**Pediastrum analysis**

import pandas as pd

from scipy.spatial.distance import braycurtis

from sklearn.decomposition import PCA

import numpy as np

import matplotlib.pyplot as plt

from scipy.ndimage import gaussian\_filter1d

# Load the dataset

pediastrum\_data = pd.read\_csv('Pediastrum.csv', delimiter=';')

# Extract species data for Bray-Curtis calculation

species\_data\_pediastrum = pediastrum\_data.iloc[:, 2:].astype(float)

# Calculate Bray-Curtis dissimilarity for each consecutive pair of samples

bray\_curtis\_dissimilarity\_pediastrum = [braycurtis(species\_data\_pediastrum.iloc[i], species\_data\_pediastrum.iloc[i+1]) for i in range(len(species\_data\_pediastrum)-1)]

# Insert the Bray-Curtis dissimilarity values into the dataframe

pediastrum\_data['Bray-Curtis Dissimilarity'] = np.nan

pediastrum\_data.loc[1:, 'Bray-Curtis Dissimilarity'] = bray\_curtis\_dissimilarity\_pediastrum

# Calculate the rate of change for Bray-Curtis Dissimilarity

pediastrum\_data['Rate of Change BC'] = pediastrum\_data['Bray-Curtis Dissimilarity'].diff()

# Smooth the Bray-Curtis Dissimilarity using a Gaussian filter

pediastrum\_data['Smoothed BC Dissimilarity'] = gaussian\_filter1d(pediastrum\_data['Bray-Curtis Dissimilarity'].fillna(0), sigma=2)

# Calculate baseline for Bray-Curtis (5th percentile)

baseline\_bc\_pediastrum = np.percentile(pediastrum\_data['Bray-Curtis Dissimilarity'].dropna(), 5)

# Function to calculate Sørensen Dissimilarity

def sorensen\_dissimilarity(a, b):

return 1 - 2 \* np.minimum(a, b).sum() / (a.sum() + b.sum())

# Calculate Sørensen Dissimilarity for each consecutive pair of samples

sorensen\_dissimilarity\_pediastrum = [sorensen\_dissimilarity(species\_data\_pediastrum.iloc[i], species\_data\_pediastrum.iloc[i+1]) for i in range(len(species\_data\_pediastrum)-1)]

# Insert the Sørensen dissimilarity values into the dataframe

pediastrum\_data['Sørensen Dissimilarity'] = np.nan

pediastrum\_data.loc[1:, 'Sørensen Dissimilarity'] = sorensen\_dissimilarity\_pediastrum

# Calculate the rate of change for Sørensen Dissimilarity

pediastrum\_data['Rate of Change Sørensen'] = pediastrum\_data['Sørensen Dissimilarity'].diff()

# Smooth the Sørensen Dissimilarity using a Gaussian filter

pediastrum\_data['Smoothed Sørensen Dissimilarity'] = gaussian\_filter1d(pediastrum\_data['Sørensen Dissimilarity'].fillna(0), sigma=2)

# Calculate baseline for Sørensen (5th percentile)

baseline\_sorensen\_pediastrum = np.percentile(pediastrum\_data['Sørensen Dissimilarity'].dropna(), 5)

# Perform PCoA (equivalent to PCA for dissimilarity matrix)

pca\_pediastrum = PCA(n\_components=2)

species\_pca\_pediastrum = pca\_pediastrum.fit\_transform(species\_data\_pediastrum.fillna(0))

# Extract the first principal component for trend analysis

pc1\_pediastrum = species\_pca\_pediastrum[:, 0]

# Plot the results

fig, axs = plt.subplots(2, 2, figsize=(15, 12))

# Bray-Curtis Dissimilarity Over Time

axs[0, 0].plot(pediastrum\_data['Age CE'], pediastrum\_data['Smoothed BC Dissimilarity'], color='orange', linewidth=2, label='Smoothed BC Dissimilarity')

axs[0, 0].scatter(pediastrum\_data['Age CE'], pediastrum\_data['Bray-Curtis Dissimilarity'], color='brown', s=30, label='Original Data')

axs[0, 0].axhline(y=baseline\_bc\_pediastrum, color='grey', linestyle='--', linewidth=1.5, label='Baseline (5th percentile)')

axs[0, 0].set\_title('Bray-Curtis Dissimilarity Over Time (Pediastrum)')

axs[0, 0].set\_xlabel('Age CE')

axs[0, 0].set\_ylabel('Bray-Curtis Dissimilarity')

axs[0, 0].legend(loc='upper left')

axs[0, 0].grid(True)

# Rate of Change in Bray-Curtis Dissimilarity Over Time

axs[0, 1].plot(pediastrum\_data['Age CE'], pediastrum\_data['Rate of Change BC'], color='gray', linewidth=2, label='Rate of Change')

axs[0, 1].axhline(y=0, color='grey', linestyle='--', linewidth=1.5, label='Baseline')

axs[0, 1].set\_title('Rate of Change in Bray-Curtis Dissimilarity Over Time (Pediastrum)')

axs[0, 1].set\_xlabel('Age CE')

axs[0, 1].set\_ylabel('Rate of Change')

axs[0, 1].legend(loc='upper left')

axs[0, 1].grid(True)

# Sørensen Dissimilarity Over Time

axs[1, 0].plot(pediastrum\_data['Age CE'], pediastrum\_data['Smoothed Sørensen Dissimilarity'], color='orange', linewidth=2, label='Smoothed Sørensen Dissimilarity')

axs[1, 0].scatter(pediastrum\_data['Age CE'], pediastrum\_data['Sørensen Dissimilarity'], color='brown', s=30, label='Original Data')

axs[1, 0].axhline(y=baseline\_sorensen\_pediastrum, color='grey', linestyle='--', linewidth=1.5, label='Baseline (5th percentile)')

axs[1, 0].set\_title('Sørensen Dissimilarity Over Time (Pediastrum)')

axs[1, 0].set\_xlabel('Age CE')

axs[1, 0].set\_ylabel('Sørensen Dissimilarity')

axs[1, 0].legend(loc='upper left')

axs[1, 0].grid(True)

# PCoA Plot (PC1 Trend Over Time)

axs[1, 1].plot(pediastrum\_data['Age CE'], pc1\_pediastrum, color='blue', linewidth=2, label='PC1 Trend')

axs[1, 1].scatter(pediastrum\_data['Age CE'], pc1\_pediastrum, color='blue', s=30)

axs[1, 1].set\_title('PCoA PC1 Trend Over Time (Pediastrum)')

axs[1, 1].set\_xlabel('Age CE')

axs[1, 1].set\_ylabel('PC1')

axs[1, 1].legend(loc='upper left')

axs[1, 1].grid(True)

plt.tight\_layout()

plt.savefig('pediastrum\_analysis\_publication\_quality.svg')

plt.show()

# Climate periods

import pandas as pd

from scipy.spatial.distance import braycurtis

from sklearn.decomposition import PCA

import numpy as np

import matplotlib.pyplot as plt

from scipy.ndimage import gaussian\_filter1d

# Load the dataset

pediastrum\_data = pd.read\_csv('Pediastrum.csv', delimiter=';')

# Define the climate periods

climate\_periods = {

"RWP": (0, 200),

"DACP": (200, 550),

"S&TWP": (550, 700),

"MCA": (900, 1400),

"LIA": (1400, 1850),

"CWP": (1850, 2023) # Assuming present year as 2023

}

# Function to filter data based on a given period

def filter\_by\_period(data, period):

start, end = period

return data[(data['Age CE'] >= start) & (data['Age CE'] <= end)]

# Filter data for each period

filtered\_data\_by\_period = {period: filter\_by\_period(pediastrum\_data, range\_) for period, range\_ in climate\_periods.items()}

# Extract species data for Bray-Curtis calculation

species\_data\_pediastrum = pediastrum\_data.iloc[:, 2:].astype(float)

# Calculate Bray-Curtis dissimilarity for each consecutive pair of samples

bray\_curtis\_dissimilarity\_pediastrum = [braycurtis(species\_data\_pediastrum.iloc[i], species\_data\_pediastrum.iloc[i+1]) for i in range(len(species\_data\_pediastrum)-1)]

# Insert the Bray-Curtis dissimilarity values into the dataframe

pediastrum\_data['Bray-Curtis Dissimilarity'] = np.nan

pediastrum\_data.loc[1:, 'Bray-Curtis Dissimilarity'] = bray\_curtis\_dissimilarity\_pediastrum

# Calculate the rate of change for Bray-Curtis Dissimilarity

pediastrum\_data['Rate of Change BC'] = pediastrum\_data['Bray-Curtis Dissimilarity'].diff()

# Smooth the Bray-Curtis Dissimilarity using a Gaussian filter

pediastrum\_data['Smoothed BC Dissimilarity'] = gaussian\_filter1d(pediastrum\_data['Bray-Curtis Dissimilarity'].fillna(0), sigma=2)

# Calculate baseline for Bray-Curtis (5th percentile)

baseline\_bc\_pediastrum = np.percentile(pediastrum\_data['Bray-Curtis Dissimilarity'].dropna(), 5)

# Plot the results for each climate period

fig, axs = plt.subplots(len(climate\_periods), 2, figsize=(15, len(climate\_periods) \* 6))

for idx, (period, data) in enumerate(filtered\_data\_by\_period.items()):

# Bray-Curtis Dissimilarity Over Time

axs[idx, 0].plot(data['Age CE'], data['Smoothed BC Dissimilarity'], color='orange', linewidth=2, label='Smoothed BC Dissimilarity')

axs[idx, 0].scatter(data['Age CE'], data['Bray-Curtis Dissimilarity'], color='brown', s=30, label='Original Data')

axs[idx, 0].axhline(y=baseline\_bc\_pediastrum, color='grey', linestyle='--', linewidth=1.5, label='Baseline (5th percentile)')

axs[idx, 0].set\_title(f'Bray-Curtis Dissimilarity Over Time ({period})')

axs[idx, 0].set\_xlabel('Age CE')

axs[idx, 0].set\_ylabel('Bray-Curtis Dissimilarity')

axs[idx, 0].legend(loc='upper left')

axs[idx, 0].grid(True)

# Rate of Change in Bray-Curtis Dissimilarity Over Time

axs[idx, 1].plot(data['Age CE'], data['Rate of Change BC'], color='gray', linewidth=2, label='Rate of Change')

axs[idx, 1].axhline(y=0, color='grey', linestyle='--', linewidth=1.5, label='Baseline')

axs[idx, 1].set\_title(f'Rate of Change in Bray-Curtis Dissimilarity Over Time ({period})')

axs[idx, 1].set\_xlabel('Age CE')

axs[idx, 1].set\_ylabel('Rate of Change')

axs[idx, 1].legend(loc='upper left')

axs[idx, 1].grid(True)

plt.tight\_layout()

plt.savefig('pediastrum\_climate\_periods\_analysis.svg')

plt.show()