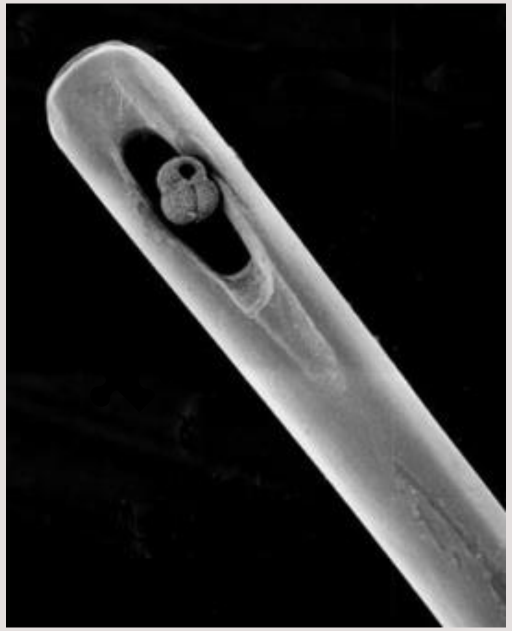
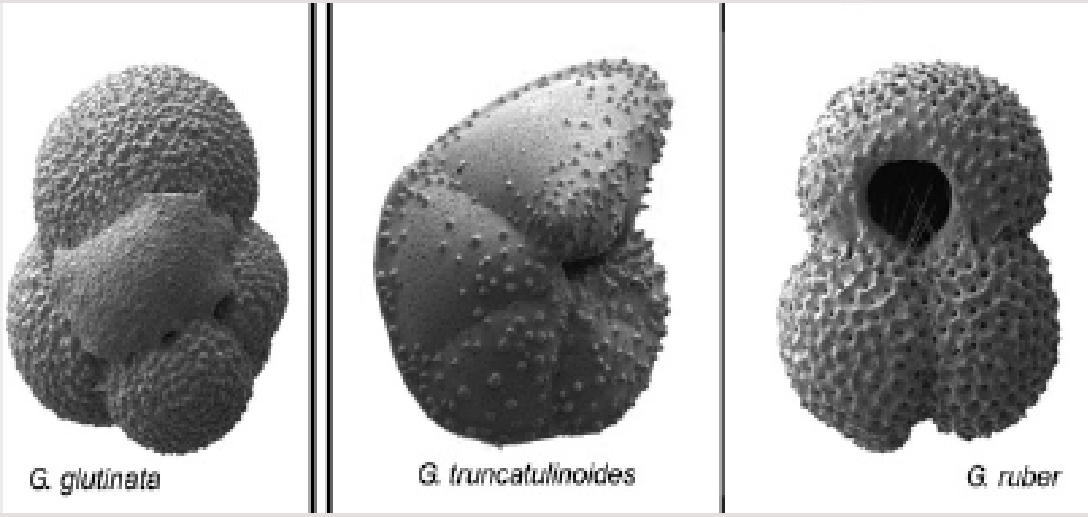





# Data Science Mini-Project Final Presentation

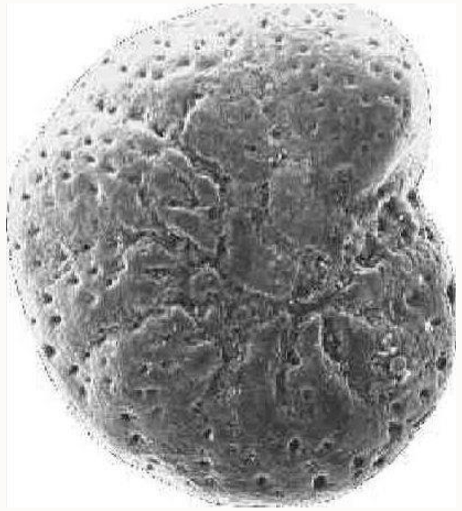
## Group M9 EMATM0050-2024

**Presented By:**

Aryan Lodha  Yitang Yuan
Somesh Raju  Puhan Wang

# Data Source

- All data files include both original and additional data, totaling 1,976,040 entries of 9-dimensional data.
- A worksheet file includes Age
- Environmental data from Cenogrid that contains the levels of Sigma 13 and Sigma 18 for each year.



# Question 1

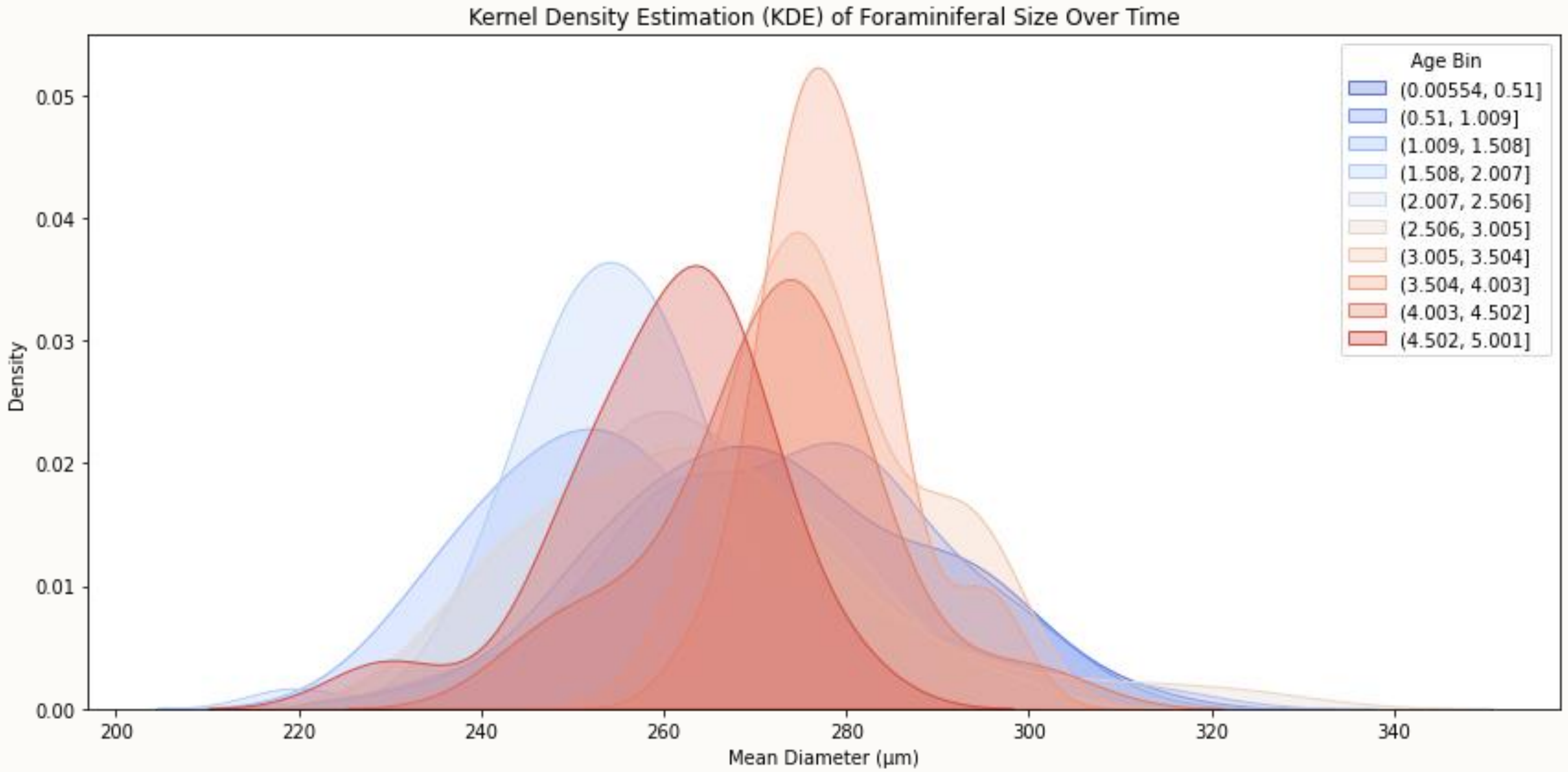
Is there a change in the shape of size distribution through time, not just its skewness, providing more insights into changes in the community structure?

## Data & Attributes

- 377 .csv morphometric files and Geological age assigned via 925 Mastersheet

## Methodology

1. **Data Integration and Cleaning**
   - Merged 377 .csv morphometric datasets with the Site 925 Mastersheet to assign accurate geological ages (Ma).
   - Standardized column names and resolved missing values for statistical continuity.
2. **Temporal Binning**
   - The time series was segmented into 5, 10, and 20 age bins to assess resolution sensitivity.
   - The 10-bin scheme was chosen for final analyses based on the best balance between interpretability and granularity.
3. **Shape Distribution Analysis**
   - Computed Skewness, Kurtosis, and Shannon Entropy for each bin to quantify distribution asymmetry, peakness, and complexity.
   - These metrics revealed evolutionary transitions and diversity shifts in community structure.



# Results & Analysis

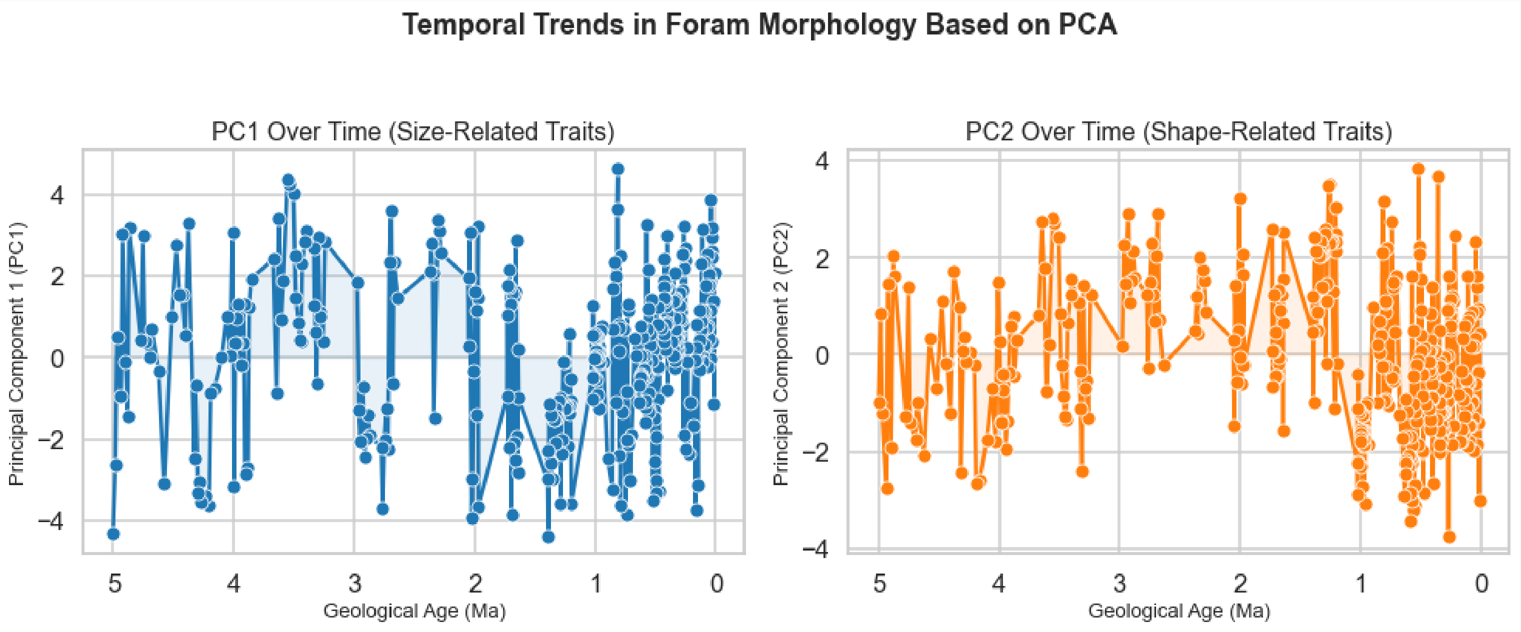
1. **Foraminiferal size distributions** show clear, non-random changes over time, reflecting ecological responses to climate variability.

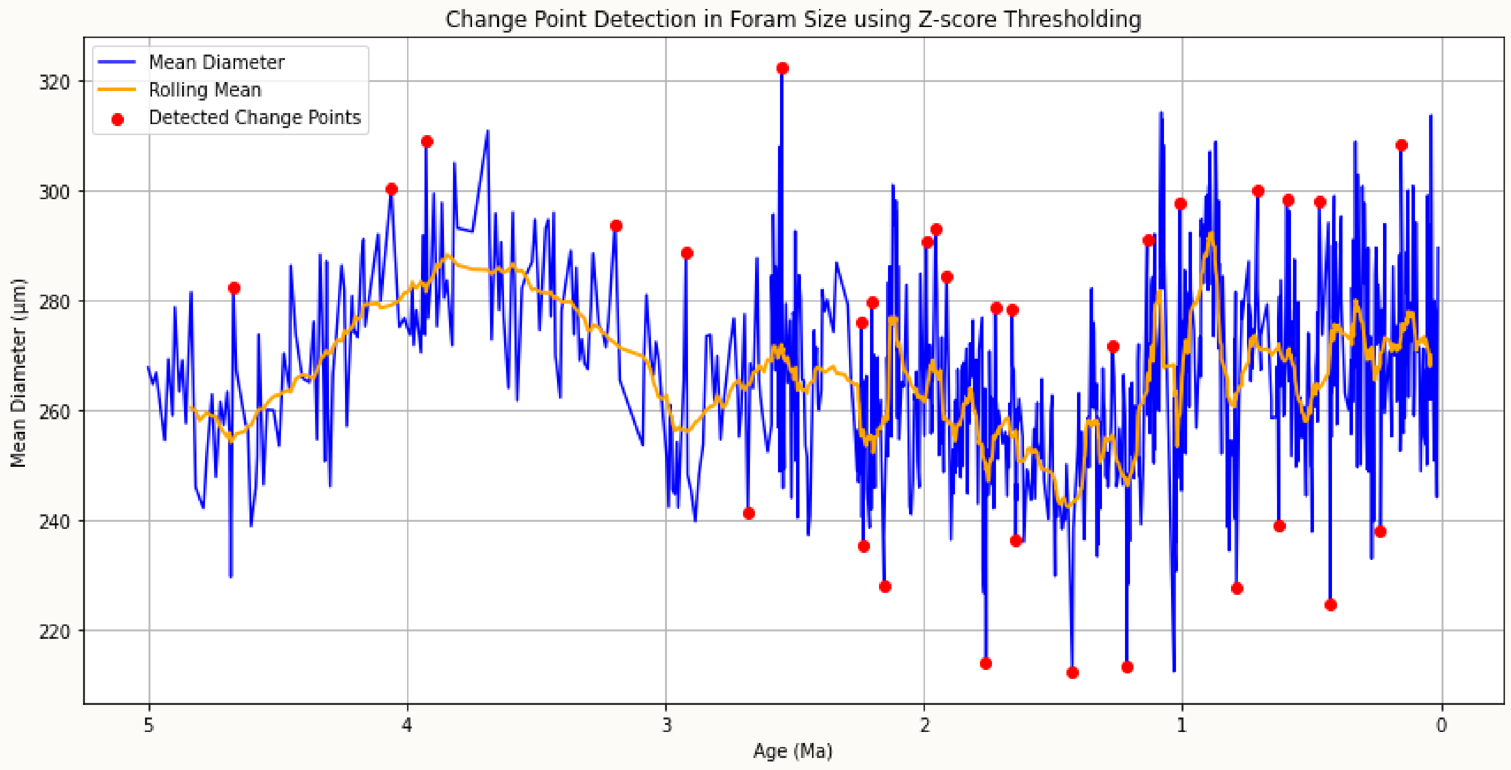
2. **KDE ridge plots** reveal:
   - **Flatter curves during warm periods** → indicative of ecological generalism.
   - **Sharper peaks during cold periods** → suggest morphotype dominance.
   - Shifts reflect changes in **kurtosis**, **entropy**, and **multimodality**, not just mean or skewness.

3. **PCA-based morphospace analysis** uncovers:
   - Temporal clustering and divergence in size and shape traits.
   - Emergence of **distinct morphological regimes**, indicating ecological restructuring.

4. **Change point detection (PELT)** identifies major shifts at:
   - **~3.2 Ma** and **~1.6 Ma**, aligning with key glacial onsets.
   - Signals **abrupt ecological reorganizations** linked to environmental forcing.

Overall, findings highlight a **strong connection between climate change and community structural transformations** in foraminifera over geologic time.



# Question 2

**Would a more sophisticated analysis allow determining changes in number of species through time?**

## Data & Attribute

- The Ceara Rise region & the 925 supplementary dataset(1,976,040 data in total)
- Mean diameter, maximum diameter, elongation, sphericity, area, grayscale intensity

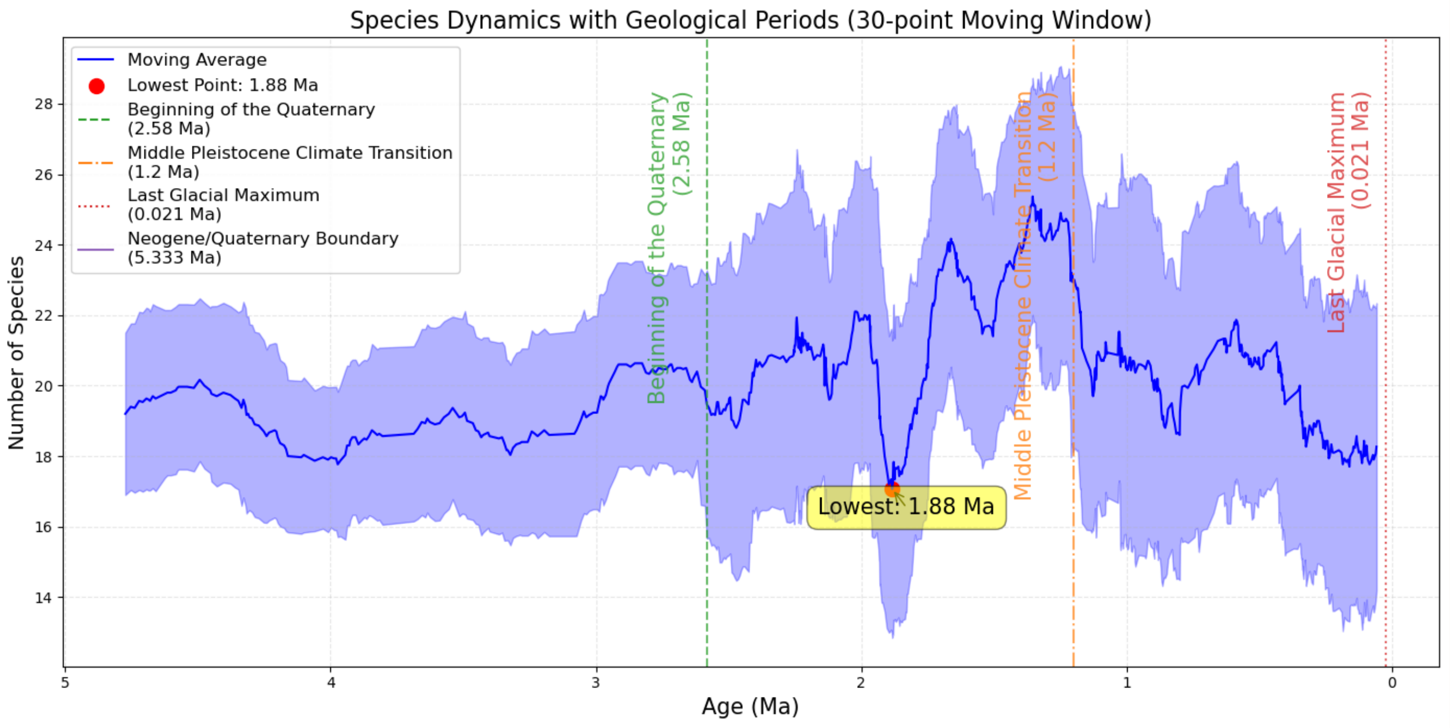
## Methodology

Gaussian mixture model

- Applied Gaussian Mixture Model (GMM) to determine the optimal number of clusters for each file.

BIC Criterion

- Evaluated models using the Bayesian Information Criterion (BIC), iterating through 2–30 clusters and selecting the number with the minimum BIC value.




| Period | Time(Ma) | Mean | Std | Points | T-test |
|---|---|---|---|---|---|
| Late Pliocene | 5.0-2.58 | 19.19 | 2.69 | 170 | -4.39 (0.00) |
| Early Pleistocene | 2.58-1.2 | 21.54 | 4.69 | 319 | 5.49 (0.00) |
| Middle Pleistocene | 1.2-0.8 | 20.41 | 4.80 | 103 | 0.67 (0.50) |
| Late Pleistocene | 0.8-0.01 | 19.41 | 4.46 | 176 | -2.04 (0.04) |

# Result & Analysis

Early Pleistocene (2.58–1.2 Ma)
- Significant species increase (mean = 21.5).
- Likely driven by accelerated ecological niche differentiation during early glacial-interglacial transitions.

Middle Pleistocene (1.2–0.8 Ma)
- Species decline (mean = 20.4).
- Potential link to extreme climate fluctuations during the Mid-Pleistocene Transition (MPT), causing local extinctions.

Species Distribution Patterns
- High diversity (≥28 species): Concentrated in 1.5–2.5 Ma (Early Pleistocene), aligned with pre-glacial warm phases favoring diversification.
- Low diversity (≤12 species): Scattered in 0.1–0.5 Ma (Late Pleistocene), possibly tied to regional extinctions from extreme cold events (e.g., Last Glacial Maximum, Heinrich events).

| Period | Time(Ma) | Mean | Std | Points | T-test |
|---|---|---|---|---|---|
| Late Pliocene | 5.0-2.58 | 19.19 | 2.69 | 170 | -4.39 (0.00) |
| Early Pleistocene | 2.58-1.2 | 21.54 | 4.69 | 319 | 5.49 (0.00) |
| Middle Pleistocene | 1.2-0.8 | 20.41 | 4.80 | 103 | 0.67 (0.50) |
| Late Pleistocene | 0.8-0.01 | 19.41 | 4.46 | 176 | -2.04 (0.04) |

# Question 3

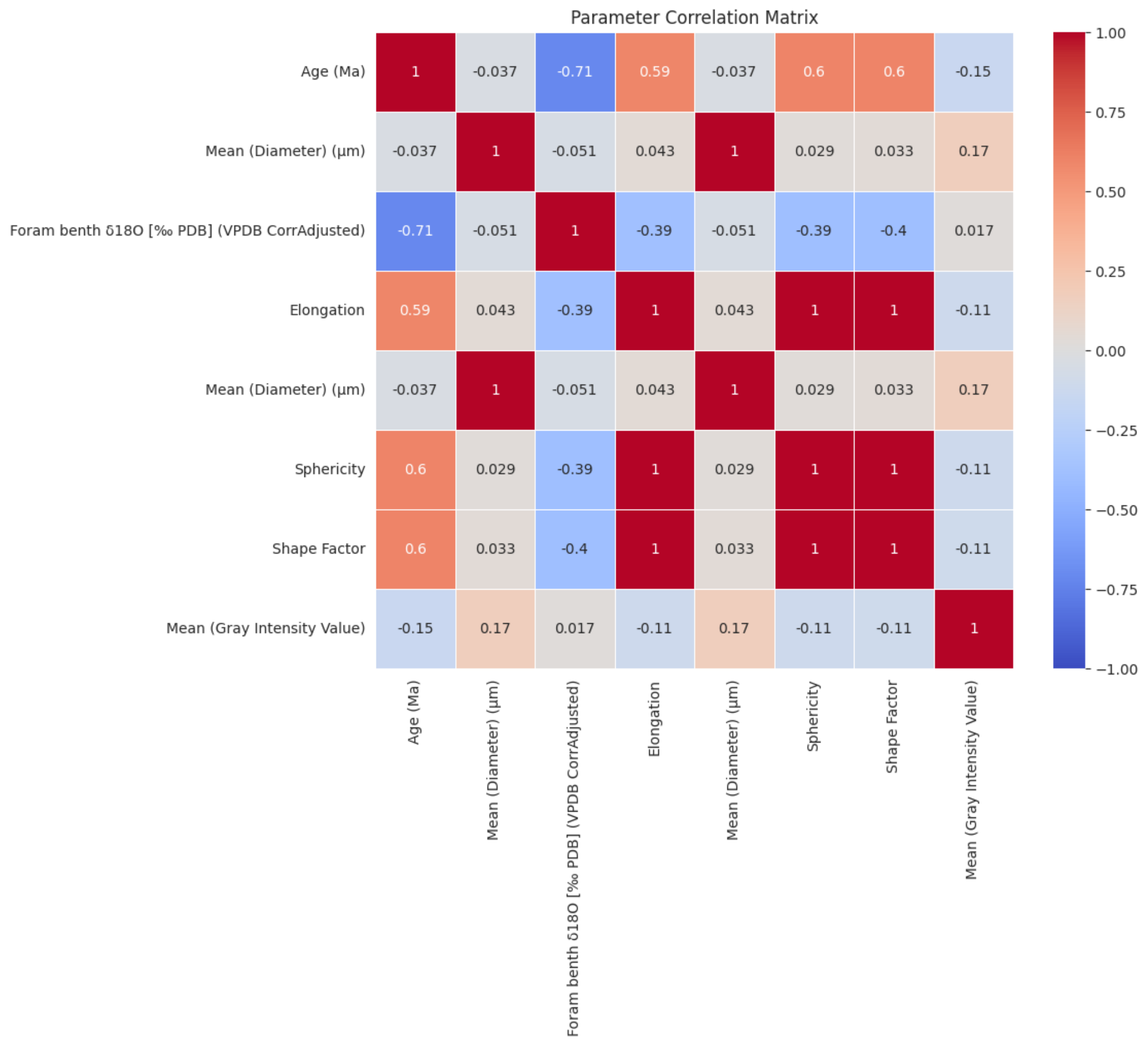
**Are the patterns of changes through time different for size than for other parameters indicating different selection pressure?**

## Data & Attribute

- The Ceara Rise dataset includes morphological measurements of fossilized foraminifera from deep-sea sediment cores, recorded across geological time.
- Key features extracted from dataset include Mean Diameter, Elongation, Sphericity, Shape Factor, Gray Intensity, Area, Perimeter, and $\delta^{18}O$ (paleotemperature proxy), all linked to specific ages.
- Data preprocessing involved removing missing values, sorting by geological age, and applying z-score normalization for consistent comparison across traits.

## Methodology

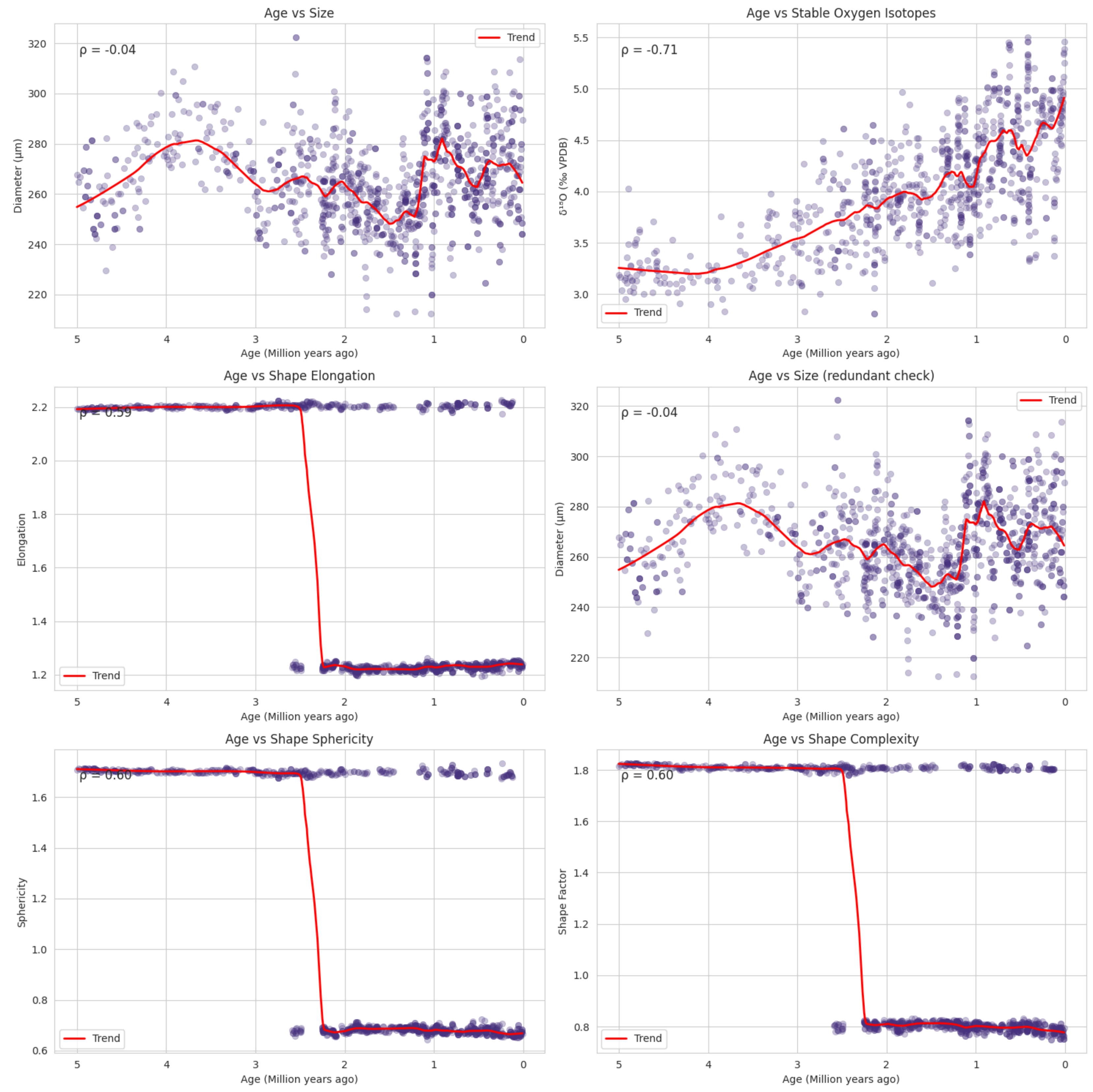
- Time-series plots were created for all traits, with time reversed (older to newer) to match geological conventions.
- Size was tracked using Mean and Max Diameter; shape with Elongation, Sphericity, Shape Factor; and intensity via Gray Intensity.
- These plots showed whether traits changed together (shared drivers) or followed different paths (separate pressures).
- A Pearson correlation matrix assessed links between size, shape, and intensity traits.
- Weak correlations suggest size and shape evolved independently, shaped by different forces.
- Comparing Mean Diameter and $\delta^{18}O$ showed partial climatic influence on size, but not on shape.




| Morphological Parameters | Correlation |
|---|---|
| Shape Factor | 0.6671 |
| Sphericity | 0.6653 |
| Elongation | 0.6570 |
| Mean (Diameter) (µm) | -0.0773 |
| Mean (Gray Intensity Value) | -0.1310 |
| Perimeter (µm) | -0.1455 |

# Result & Analysis

- **Size traits** (Mean/Max Diameter) show weak, irregular changes over time with little correlation to age (ρ ≈ -0.04), highlighting their fluctuating and environmentally sensitive nature.
- **Shape traits** (Elongation, Sphericity, Shape Factor) undergo abrupt, synchronized shifts around **3 million years ago**, indicating structurally stable yet episodically shifting features under different evolutionary control.
- **$\delta^{18}$O values** correlate more strongly with size (ρ ≈ -0.71), suggesting that size responds more directly to long-term climatic trends, while shape appears decoupled from these environmental signals.
- Indeed, the patterns of change over time are different: **"size fluctuates irregularly with little correlation to time (ρ = -0.04)"**, whereas **"shape parameters (elongation, sphericity, and complexity) show abrupt and synchronized shifts around 3 Ma"**, indicating **"distinct selective pressures"** acting separately on size and shape traits.



# Question 4

**Is there are periodicity in the data and is this different for different parameters?**

## 1 Data & Attribute

The attributes include metrics such as:

- **Size measurements**: Minimum, maximum, and mean diameters (µm)
- **Shape descriptors**: Elongation, sphericity, shape factor
- **Optical properties**: Mean gray intensity
- **Geometric properties**: Area ($\mu m^2$) and perimeter (µm)

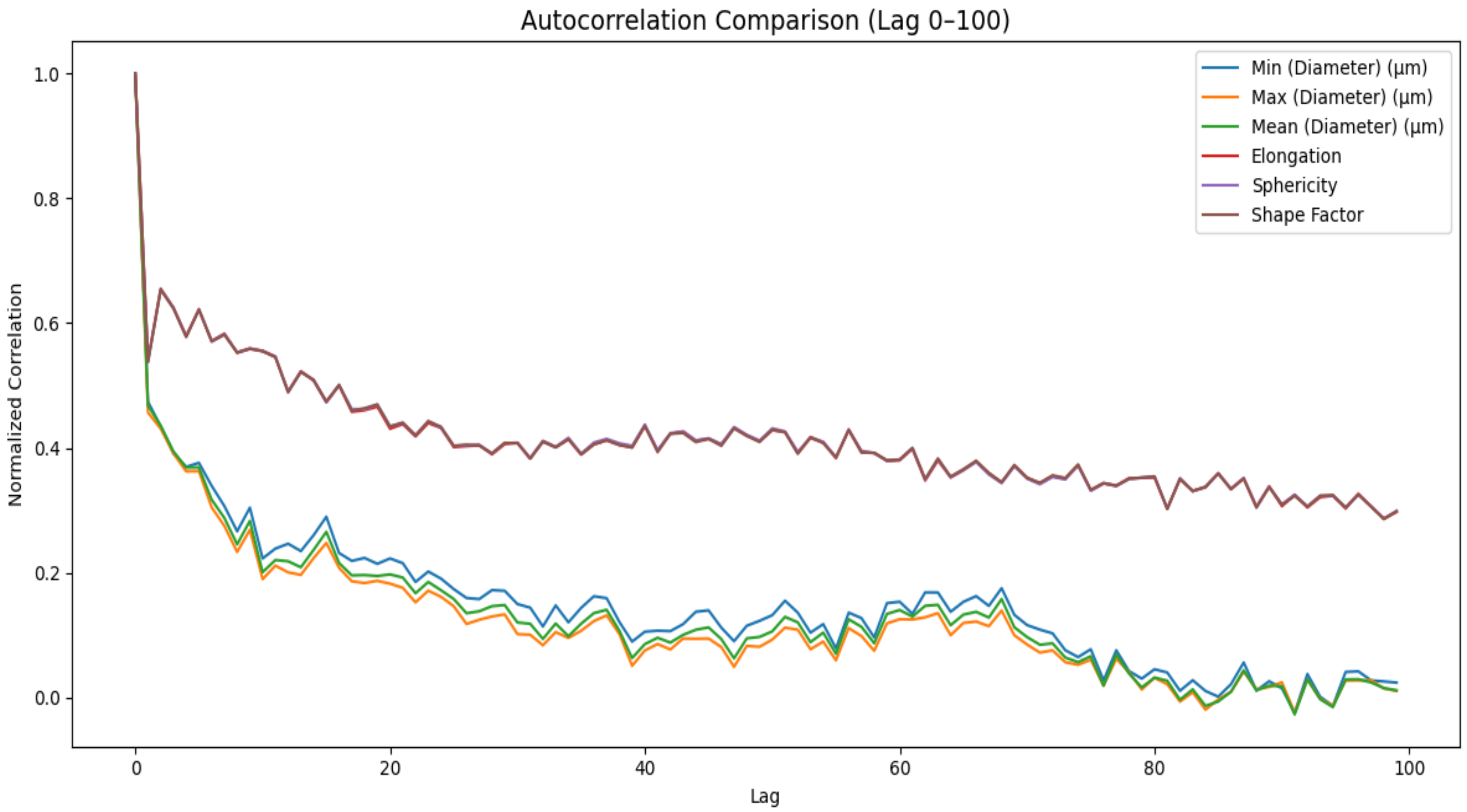
These parameters reflect physical and compositional characteristics of particles across geological time, allowing for potential detection of periodic trends linked to environmental or sedimentary cycles.

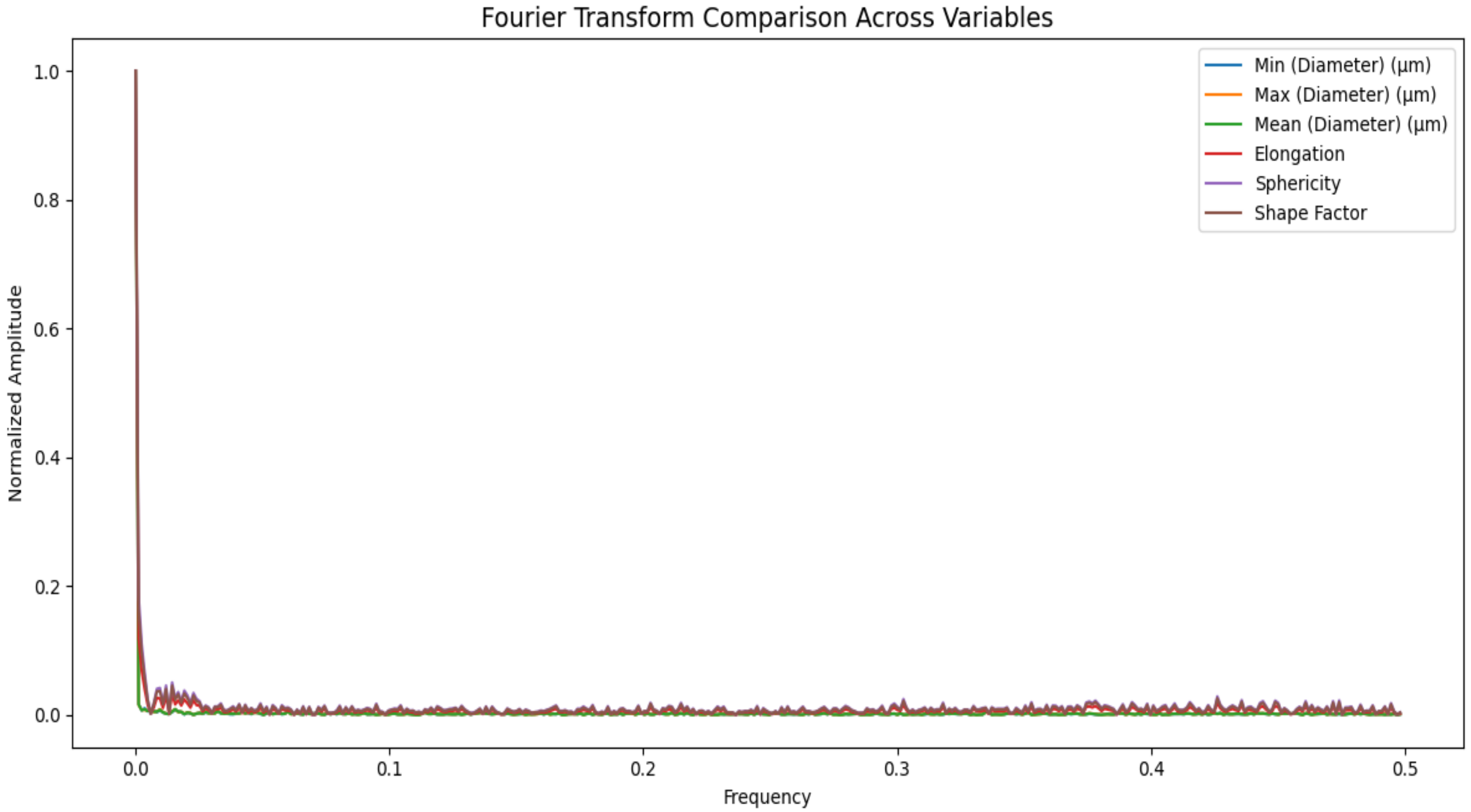
## 2 Methodology

The Fourier and Wavelet transforms are used when dealing with simpler an

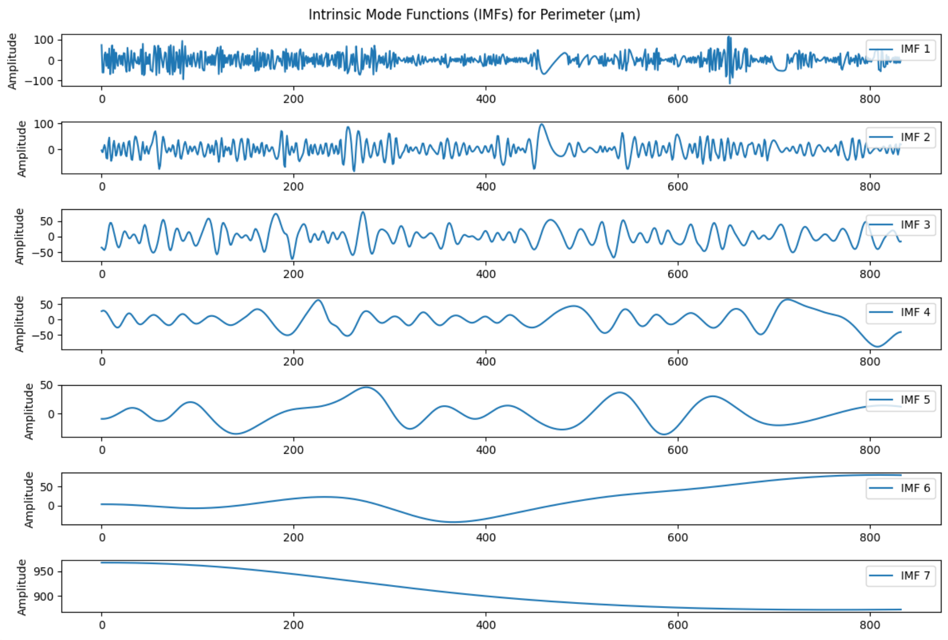
## 3 Mid-term revisions and the problems faced

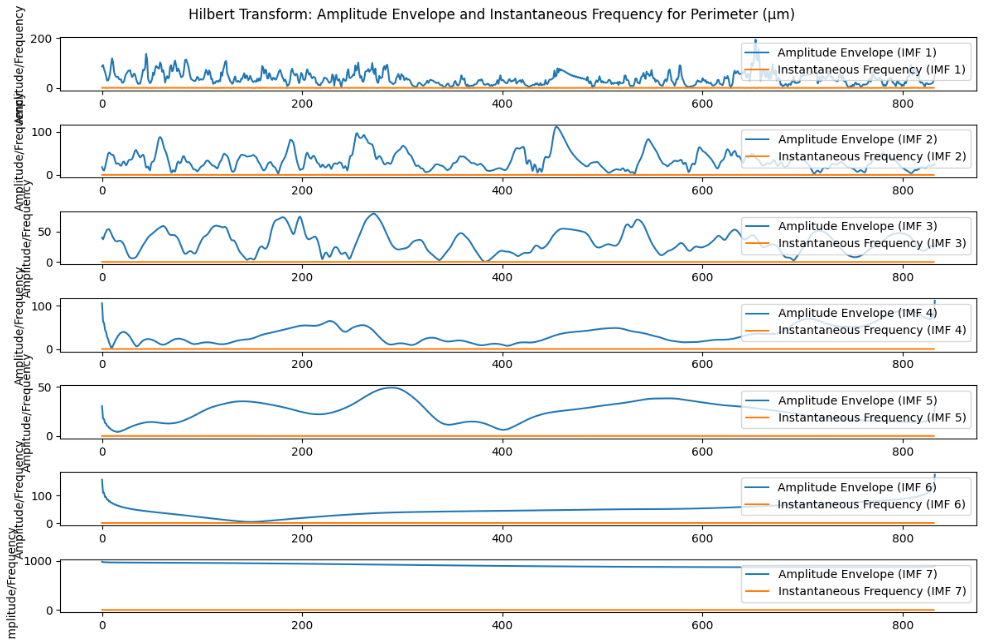
Fourier transform is not suitable for processing complex data and cannot handle data very well



# Whether the data is periodic and whether the periodicity is the same?




## Result & Analysis

Through our analysis of the images, we found that there are three groups of data with the same periodicity, namely: Min Diameter, Max Diameter, and Mean Diameter are the first group; The Shape Factor and Mean (Gray Intensity) are the second group,δ 13C and δ18O are the third group, while the other data either have no periodicity or have different periodicity and are relatively independent.

| Data Name | IMF 1~3 | IMF 4~6/7 | IMF 7/8/9 |
|---|---|---|---|
| Area | No | Weak/Quasi-periodic | No |
| Perimeter | No | Weak/Quasi-periodic | No |
| d13C_VPDB | No | Weak/Quasi-periodic | No |
| d18O_VPDB | No | Weak/Quasi-periodic | No |
| Min Diameter | No | Weak/Quasi-periodic | No |
| Max Diameter | No | Weak/Quasi-periodic | No |
| Mean Diameter | No | Weak/Quasi-periodic | No |
| Elongation | No | No periodicity | No |
| Sphericity | No | No periodicity | No |
| Mean (Gray Intensity) | No | Weak/Quasi-periodic | No |
| Shape Factor | No | Weak/Quasi-periodic | No |

# Question 5

**Would a more sophisticated analysis allow determining changes in number of species through time?**

## Data & Attribute

- Mean value of each age(769 data in total)

## Methodology

- Mixed-effects model
- Calculate the correlation with a lag, compute the maximum correlation from 10 periods ahead to 10 periods behind.

| Ind. | Lag (Periods) | Coef. r |
|---|---|---|
| Shape Factor vs δ 13C | 2 | +0.19 |
| Shape Factor vs δ 18O | 3 | -0.36 |
| Max Diameter vs δ 13C | 8 | -0.11 |
| Max Diameter vs δ 18O | 8 | +0.11 |
| Mean Diameter vs δ 13C | 8 | -0.09 |
| Mean Diameter vs δ 18O | 8 | +0.05 |
| Elongation vs δ 13C | 5 | -0.24 |
| Elongation vs δ 18O | 5 | +0.51 |
| Sphericity vs δ 13C | 4 | +0.24 |
| Sphericity vs δ 18O | 5 | -0.49 |

| Morphological Indicator | δ 13C (coef) | p | δ 18O (coef) | p |
|---|---|---|---|---|
| Shape Factor | -0.115 | 0.008 | -0.416 | <0.001 |
| Max Diameter | -0.057 | 0.223 | +0.013 | 0.779 |
| Mean Diameter | -0.080 | 0.087 | -0.058 | 0.208 |
| Elongation | +0.167 | <0.001 | +0.537 | <0.001 |
| Sphericity | -0.171 | <0.001 | -0.498 | <0.001 |

# Result & Analysis

**Key Findings**

- **Significant Correlations**
  - **Shape Factor**: Strong negative correlations with $\delta^{13}C$ (coef = -0.115, p = 0.008) and $\delta^{18}O$ (coef = -0.416, p < 0.001).
  - **Elongation**: Strong positive correlations with both isotopes ($\delta^{13}C$: +0.167, p < 0.001; $\delta^{18}O$: +0.537, p < 0.001).
  - **Sphericity**: Negative correlations ($\delta^{13}C$: -0.171; $\delta^{18}O$: -0.498, p < 0.001).
  - **Size Metrics** (Max/Mean Diameter): No significant associations (p > 0.05).
- **Environmental Drivers**
  - **Shape plasticity** (flattening, elongation, reduced sphericity) is driven by environmental shifts, not size changes.
  - $\delta^{13}C$ (paleoproductivity/carbon cycling) and $\delta^{18}O$ (paleotemperature/ice volume) alter water chemistry/physics, triggering adaptive morphological adjustments.

**Lagged Responses & Environmental Dynamics**

- **Cross-Correlation Analysis**
  - Morphological responses peak after **2–8 periods** (≈10,000–40,000 years), indicating delayed adaptation.
  - **Strongest lags**:
    - Shape Factor vs. $\delta^{13}C$: **2 periods** (r = +0.19); vs. $\delta^{18}O$: **3 periods** (r = -0.36).
    - Elongation vs. $\delta^{18}O$: **5 periods** (r = +0.51).
    - Sphericity vs. $\delta^{18}O$: **5 periods** (r = -0.49).
- **Isotopic Influence Hierarchy**
  - $\delta^{18}O$ (temperature/ice volume) has a **stronger impact** on morphology ($|r| \approx 0.5$) than $\delta^{13}C$.

| Morphological Indicator | δ 13C (coef) | p | δ 18O (coef) | p |
|---|---|---|---|---|
| Shape Factor | -0.115 | 0.008 | -0.416 | <0.001 |
| Max Diameter | -0.057 | 0.223 | +0.013 | 0.779 |
| Mean Diameter | -0.080 | 0.087 | -0.058 | 0.208 |
| Elongation | +0.167 | <0.001 | +0.537 | <0.001 |
| Sphericity | -0.171 | <0.001 | -0.498 | <0.001 |

| Ind. | Lag (Periods) | Coef. r |
|---|---|---|
| Shape Factor vs δ 13C | 2 | +0.19 |
| Shape Factor vs δ 18O | 3 | -0.36 |
| Max Diameter vs δ 13C | 8 | -0.11 |
| Max Diameter vs δ 18O | 8 | +0.11 |
| Mean Diameter vs δ 13C | 8 | -0.09 |
| Mean Diameter vs δ 18O | 8 | +0.05 |
| Elongation vs δ 13C | 5 | -0.24 |
| Elongation vs δ 18O | 5 | +0.51 |
| Sphericity vs δ 13C | 4 | +0.24 |
| Sphericity vs δ 18O | 5 | -0.49 |

# Thank You