NJIT Grace Hopper
AI Research Institute
Seed Program

Understanding Fluid Flux: Topological Data Analysis of Porous Media

Ebru Dagdelen, Aakash Karlekar, Matthew Illingworth, Jonathan Jaquette, Linda Cummings, Lou Kondic

NSF Grant
DMS-2201627

Introduction

Porous media are fundamental in many natural and industrial processes, such as environmental remediation, secondary and tertiary oil recovery and CO_2 sequestration. The structure of these materials, with their interconnected pores and channels, directly influences the way fluids behave within them, making their characterization essential for optimizing processes. To study these structures, we use Topological Data Analysis (TDA), which captures information about the connectivity and geometry of the pore space through tools like persistent homology.However, when working with real-world data, especially from experimental images of rock samples, we encounter significant computational challenges. These datasets often contain a high level of complexity due to the intricate nature of the porous structure, as well as noise introduced during the imaging process. These factors can make it difficult to extract meaningful insights from the data. To address these challenges, it is essential to develop methods that reduce computational complexity and mitigate noise while preserving the key features.

Methodology

We propose to apply Gaussian convolution, across various smoothing levels, to noisy data as our denoising strategy. By evaluating the extent to which topological structure is preserved across the smoothing range, we aim to find to find an optimum smoothing level that minimizes the noise while preserving key features.

Datasets: Our datasets are 3D grayscale images, where each pixel is assigned with an integer value between 0 (darkest) and 255 (brightest) characterizing its brightness, assumed proportional to local material density. We first present results for artificial data (Figure 1), generated via 3D Fourier series with random coefficients before discussing the extension to real experimental data.

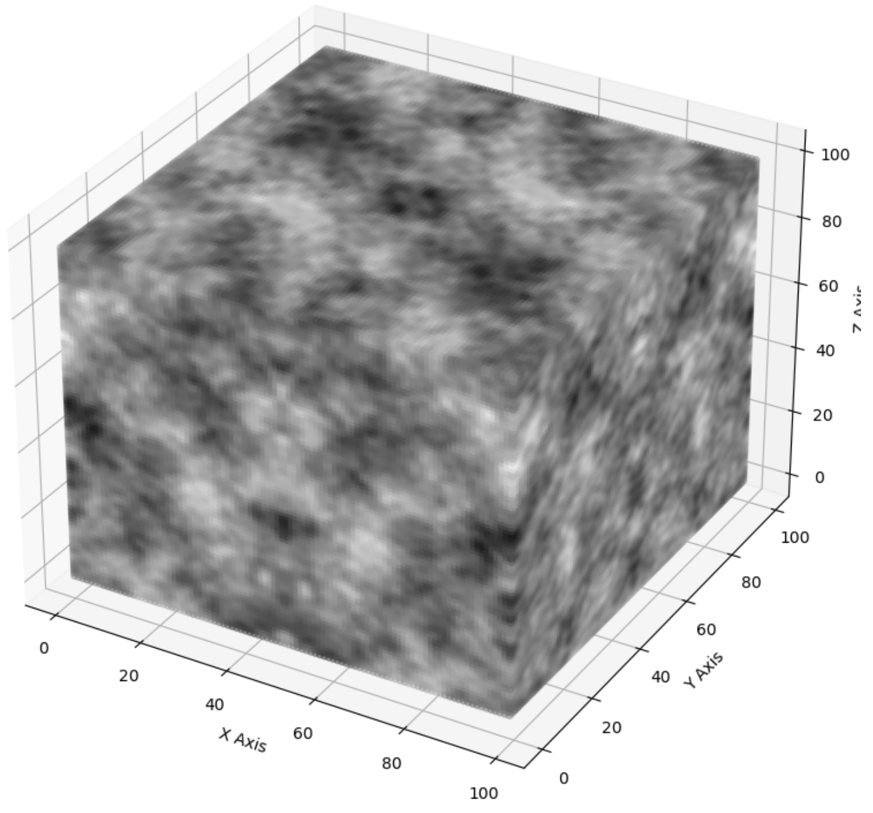
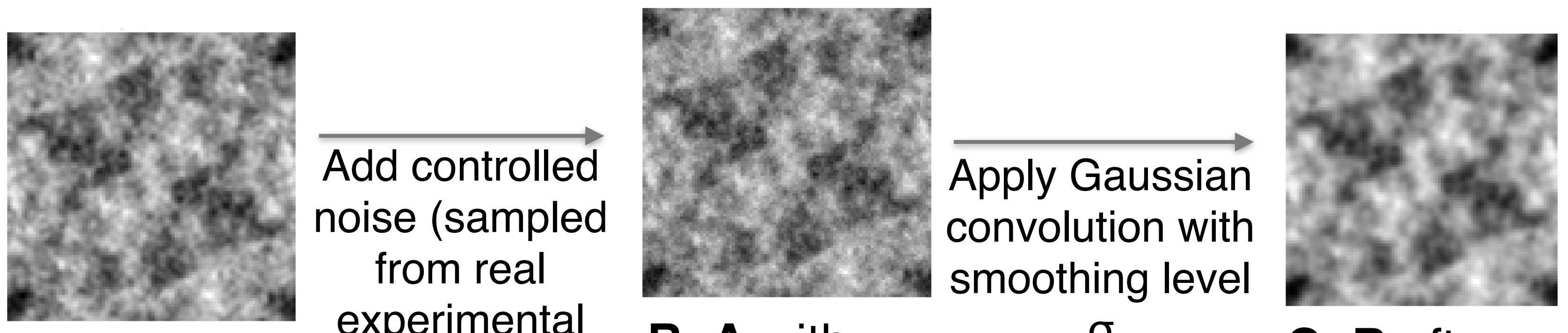


Figure 1: Simulated 3D Fourier dataset, 100 x 100 x 100 pixels.



A: 2D cross-section of original 3D dataset

Add controlled noise (sampled from real experimental scans)

B: A with added noise (uniform distribution)

Apply Gaussian convolution with smoothing level σ

C: B after smoothing

→ Compute topological data properties of 3D images using Gudhi [2], and Cubicle [3].

→ Compare measures utilizing the topological data properties of original (A) and smoothed (C) data.

Analysis

The persistence diagram (Figure 2) highlights the most persistent topological features in our data, which could indicate regions of high fluid flow through the porous media.

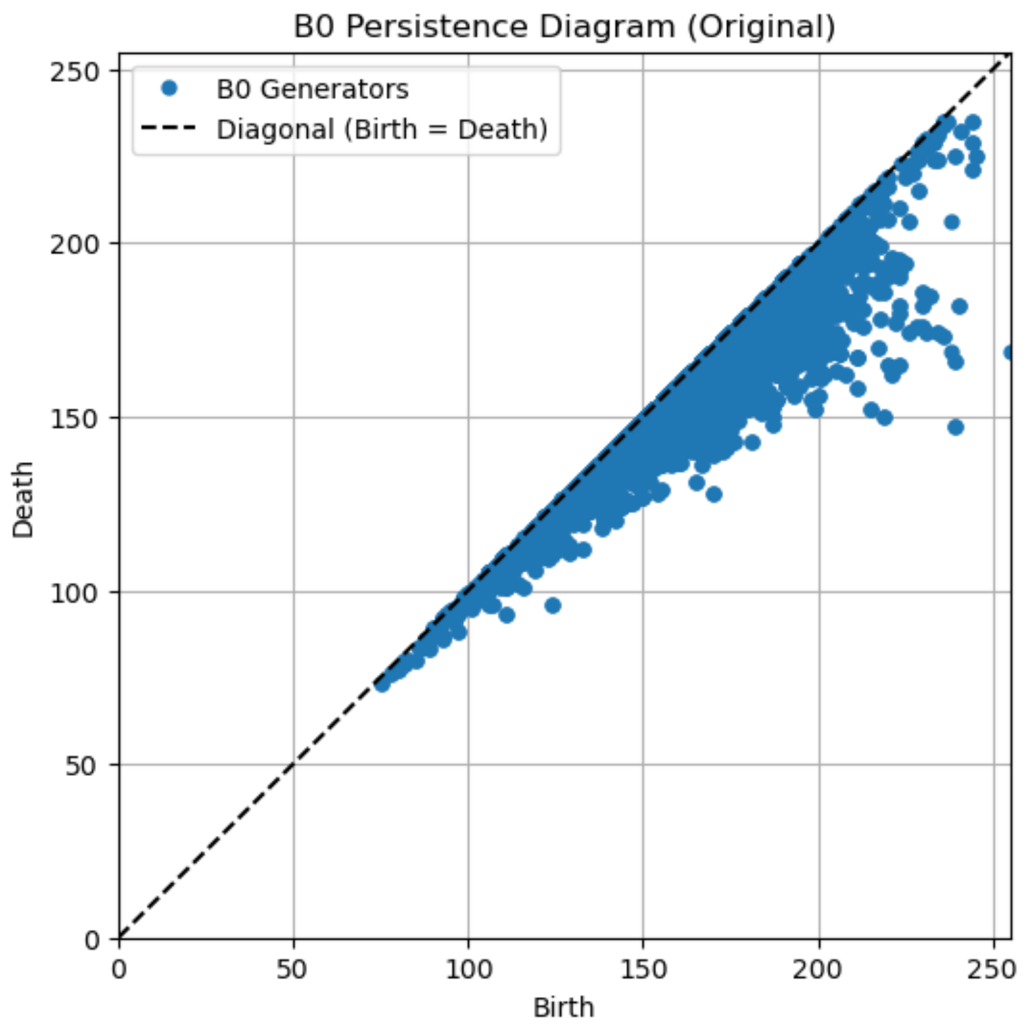


Figure 2: Persistence Diagram (connected components)
Each point represents a topological feature's birth and death over time.

Overview of Persistence Diagrams (PDs):
Grayscale images are analyzed using thresholding based on the pixel brightness, assumed to correlate with material density. To study the topological structure, **Betti numbers** are computed to quantify different topological features: β_0 represents the number of connected components, β_1 corresponds to the number of loops, and β_2 represents the number of cavities. PDs are created to visualize the birth and death of these features as the brightness threshold is lowered, providing insight into their significance within the sample.

The **birth** of a feature refers to the first pixel brightness threshold at which it appears, while its **death** occurs when the feature either fills in (for loops) or merges with another component. Features that live for a short period appear close to the diagonal of the PD and are often considered less significant (or to be due to noise). In contrast, long-lived features are farther from the diagonal and are considered more important or persistent.

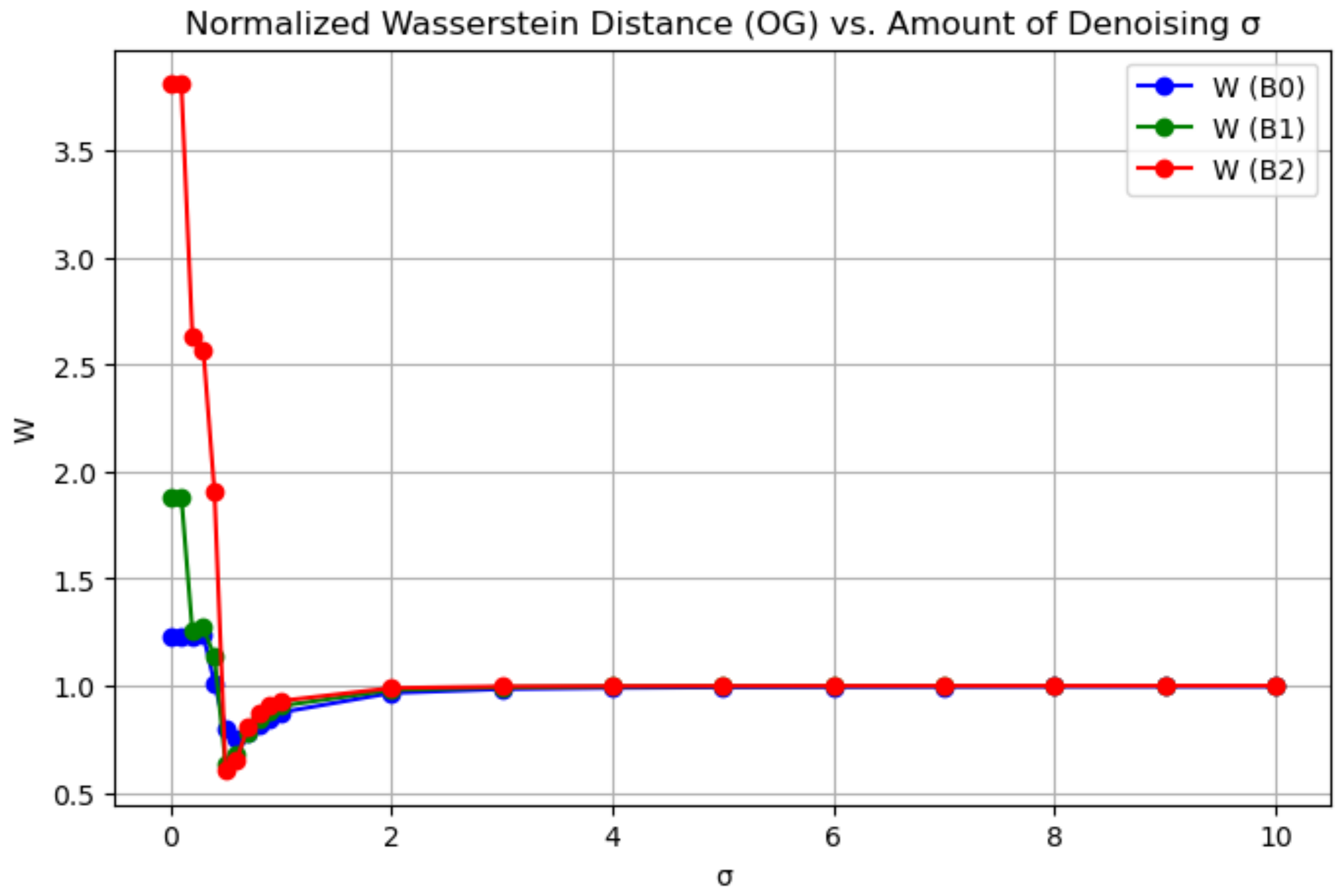


Figure 3: Normalized Wasserstein Distance vs. Amount of denoising

To quantify how well our denoising process works, we use the Wasserstein distance metric [2] to measure the similarity between the PDs for the original and the smoothed datasets (Figure 3) [2]. The Wasserstein distance compares two PDs by finding an optimal pairing between their points and summing the distances between each pair. This provides a global measure of difference that captures all changes between diagrams. For each Betti type, the Wasserstein distances computed between the original and denoised PDs are normalized by the Wasserstein distance between the PDs for the original dataset and a zero baseline. This normalization allows for meaningful comparison across Betti types and denoising levels.

Results

Our methodology has yielded the following outputs:

- Certain σ-values (characterizing the smoothing level) minimize the difference between the original and the noisy datasets.
- Preliminary results suggest a reliable denoising approach for future use on computationally complex grayscale data.

Future Work

We plan to extend our method to more complex datasets, such as the image of the coquina sample shown in Figure 7, provided by experimental collaborators [4]. These structures present greater computational challenges due to their size and complexity. By applying our method to these datasets, we aim to investigate whether strong correlations exists between fluid flux through the porous materials and their topological properties.

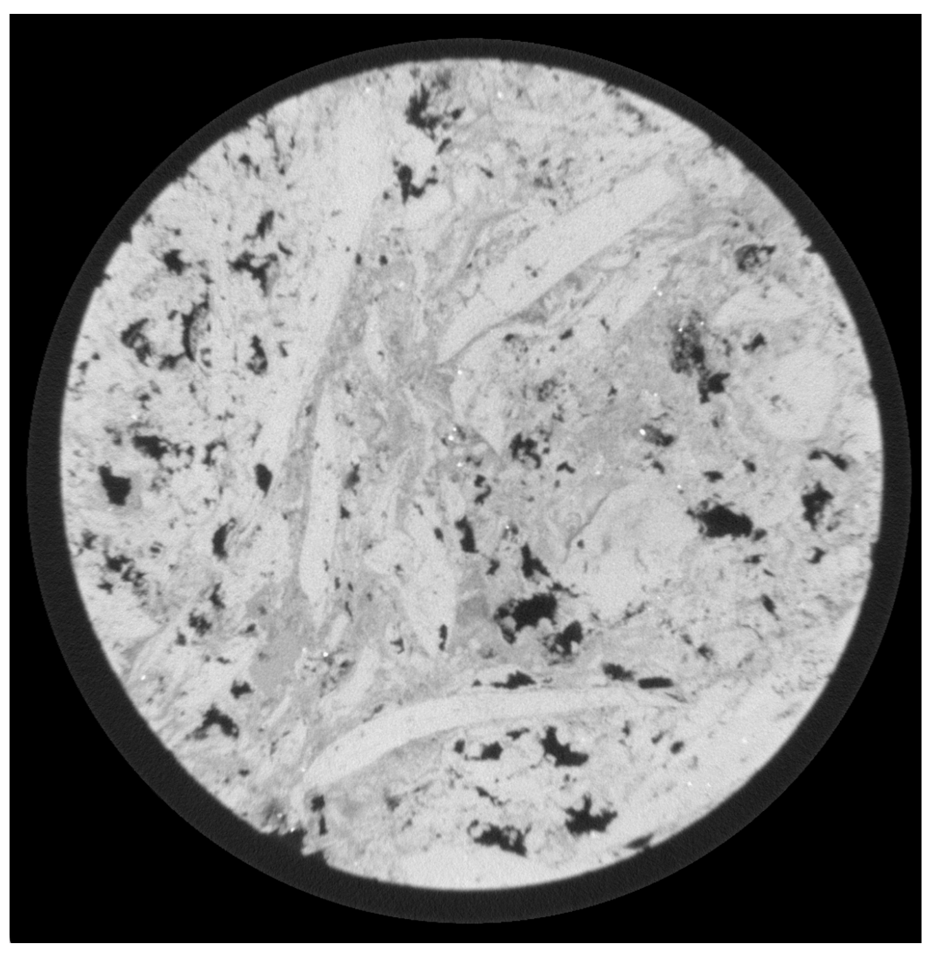


Figure 4: Cross-section of man-made coquina sample from PUC-Rio Brazil; image resolution: 40 microns. 1001 x 1024 pixels.

References

[1] Santos, J. E. 3D Dataset of Simulations. Available at: https://doi.org/10.17612/93pd-y471

[2] The GUDHI Project. *GUDHI User and Reference Manual.* GUDHI Editorial Board, Edition 3.11.0, 2025. Available at: https://gudhi.inria.fr/doc/3.11.0/

[3] Wagner, Hubert. Slice, Simplify and Stitch: Topology-Preserving Simplification Scheme for Massive Voxel Data. In 39th International Symposium on Computational Geometry (SoCG 2023), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

[4] M. Carvalho, PUC - Rio de Janeiro, Brazil - private communication