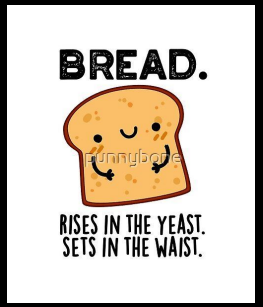
ETH zürich

Quantitative Big Image Spring'24
30th May, 2024

QUANTIFYING PORE DISTRIBUTIONS OF BREAD SAMPLES USING X-RAY MICROTOMOGRAPHY



Saksham Pruthi | spruthi@student.ethz.ch

Mentored by Dr. Anders Kaestner, PSI

INTRODUCTION

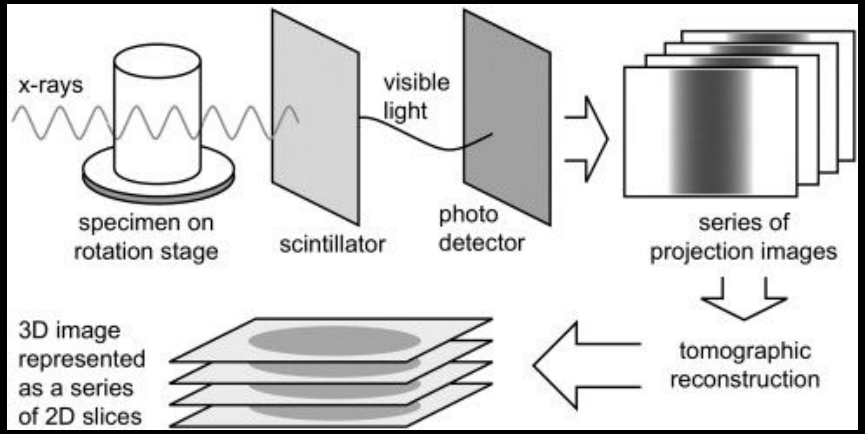
The texture and quality of bread are significantly influenced by the distribution and characteristics of pores formed during the baking process.

In our study, we try to analyze the impact of sourdough culture on pore development, by employing the advanced imaging technique of x-ray microtomography.

CONTENTS

- Image Analysis
  - Preprocessing
  - Pore Extraction
  - Segmentation
- Pore Analysis
  - Porosity, Thickness maps
  - Quantification of Metrics
    - Size (Volume), Surface Area, Length (Bounding Box), Sphericity
    - Size Distribution, Radial Density Distribution
- Statistics
  - Comparison of Distributions
  - Clustering (Classification)

X-RAY TOMOGRAPHY

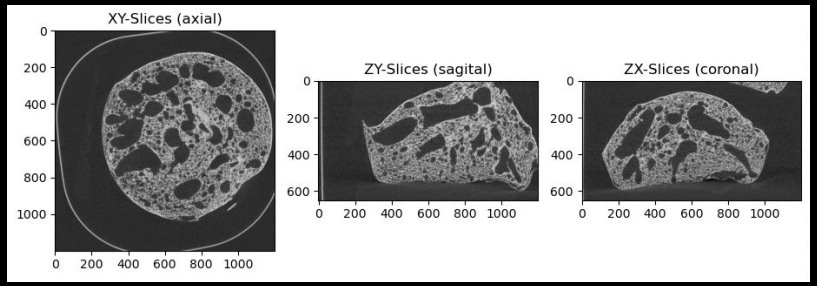


- Uses X-rays for cross-sectional 2d images of objects non-destructively
- 2d projections are used to reconstruct 3d object
- Higher resolution than CT scans

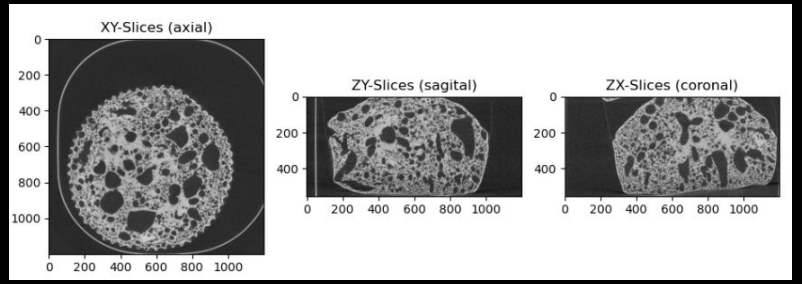
- Absorption of light is a function of the absorptivity of the material - intensities related through the line integral of material absorptivities along path

(Landis & Keane, 2010)

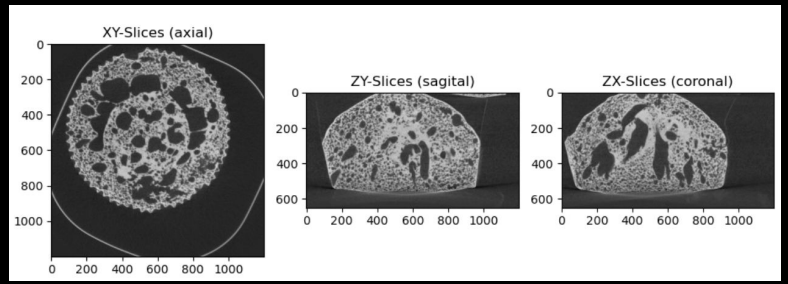
SAMPLES



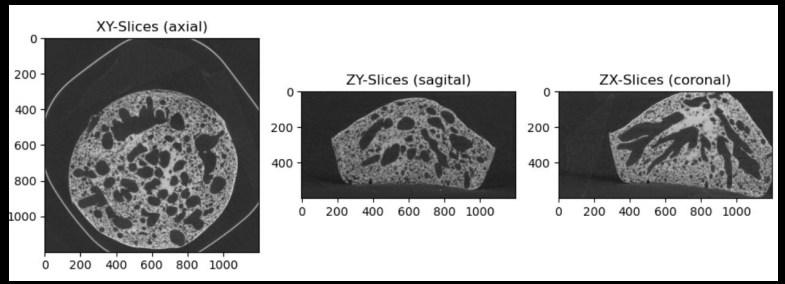
Ro1a (150:799 & 1420:2069)



Ru1a (299:853 & 1440:1994)



V1a (180:829 & 1420:2069)



W1a (300:899 & 1440:2039)

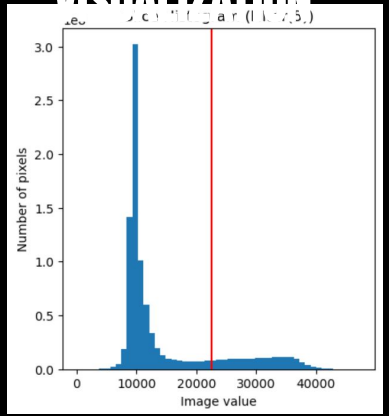
PREPROCESSING PIPELINE

Anisotropy Interference → Down-sampling → Filtering → Thresholding

CHALLENGES IN PREPROCESSING

- Anisotropy
- Container
- Merge errors
- Loss of small pores

THRESHOLD VISUALIZATION

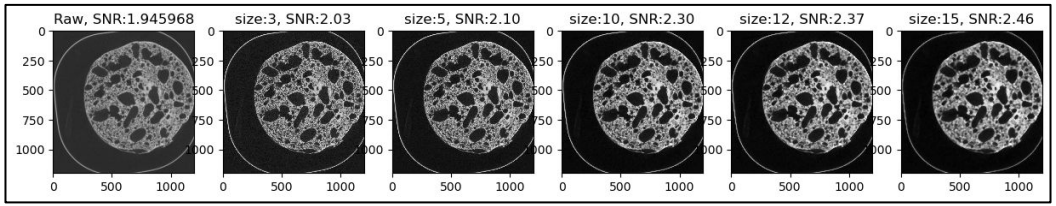


SPECIFIC PREPROCESSING PIPELINE

No Downsampling

FFT Anisotropy

Median Filtering (Size window = 5)



Ro1a

Hysteresis Thresholding

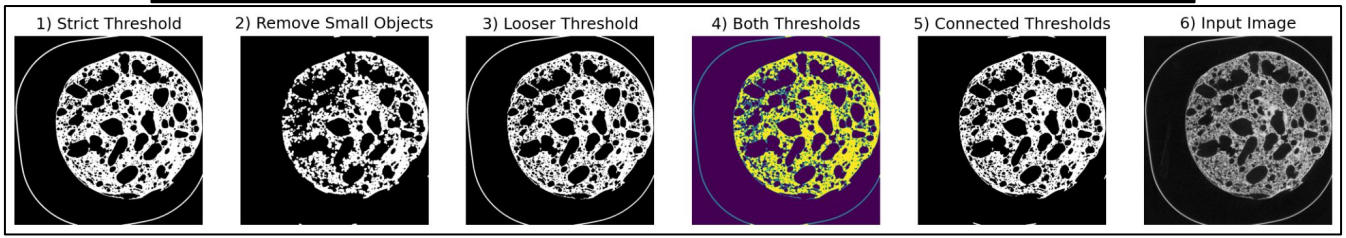
HYSTERESIS THRESHOLDING

Otsu Threshold - Range of Threshold

Morphological Opening of Strict Threshold

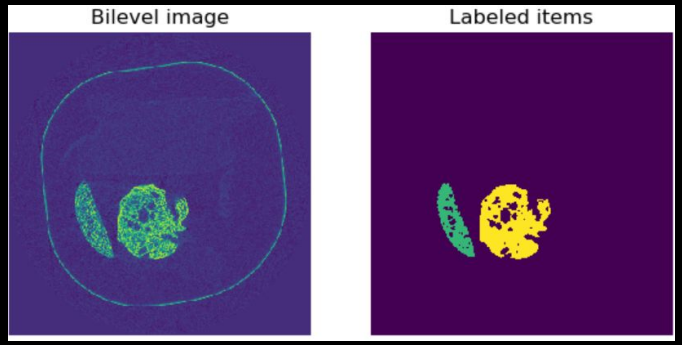
Both Thresholds = Linear combination of looser and remove small objects

Linear combination of looser and morphologically dilated remove small objects

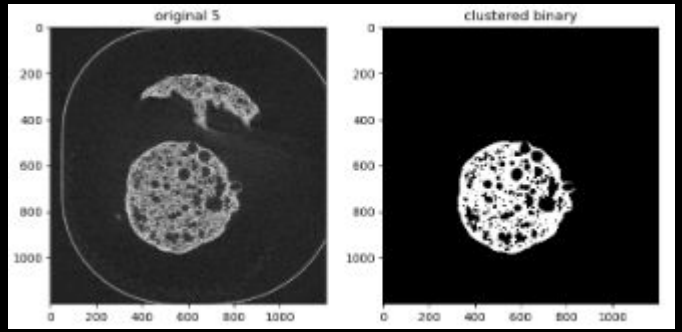


Ro1a

CLUSTERING SAMPLES - EXTRACTION

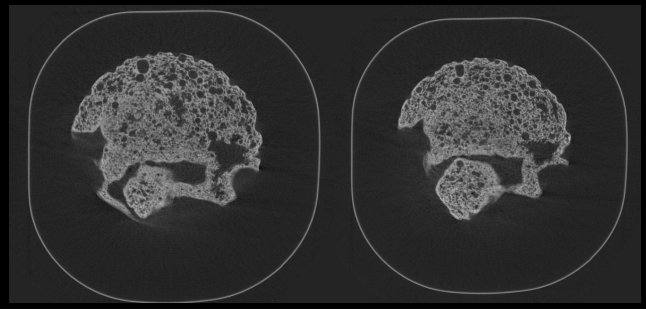


Ro1a

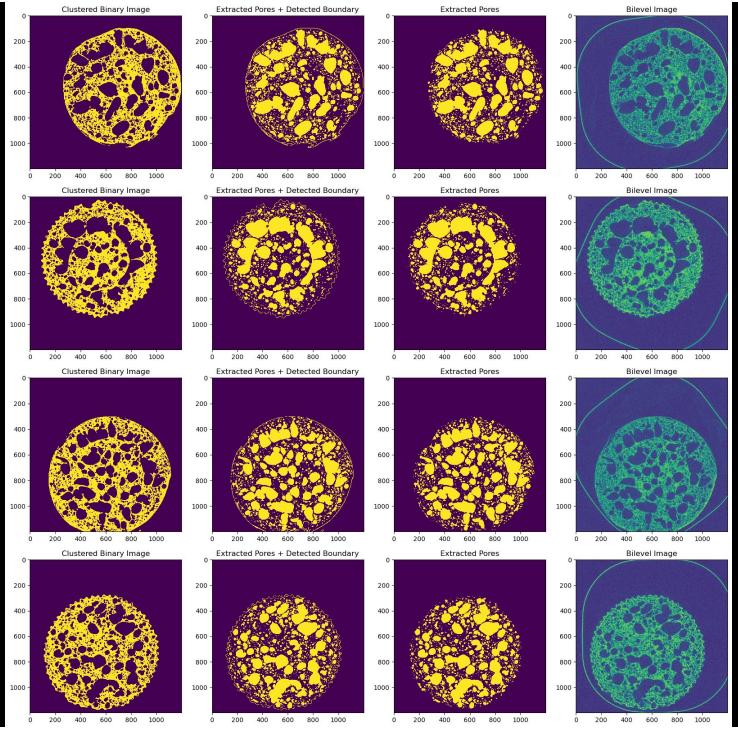


Ru1a

- K-means clustering approach
- Extract single label - connected components labelling approach
- Extract multiple labels - connected components labelling & area sorting approach
- Problem



PORE EXTRACTION



- Fill Holes - Filled Pores
- Sobel Filter - Edge Detection
- XOR Operation - Actual Pore Extraction within Sample Volume

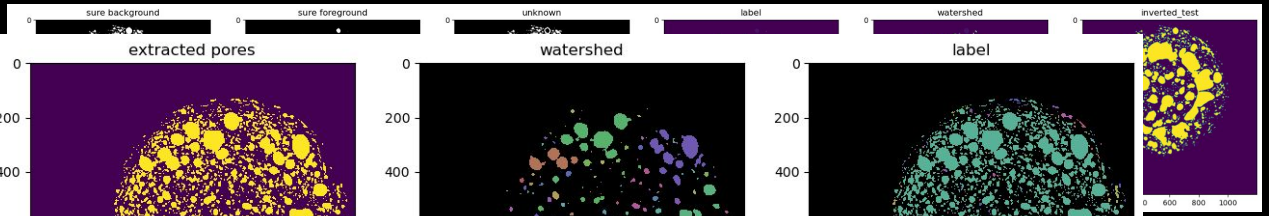
- Structure given to binary_fill_holes
  - Larger - Missed smaller holes
  - Smaller - Merge errors
- Other methods - missed larger pores, randomly added smaller pores

SEGMENTATION OF PORES

LABEL

Automa
Compo
labellin
scipy.ndimage

WATERSHED

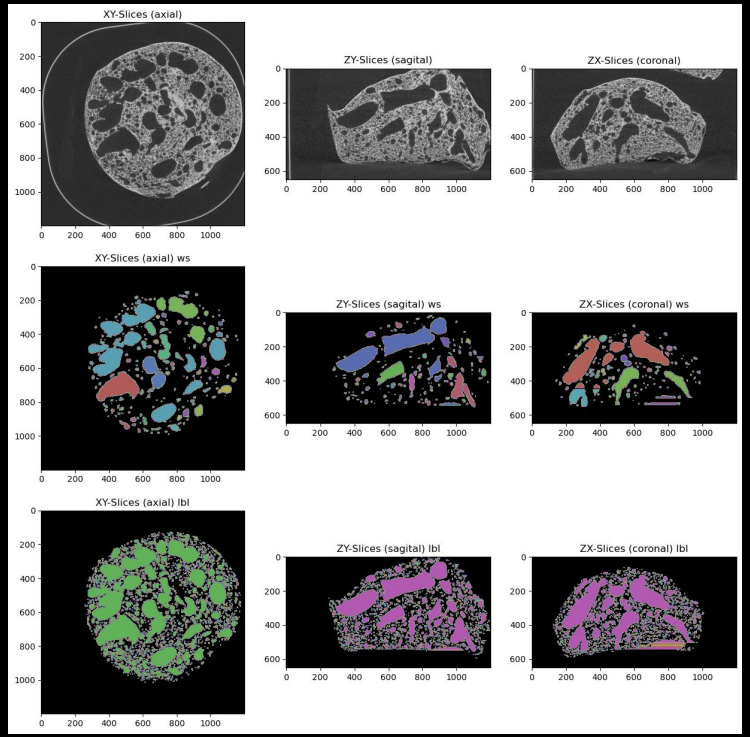


- Unknown set as sink (0) in labelled sure foreground
- Labelled sure foreground used as Marker during Watershed

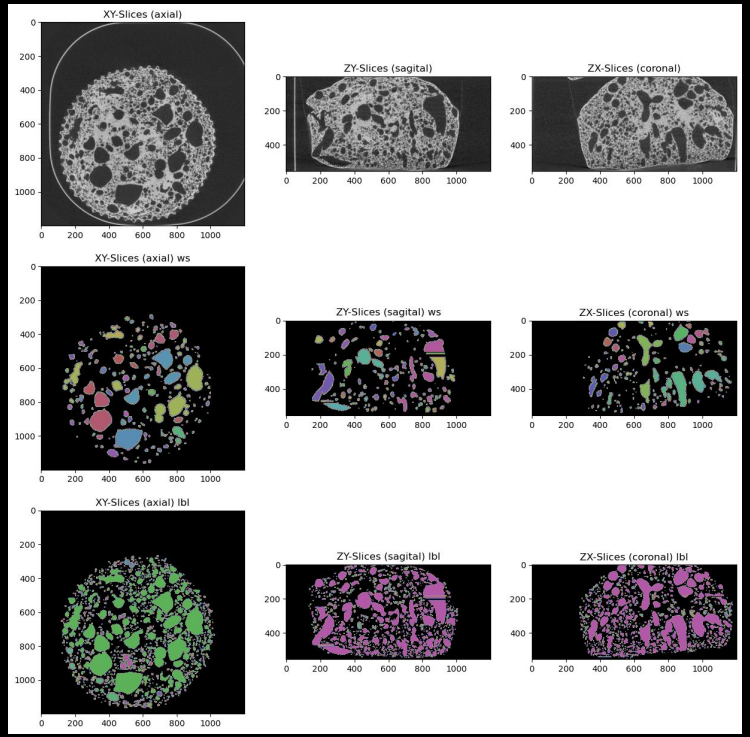
(https://docs.opencv.org/4.x/d3/db4/tutorial_py_watershed.html)

SEGMENTED SAMPLES

Ro1a (150:799 & 1420:2069)

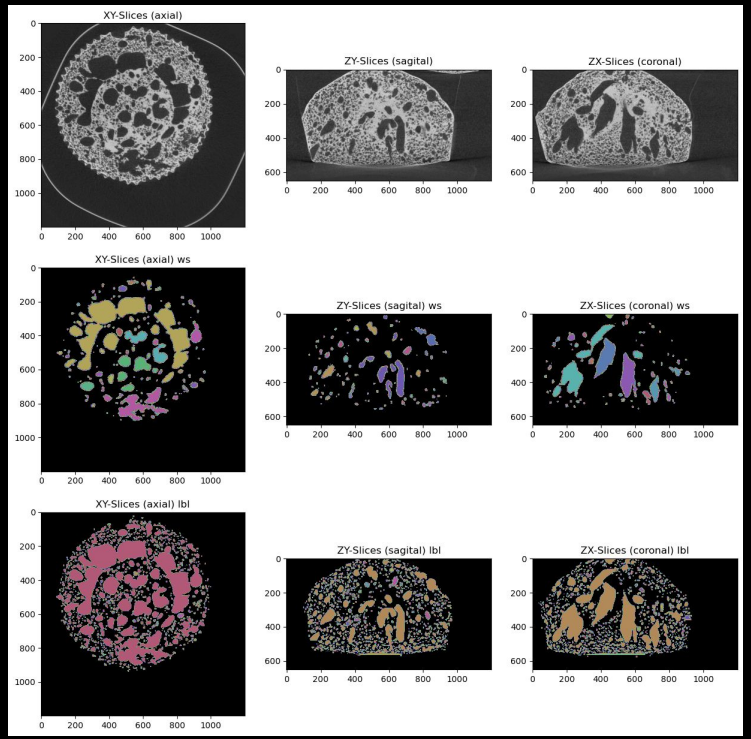


Ru1a (299:853 & 1440:1994)

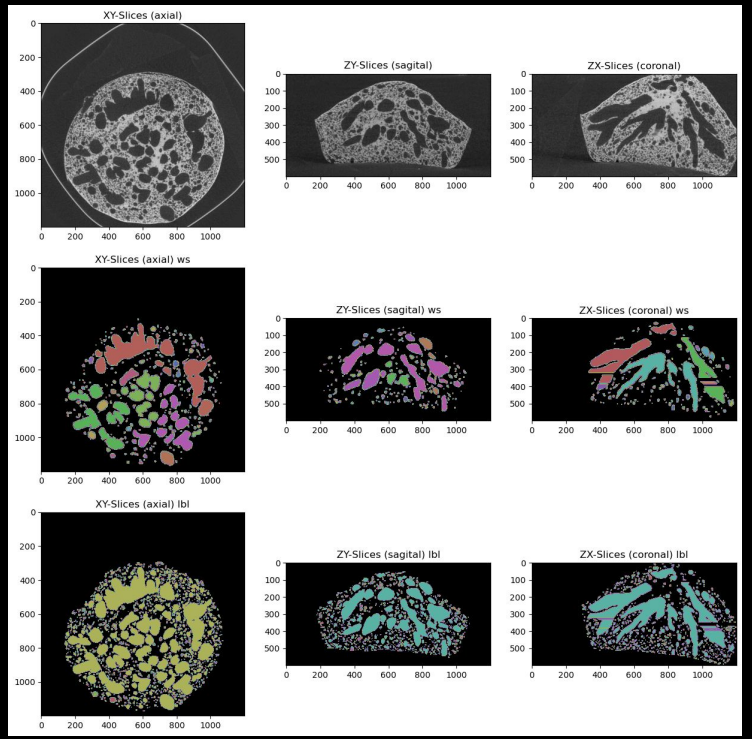


SEGMENTED SAMPLES

V1a (180:829 & 1420:2069)



W1a (300:899 & 1440:2039)



FILTERING SEGMENTS

SIZE-BASED FILTERING

- Using a specified size threshold (number of voxels) to reduce smaller pores
- Currently using size-thresh = 10 (ad-hoc)
- About 5-7% regions removed for ws segmentation
- About 25-30% regions removed for lbl segmentation

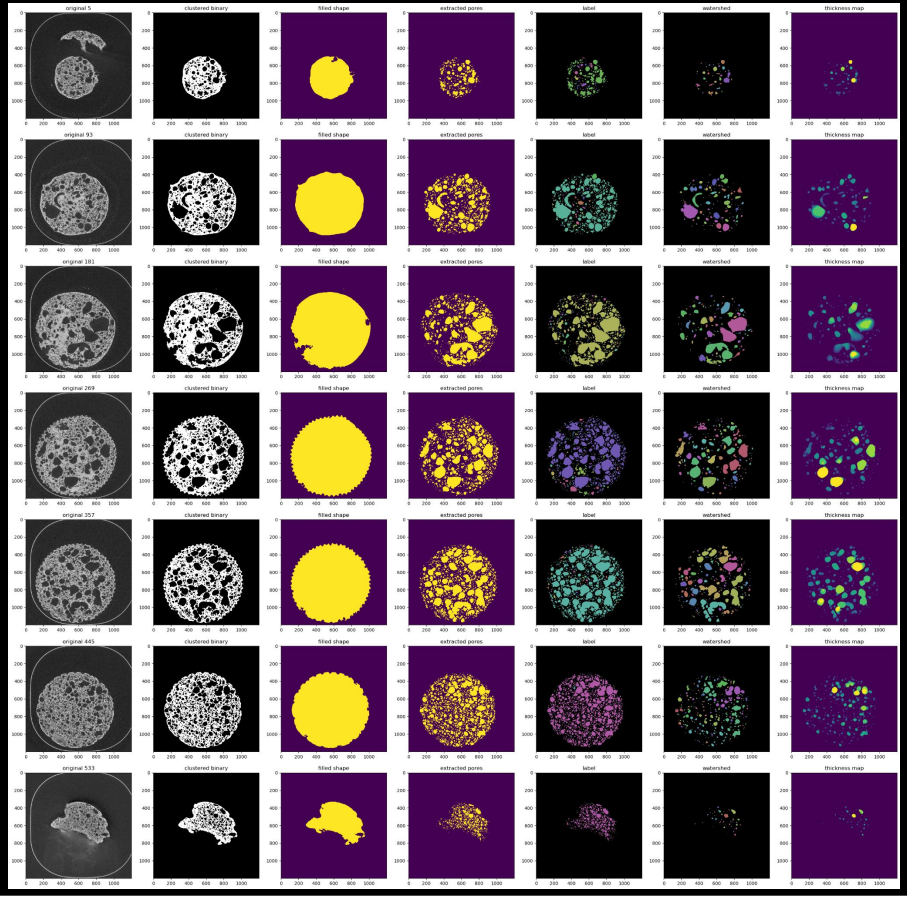
MERGING REGIONS

- Preliminary agglomeration step
- Using a specified distance based threshold to combine segmented pores
- Find objects inside labelled pores - distance transform
- If the distance transform value per centroid of pore < distance thresh => get 26-neighbours and combine if labelled too
- Currently not using due to problem of merge errors

ETH zürich

QBI'24

BREAD IMAGE - PIPELINE AT A GLANCE



Ru1a

COMPARING SEGMENTATION

	Ro1a	Ru1a	V1a	W1a
Dice Coefficient	0.697	0.684	0.664	0.748
Jaccard Score	0.780	0.756	0.788	0.789
Jaccard Index	0.536	0.520	0.497	0.598

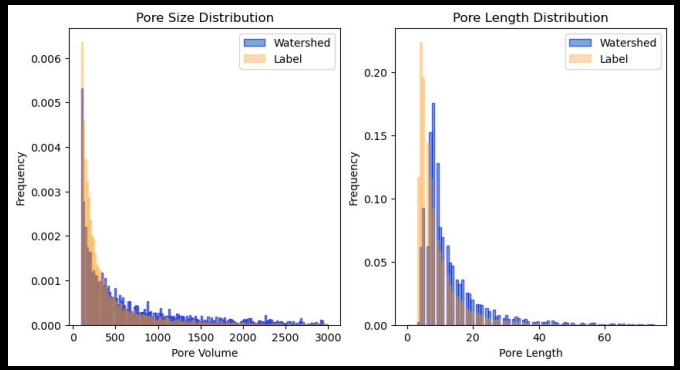
- Dice Coefficient - Gauge similarity of 2 samples - 2*intersection/sum of areas
- Jaccard score - Intersection / Union
- Jaccard Index (scikit learn's implementation) - micro averaging - Global calculation

POROSITY

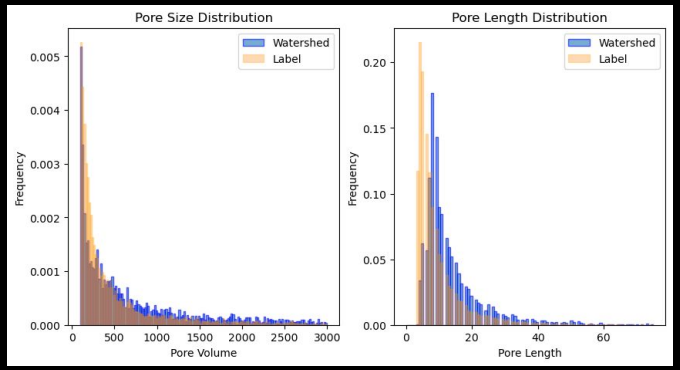
	Ro1a	Ru1a	V1a	W1a
WS segmentation	23.015%	19.021%	21.605%	24.821%
LBL segmentation	41.67%	37.983%	38.961%	41.965%

- Porosity = total number of voxels labelled / Total number of voxels inside sample

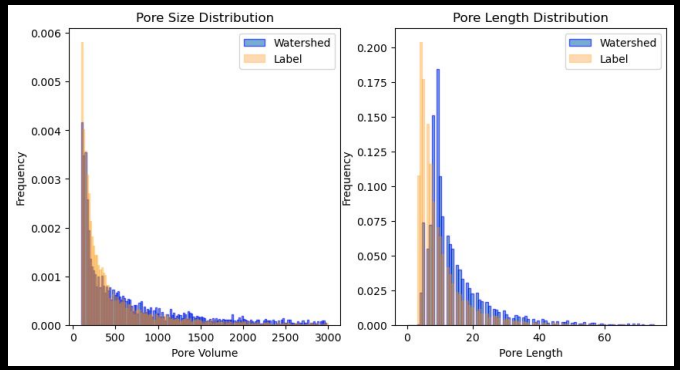
PORE SIZE DISTRIBUTIONS



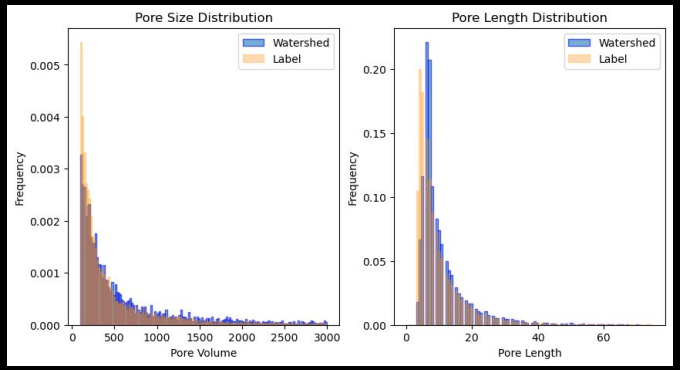
Ro1a (150:799 & 1420:2069)



Ru1a (299:853 & 1440:1994)

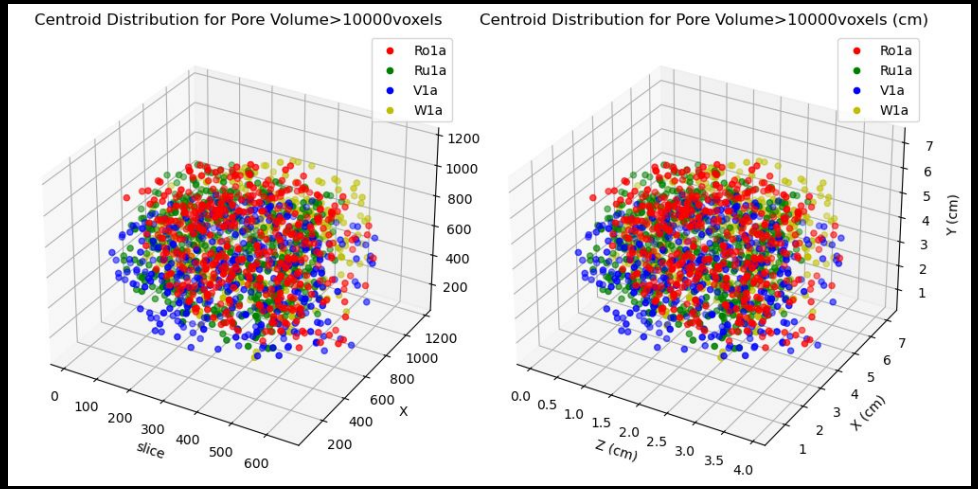


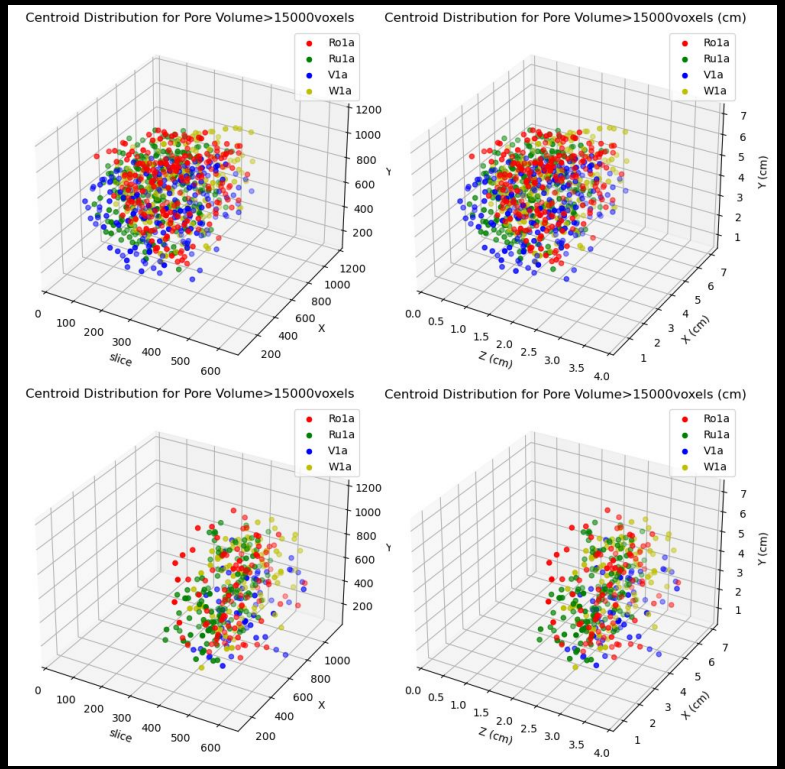
V1a (180:829 & 1420:2069)



W1a (300:899 & 1440:2039)

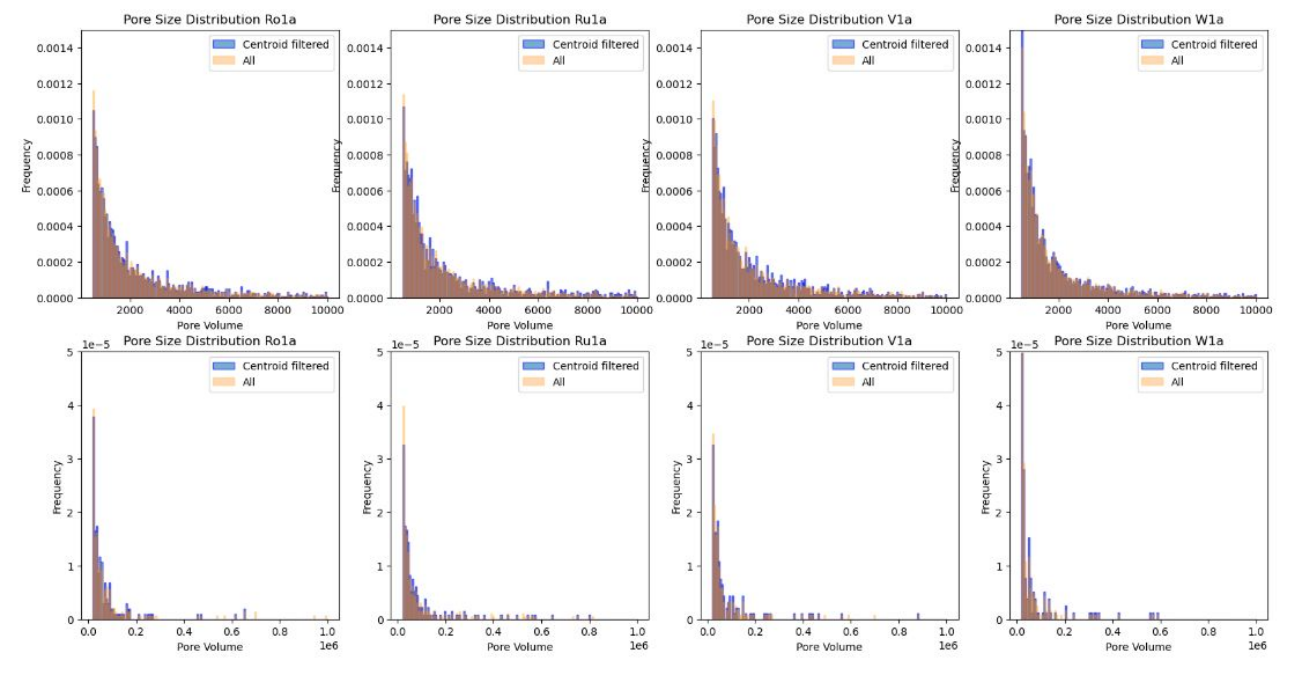
CENTROID PLOTS





PORE SIZE DISTRIBUTIONS - Z FILTERED

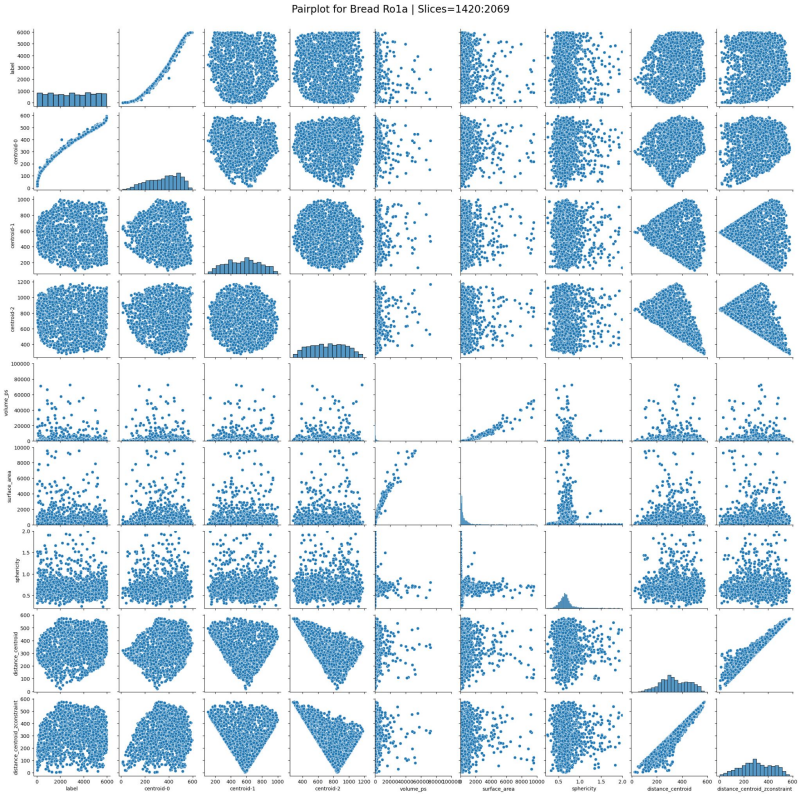
```
Number of large (>15000) pores after z-filtering for bread Ro1a (%) = 69.8051948051948%
Number of large (>15000) pores after z-filtering for bread Ru1a (%) = 66.75461741424802%
Number of large (>15000) pores after z-filtering for bread V1a (%) = 73.4126984126984%
Number of large (>15000) pores after z-filtering for bread W1a (%) = 63.984674329501914%
```

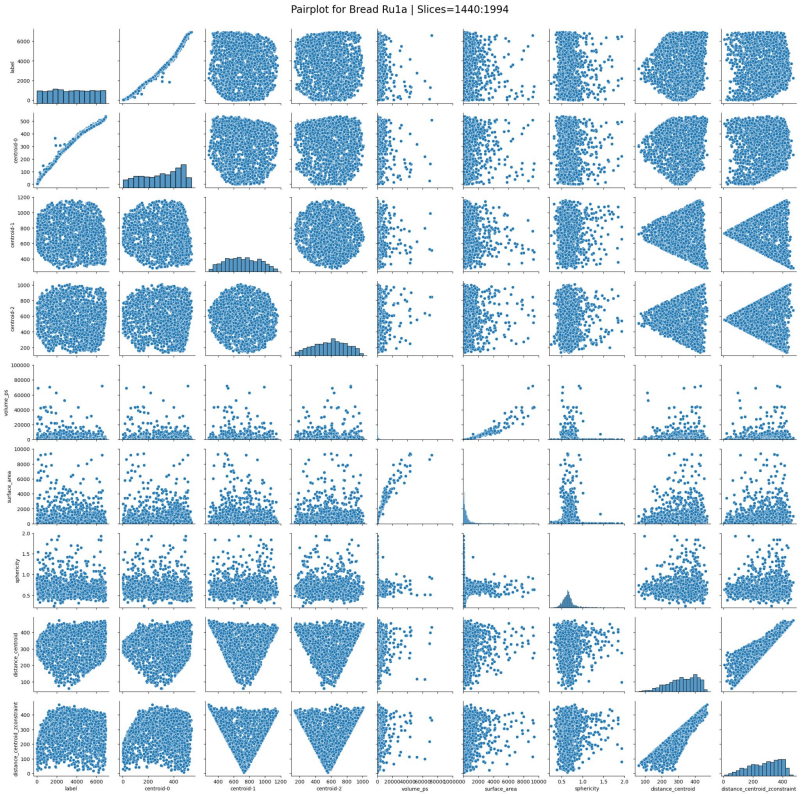


PAIRPLOTS

Ro1a (150:799 & 1420:2069)

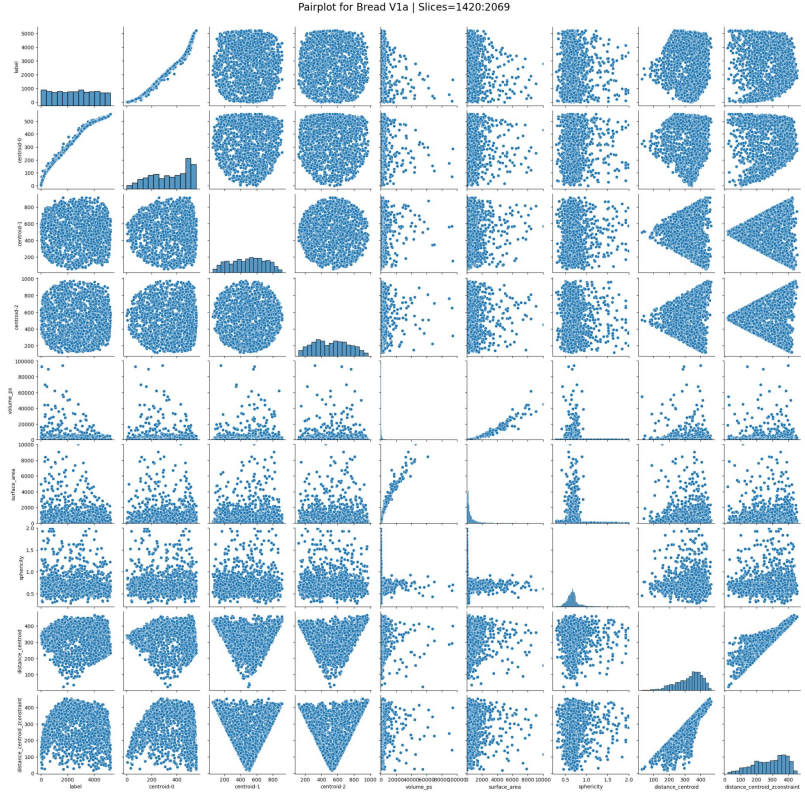
Ru1a (299:853 & 1440:1994)



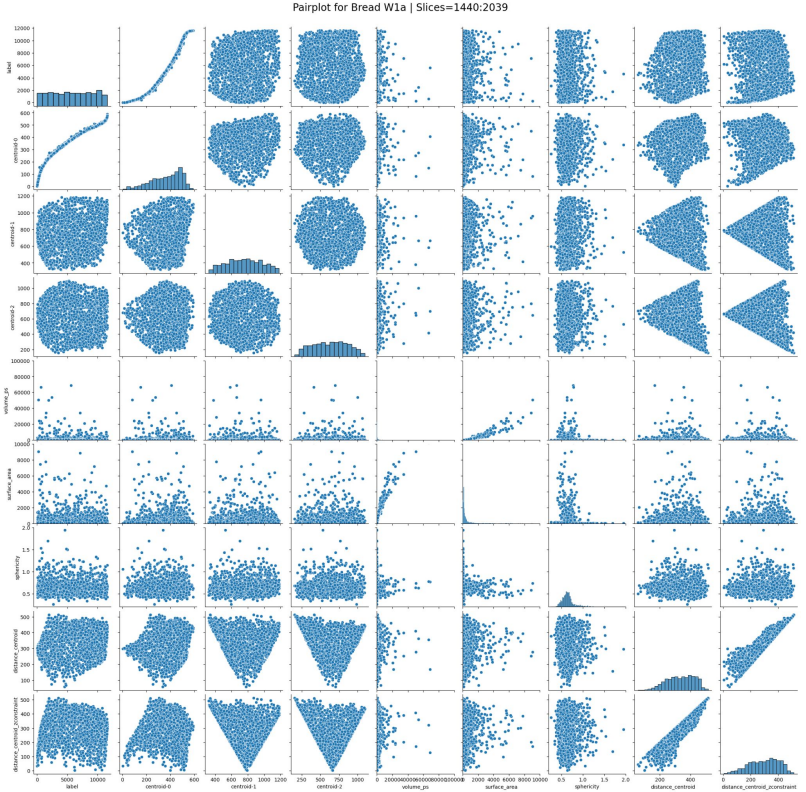


PAIRPLOTS

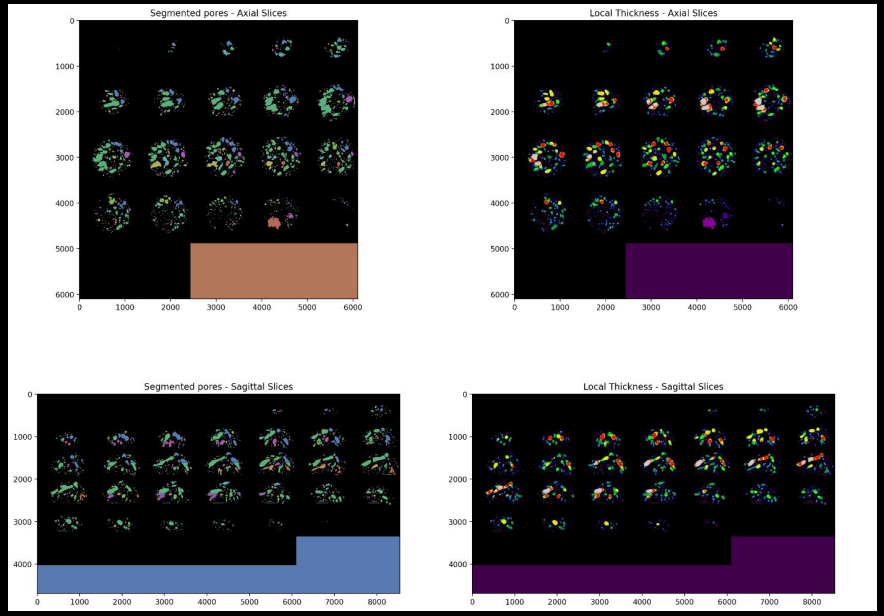
V1a (180:829 & 1420:2069)



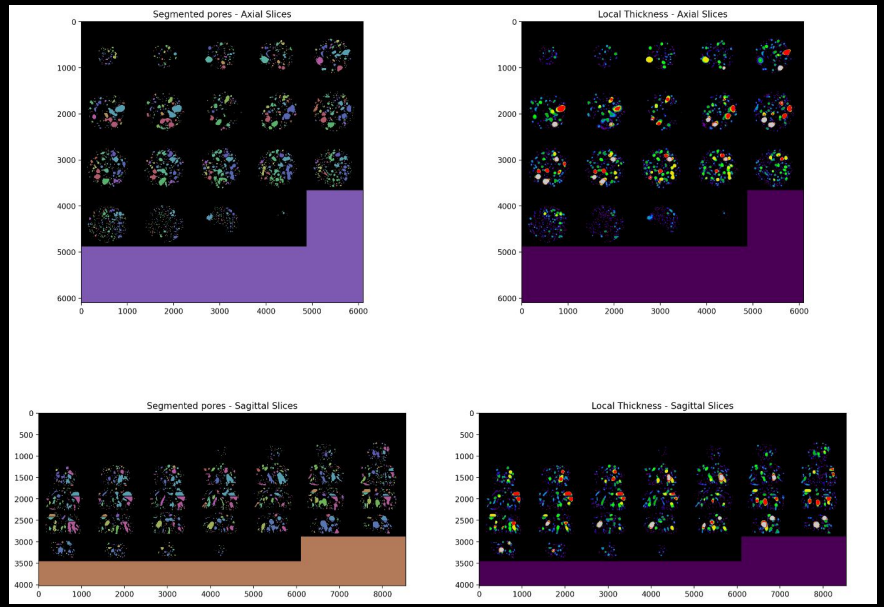
W1a (300:899 & 1440:2039)



LOCAL THICKNESS MAPS

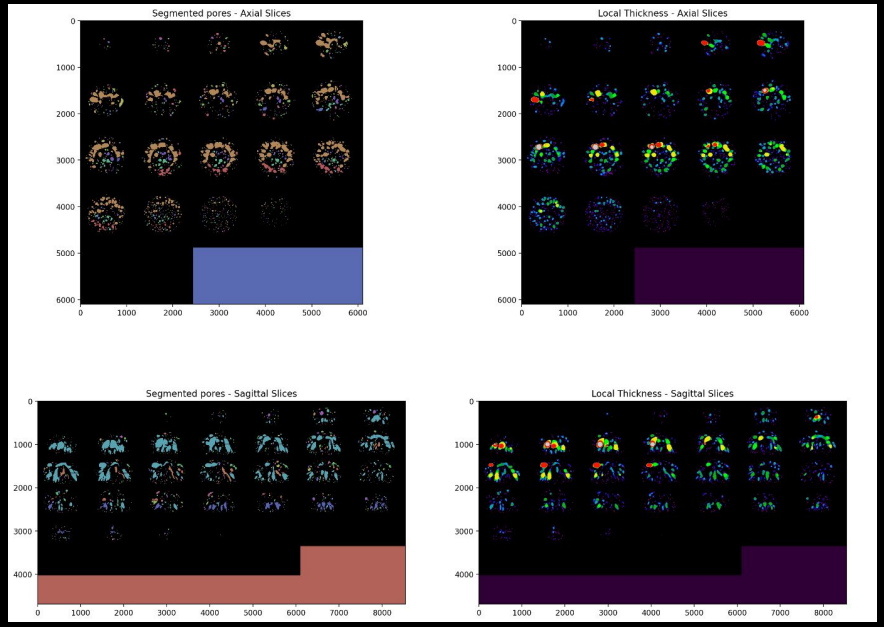


Ro1a (1420:2069)

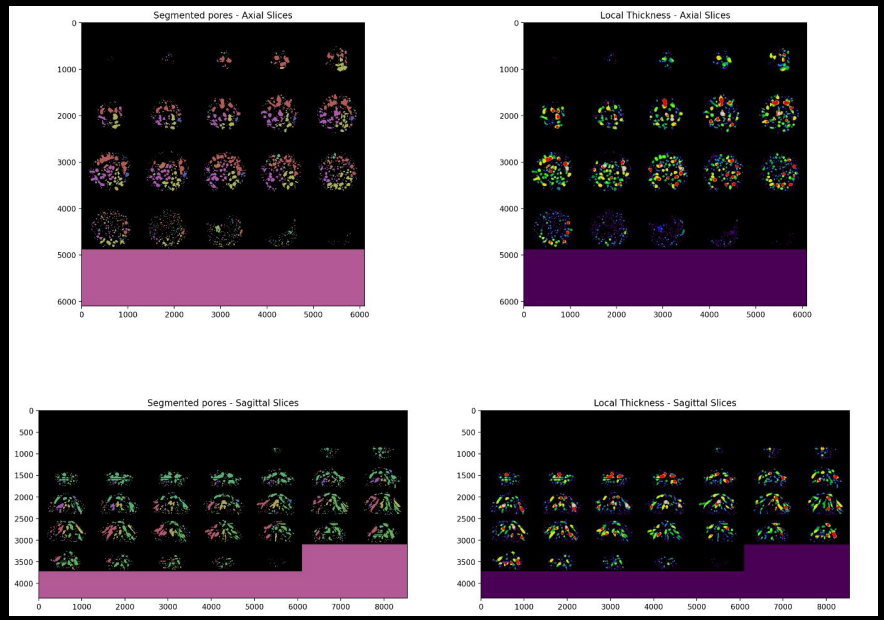


Ru1a (1440:1994)

LOCAL THICKNESS MAPS



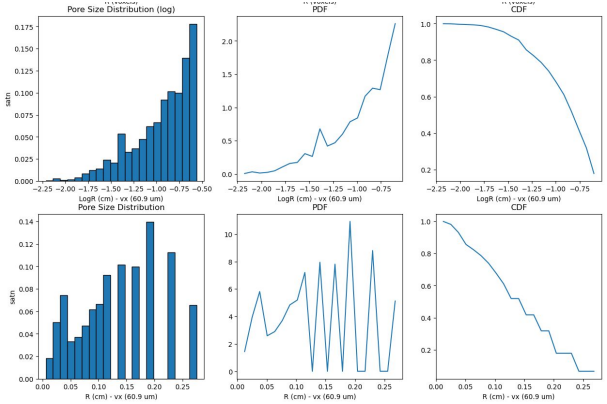
V1a (1420:2069)



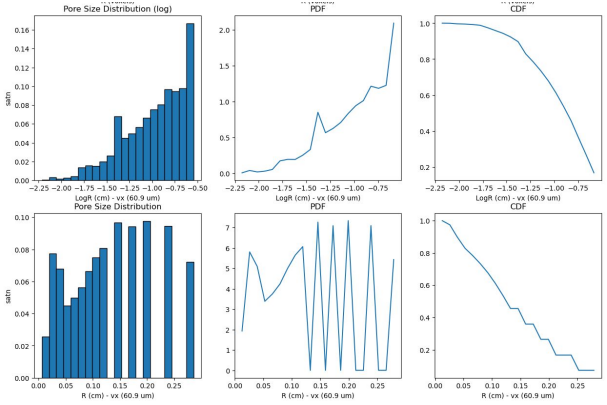
W1a (1440:2039)

PORE SIZE DISTRIBUTION

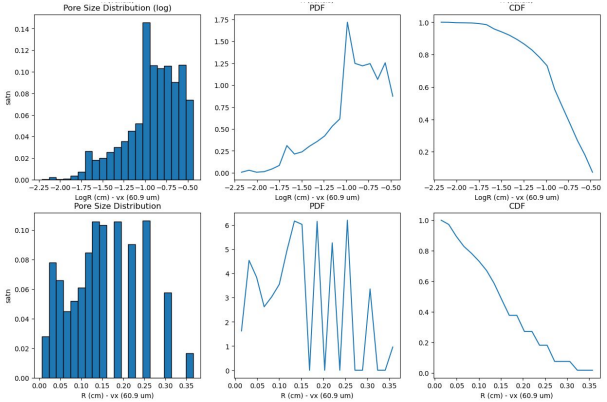
Ro1a (1420:2069)



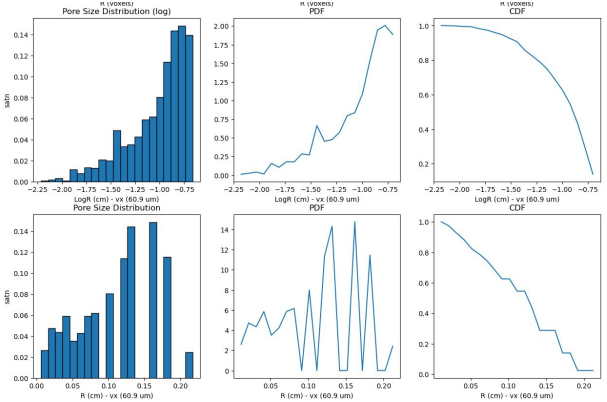
Ru1a (1440:1994)



V1a (1420:2069)

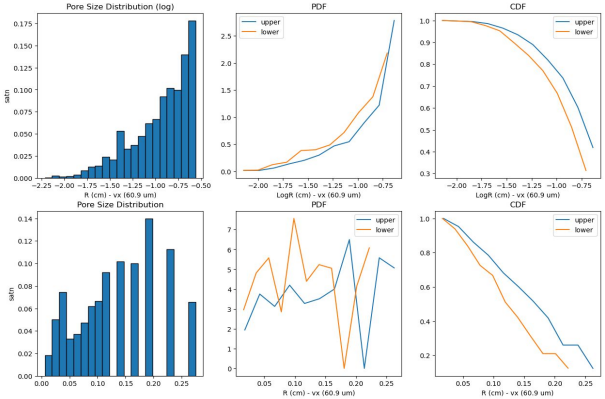


W1a (1440:2039)

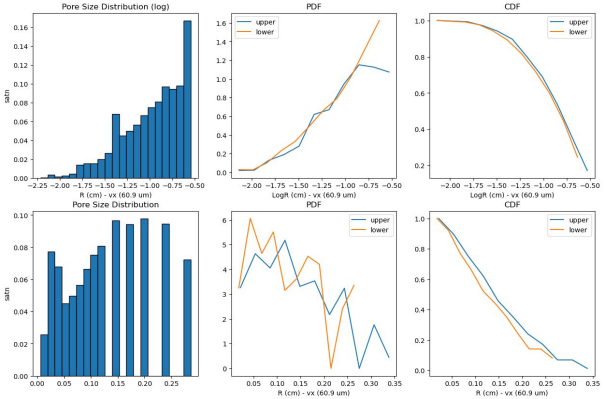


PORE SIZE DISTRIBUTION

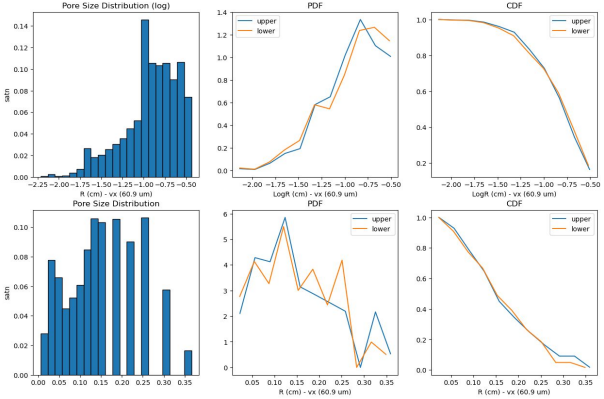
Ro1a (1420:2069)



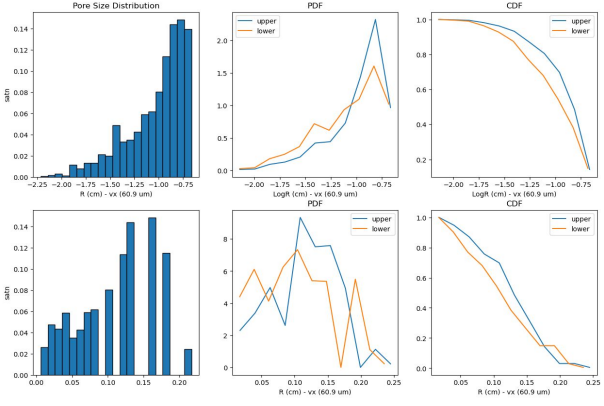
Ru1a (1440:1994)



V1a (1420:2069)

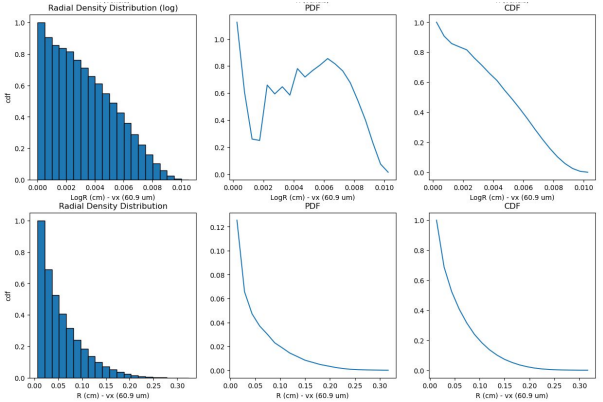


W1a (1440:2039)

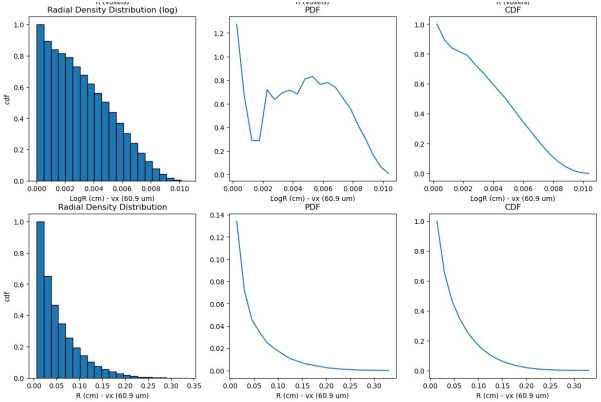


RADIAL DENSITY DISTRIBUTION

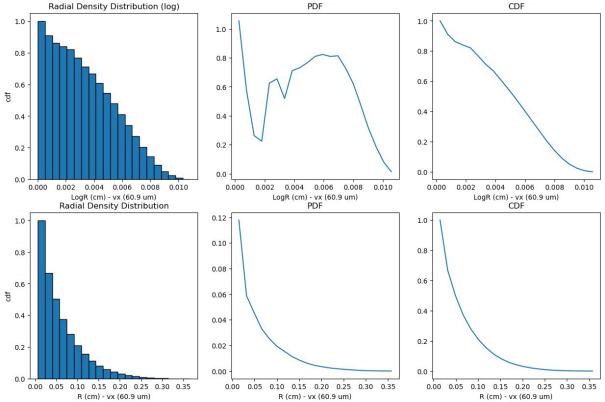
Ro1a (1420:2069)



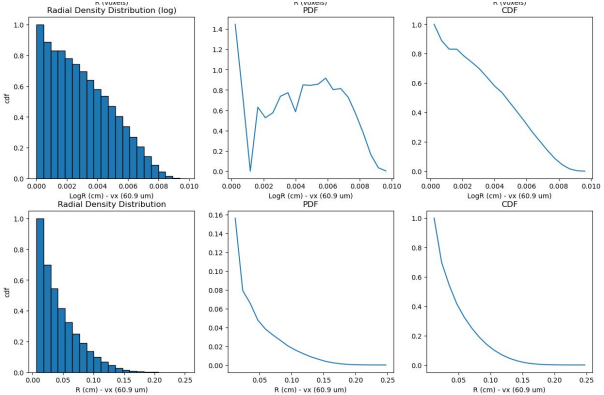
Ru1a (1440:1994)



V1a (1420:2069)

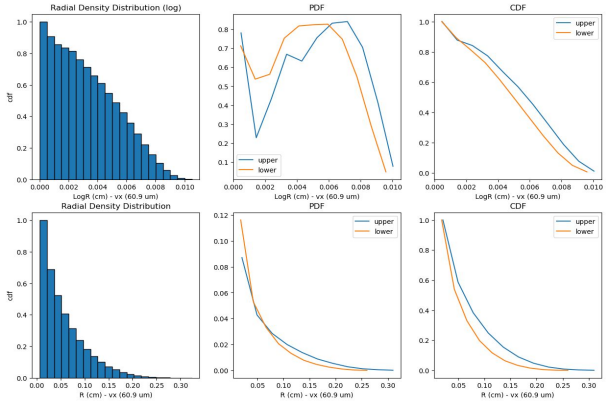


W1a (1440:2039)

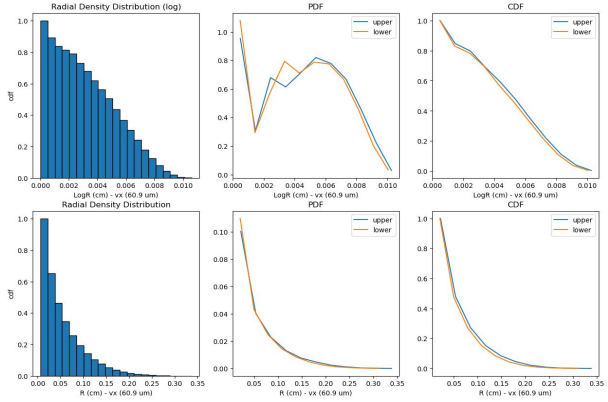


RADIAL DENSITY DISTRIBUTION

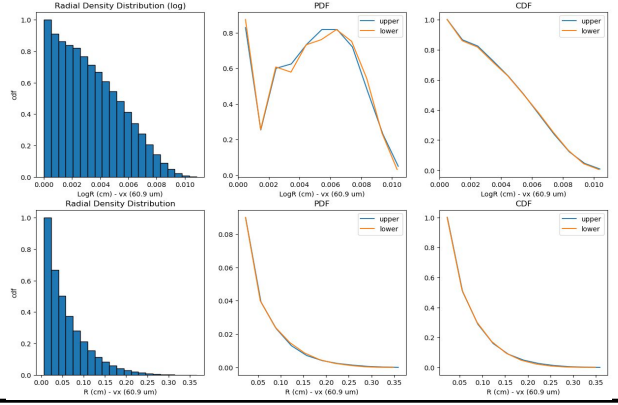
Ro1a (1420:2069)



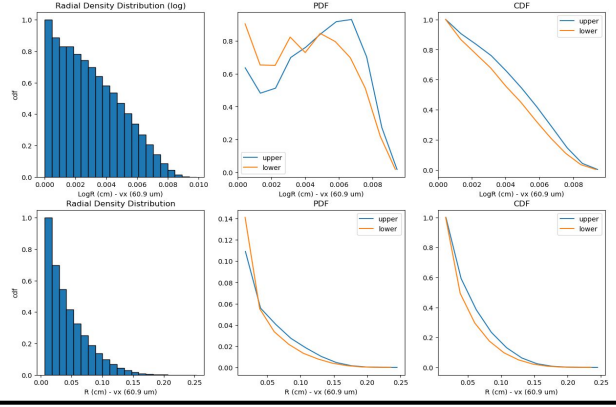
Ru1a (1440:1994)



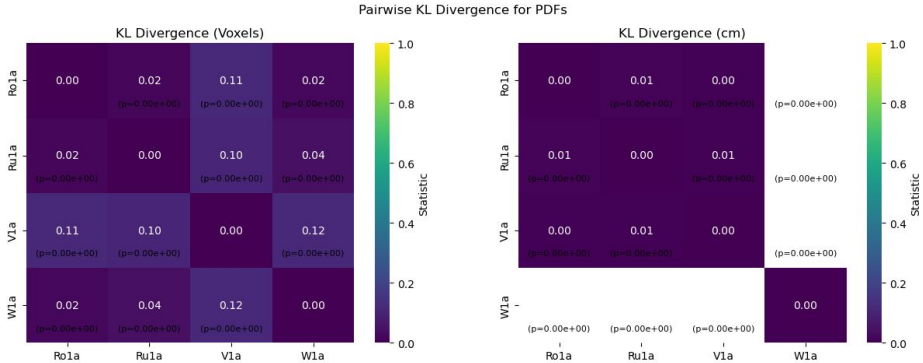
V1a (1420:2069)

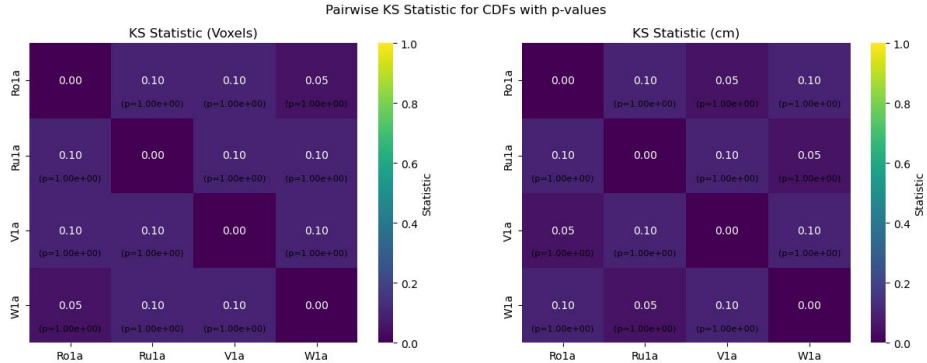


W1a (1440:2039)



COMPARING PS AND RD DISTRIBUTIONS





Comparing Probability Density Functions using KL Divergence -

1. Pore Size Distribution
2. Radial Density Distribution

Comparing Cumulative Density Functions using KS Test -

1. Pore Size Distribution
2. Radial Density Distribution

KMEANS CLUSTERING

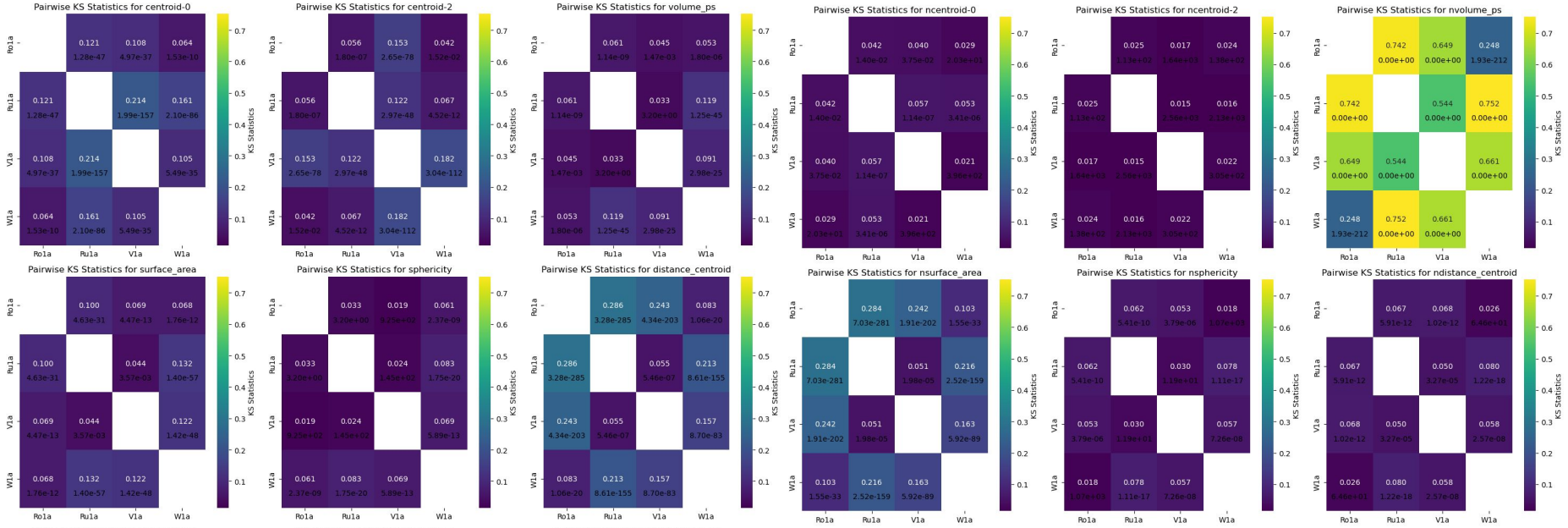
```
kmeans= KMeans(4)
#x=kmeans.fit_predict(test_df_pore[['nlabel','ncentroid-0','ncentroid-1','ncentroid-2','nvolume_ps','nsphericity']].values)
x=kmeans.fit_predict(test_df_pore[['nlabel','ncentroid-0','nvolume_ps','nsurface_area']].values)
test_df_pore['group']=x.tolist()
test_df_pore.sample(5,random_state=30)
```

```
Group 0: Ro1a=7009, Ru1a=5370, V1a=4877, W1a=9850
Group 1: Ro1a=19, Ru1a=31, V1a=16, W1a=21
Group 2: Ro1a=4495, Ru1a=3633, V1a=3107, W1a=6000
Group 3: Ro1a=6, Ru1a=5, V1a=5, W1a=7
```

- Low sample size?, Preprocessing and segmentation quality? Relevant metrics?
- Different enough distributions?

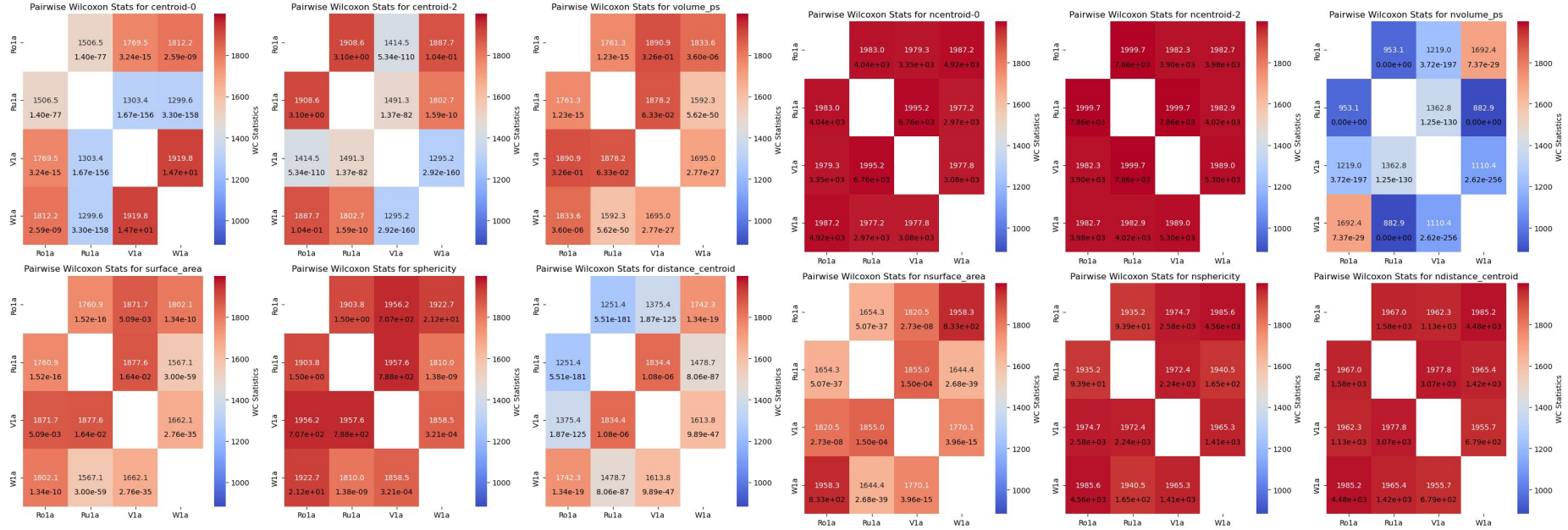
(Goncalves et al., 2022; Fan et al., 2016)

STATISTICAL TESTS - KOLMOGOROV-SMIRNOV



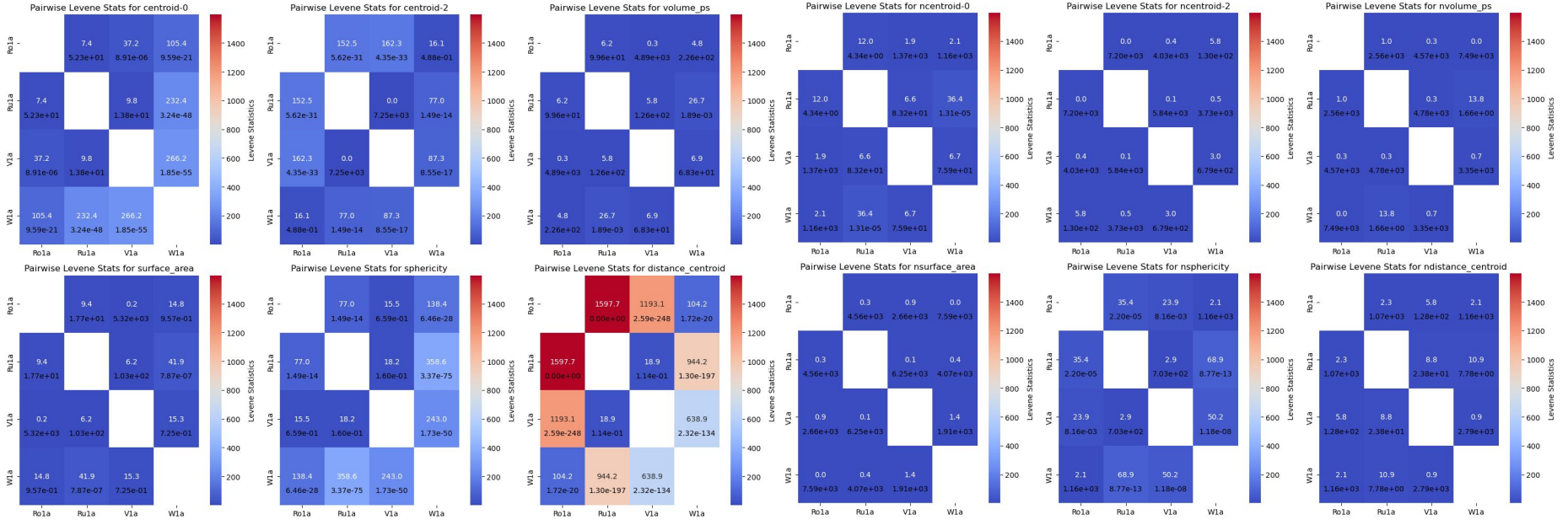
- Non-parametric test to compare 2 distributions

STATISTICAL TESTS - WILCOXON SIGNED-RANK



- Non-parametric test to compare the locations (medians) of two related samples

STATISTICAL TESTS - LEVENE'S



- Non-parametric test to compare the variances of two samples

LIMITATIONS

- Logistical
  - Memory and Processor Requirements
  - 2D Visualisations and Visual Inspections made
- Methodological
  - Absence of Ground Truth - Segmentation Quality
  - Lower sample sizes
  - Ad hoc constants - filter size window, threshold value, structures for pore extraction, size and distance filters
- Scientific
  - Wrong metrics chosen (probably)
  - Clustering approach - Feature rescaling and Representation
  - Heterogeneity and Material Anisotropy

CONCLUSIONS

There are trends in the data that show a distinction between sourdough starters can be made

(Normalised) pore volumes in the data arise from different distributions indicating a potential difference in pore spatial distributions across sourdough starters. Difference in median metrics for almost all normalised metrics

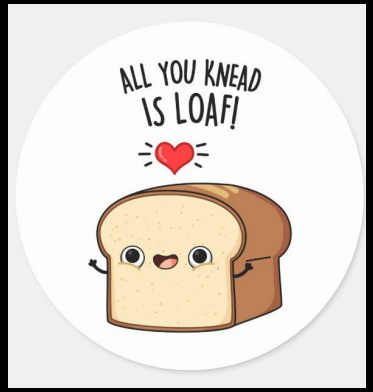
Larger pores are more concentrated towards the top of the bread sample - however probably not statistically significant due to dominance of smaller pores

We have a nice general pipeline for processing samples obtained from X-ray tomography that can serve as a baseline for future developments

FUTURE PROSPECTS

- Logistical
  - Memory and Processor Requirements - Run in Clusters (HPCs), Dask, etc
  - Higher Sample Size
  - Fully automated standardised pipelines for industrial products
- Methodological
  - Creation of Ground Truth - Assessing Segmentation Quality and Development
  - Use of ImageNets, Unets (CNNs), Autoencoders, and Spatial Clustering approaches
  - Controlled Baking environment and approach
- Scientific
  - Multimodal imaging approaches to same problem
  - Assessing Heterogeneity and Material Anisotropy - Mechanical Properties
  - Scalable Applications to Food Industry and Other products

THANK YOU



QUESTIONS AND FEEDBACK WOULD BE MUCH APPRECIATED

https://github.com/Consilium5128/QBI_BreadAnalysis/tree/main

Saksham Pruthi | spruthi@student.ethz.ch