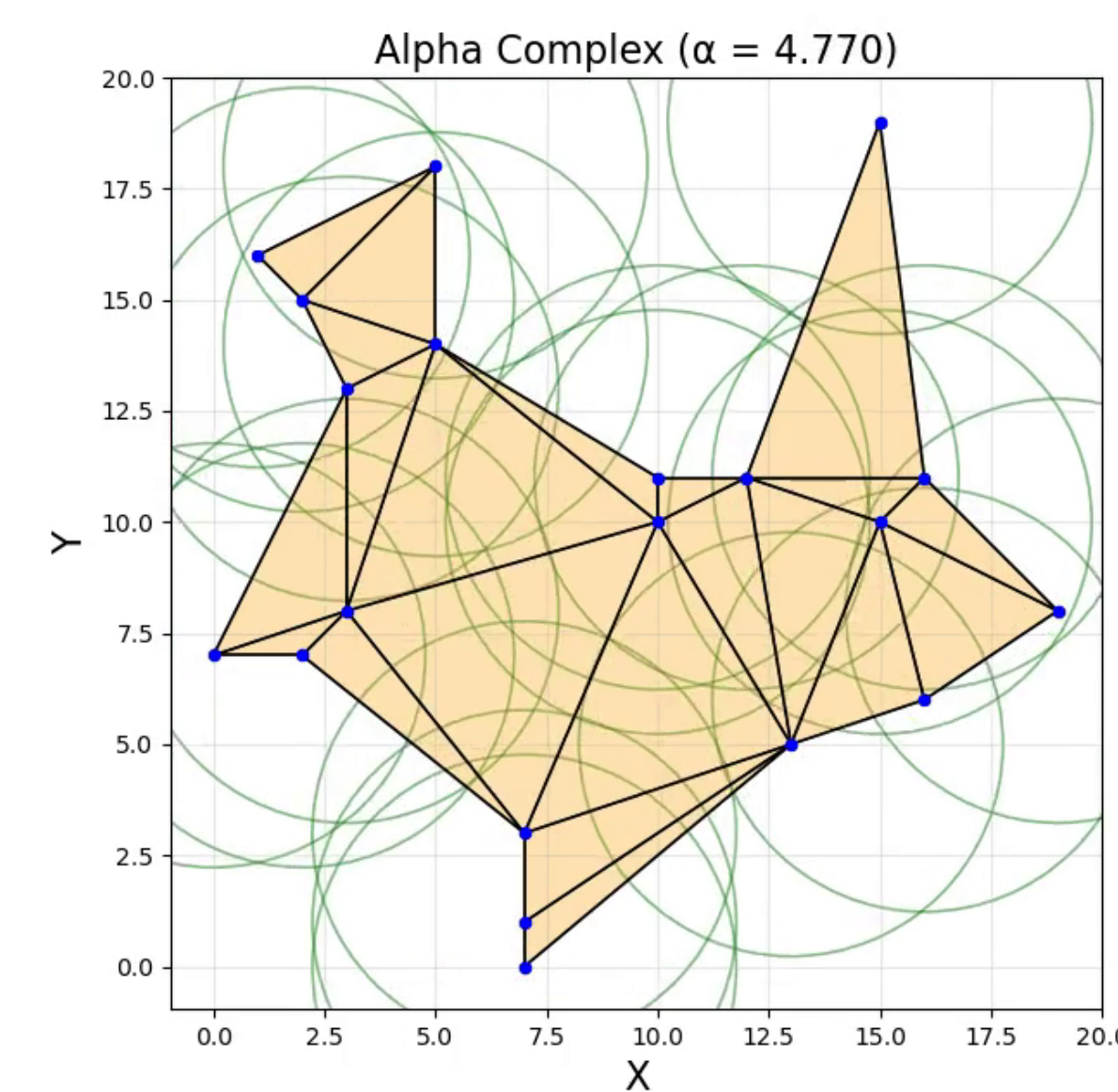


Figure 2. A 2D illustrative example (not the Fig. 1 data): as balls of increasing radius grow around randomly sampled points, simplices form and β_0 (components) and β_1 (loops) appear.

Distance-Transform PH (via HomCloud [3])

A signed Euclidean distance transform labels each voxel by its distance to the nearest solid-void interface (positive inside solid, negative in void). Sweeping a distance threshold from high to low (a *superlevel* filtration) progressively grows the solid phase; loops and cavities appear and are later filled, and the thresholds at which they appear/disappear are recorded in the persistence diagram. Figure 3 shows the solid phase as it appears at an intermediate value of the distance threshold.

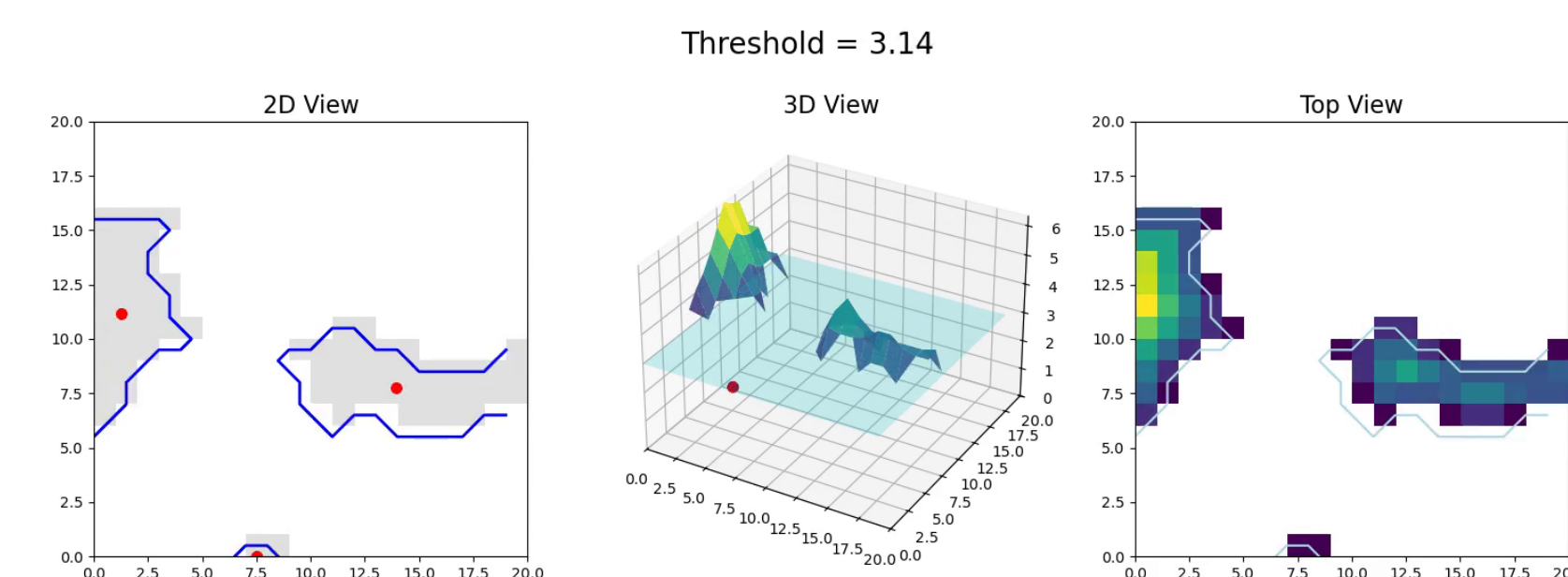


Figure 3. 2D, 3D, and top views of the distance-transform filtration at one threshold.

Normalization & Metrics

Per cube we count finite persistence pairs (“generators”) per dimension and sum lifespans to obtain *total persistence*, then normalize by volume $V = L^3$ to obtain feature densities, comparable across domain sizes.

Results

Convergence of topological metrics

Both feature density and total-persistence density converge to constant values as the cube side length L grows, plateauing by $L \approx 200$ –250. Both PH approaches (alpha complexes and the distance transform) show the same plateau. We take this converged length as the representative domain size, $D \approx 250$.

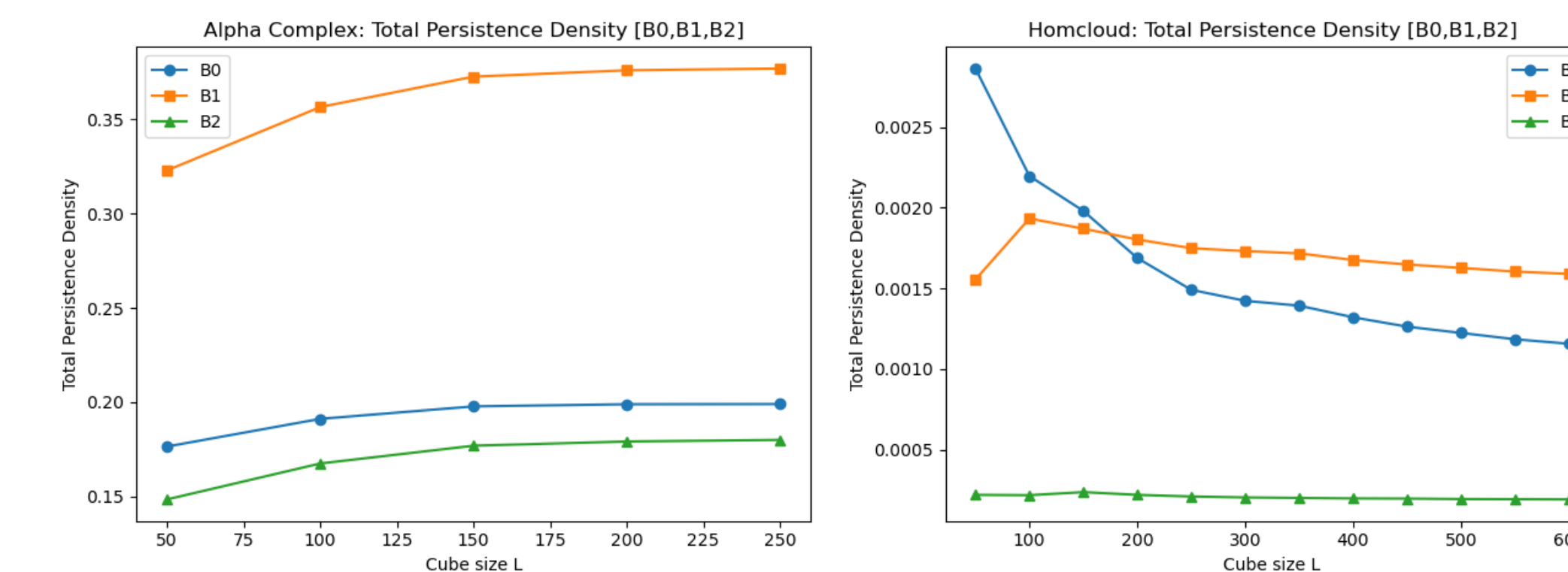


Figure 4. Total persistence density vs cube side length L (overlapping cubes): alpha complexes (left) and the Euclidean distance transform (right). Both level off, and confirm that the plateau is reached by $L \approx 200$ –250. (B_0 , B_1 , B_2 denote β_0 , β_1 , β_2 .)

Consistency across locations

Beyond a single expanding window, we compare separate non-overlapping subdomains taken at different locations; their topological measures are nearly identical, so we conclude that the structure is uniform in space.

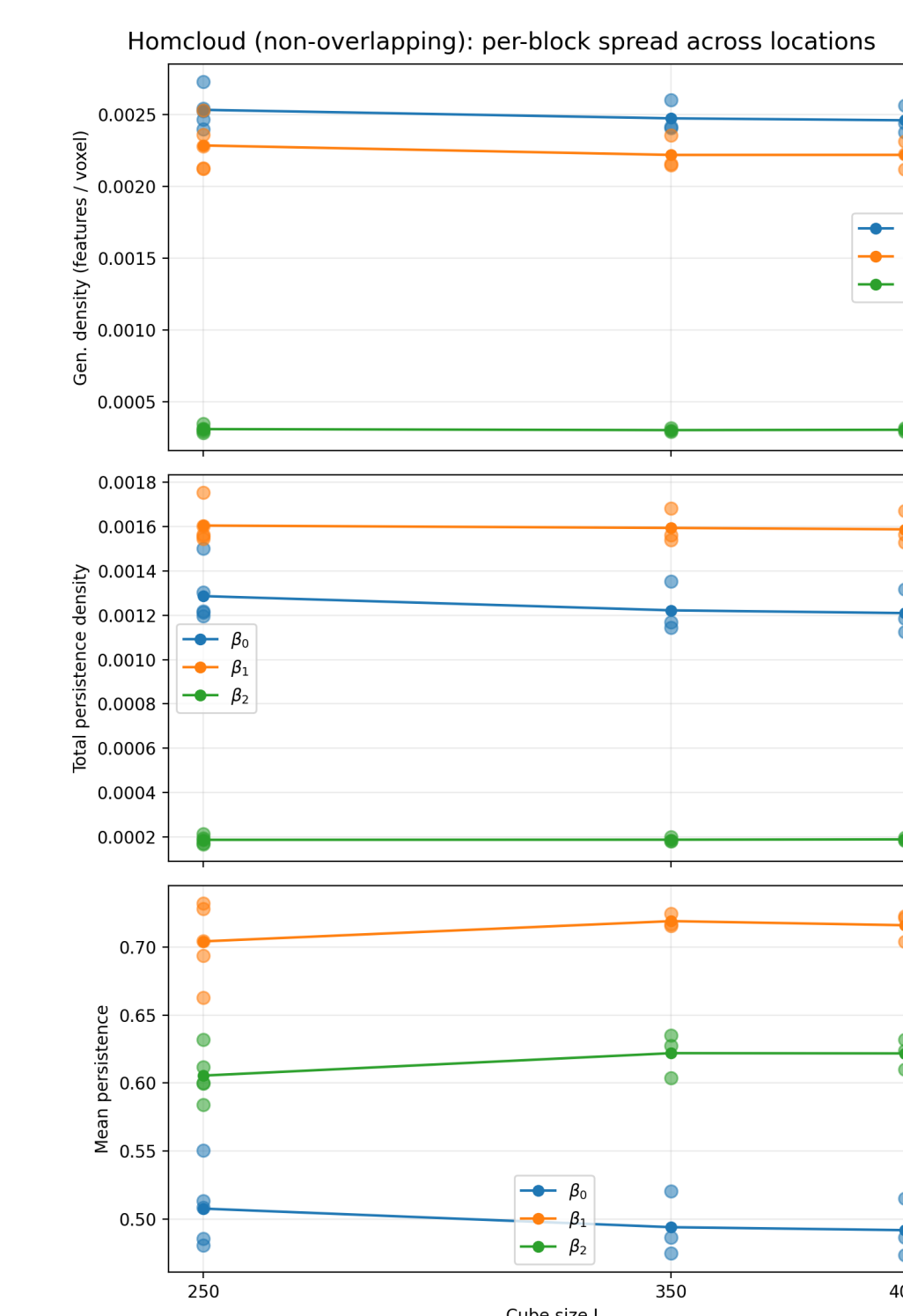


Figure 5. Per-block metrics for non-overlapping cubes (5 blocks at $L=250$, 3 each at 350, 400) plotted at their cube size (distance transform); each dot is one z -block, lines are per-size means. The block-to-block spread is only a few percent (and shrinks with L), so the topological measurement does not depend on where the subvolume is taken — direct evidence of homogeneity.

Homogeneity confirmed

The two very different PH approaches (alpha complexes vs. distance-based) and the stabilization of all metrics support the hypothesis that the material is topologically homogeneous.

Characteristic Length L_c

To attach a physical length scale to the pore structure, we sample straight lines through the binary images; the mean spacing between solid-void boundaries defines a characteristic length:

$$L_c = \frac{\text{Line Length}}{\text{Number of Transitions}}, \quad (1)$$

where transitions are solid-void boundaries. Averaging over many lines and orientations, the YTEC data give $L_c = 20.71 \pm 2.56$ voxels.

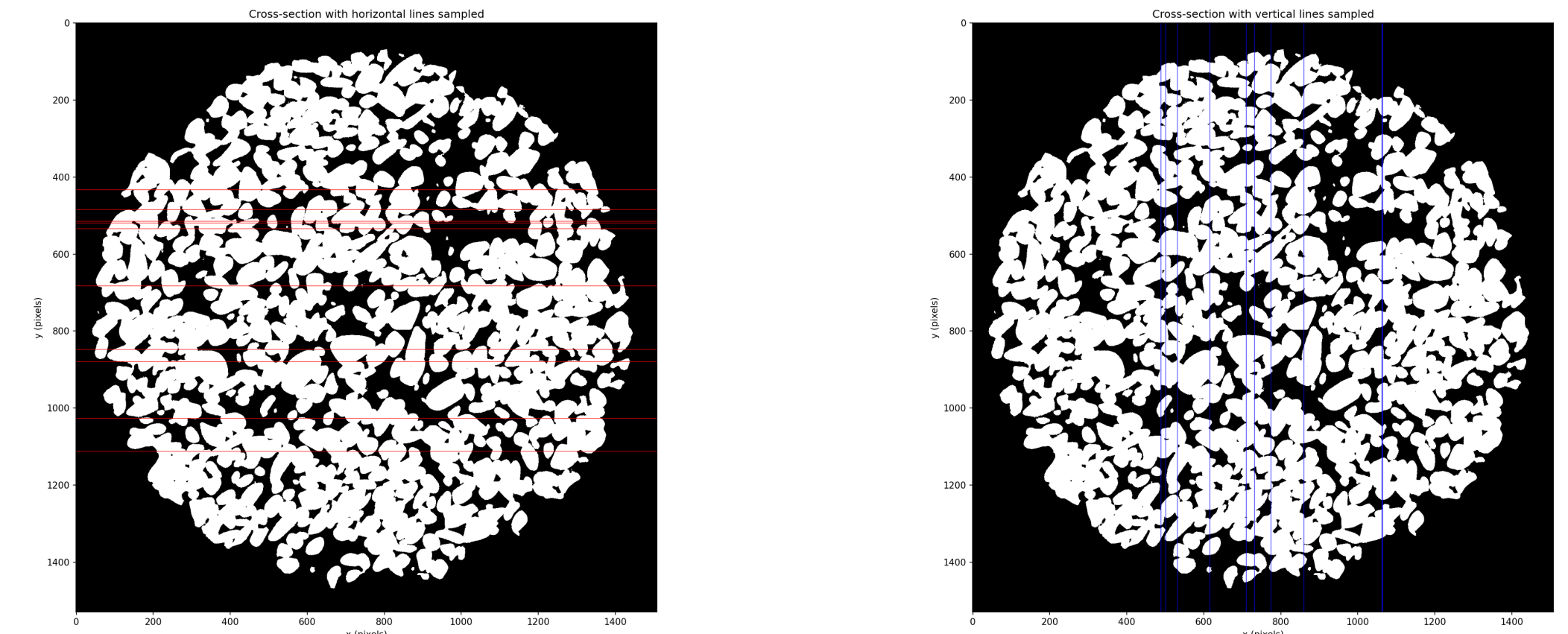


Figure 6. Cross-section with horizontal (left) and vertical (right) sample lines used to estimate L_c .

Toward a Universal R

We hypothesize a universal relationship between the characteristic length L_c and the domain size D at which topological measures converge:

$$R = \frac{D}{L_c}. \quad (2)$$

For YTEC data, $D \approx 250$ and $L_c = 20.71 \pm 2.56$ voxels give $R = 250/20.71 \approx 12$. If R is universal, one measures L_c , sets $D = R L_c$, and analyzes a single D^3 cube instead of the full billion-voxel image — estimating structure (and ultimately transport) from a representative subdomain.

Current Work

To verify the universality of R across different materials, ongoing work applies this analysis to additional datasets:

- **Synthetic data:** computer-generated porous media with controlled properties using the NASA-produced software PuMA [4].
- **Experimental data** from an external group of collaborators, led by Dr. Joaquín Jiménez-Martínez, who have shared with us the data from their recent work [5].

Summary

- Densities **converge by** ~ 200 –250³ ($D \approx 250$): a representative elementary volume for topology.
- Two independent PH methods agree across locations — **homogeneity confirmed**.
- Chord-length analysis gives $L_c \approx 20.7$ voxels, so $R \approx 12$ for YTEC data.
- **Ongoing:** test the universality of $R = D/L_c$ across synthetic and experimental datasets.

References

- [1] Data courtesy of Dr. E. Domene, YTEC, Argentina.
- [2] The GUDHI Project (2025). *GUDHI User and Reference Manual*, Ed. 3.11.0.
- [3] Suzuki, A., et al. (2021). Flow estimation solely from image data through persistent homology analysis. *Sci. Rep.*, 11, 17948.
- [4] Ferguson, J.C., et al. (2021). Update 3.0 to “PuMA: the porous microstructure analysis software”. *SoftwareX*, 15, 100775.
- [5] Velásquez-Parra, A., et al. (2024). Phase saturation control on vorticity enhances mixing in porous media. *Water Resour. Res.*, 60, R036628.