Analysis of Speleothem Data Using SISALv3

Introduction

Speleothems, such as stalagmites and stalactites, serve as crucial archives of past climate conditions. By analyzing their isotopic compositions, valuable historical climate data can be inferred. This report outlines the methodology and findings from our analysis using the SISALv3 database, focusing on stable isotope compositions (δ18O and δ13C) in speleothem samples.

Methodology

Data Extraction

1. **Database Connection:** Connected to the SISALv3 MySQL database using SQLAlchemy. Key tables included were sample, d18O, and d13C.

2. **SQL Query:** Extracted sample_id, depth_sample, d18O_measurement (renamed as d18O), and d13C_measurement (renamed as d13C) from the respective tables, ensuring no NULL values were included.

3. **Data Loading:** Loaded the queried data into a Pandas DataFrame for analysis.

Data Processing

1. **Sorting and Splitting:** Sorted the data by depth_sample to maintain chronological order and split the dataset into training (80%) and testing (20%) sets.
2. **Time Series Analysis:**
    a. **ARIMA Model:** Fitted an ARIMA model (order 5,1,0) to the d18O values in the training set. Forecasted d18O values for the testing set, and added predictions to the DataFrame.
    b. **Performance Metrics:** Calculated Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to evaluate model performance.
3. **PCA Analysis:**
    a. Standardized the depth_sample, d18O, and d13C columns.
    b. Performed Principal Component Analysis (PCA) to reduce dimensionality and identify principal components.

Visualization

1. **Time Series Plot:**
    a. Plotted the d18O values from the training, testing sets, and model predictions against depth_sample.
    b. Included a legend to distinguish between training data, testing data, and predictions.
2. **PCA Plot:**
    a. Created a scatter plot of the first two principal components, color-coded by depth_sample.

Findings

Descriptive Statistics

- The dataset included 266,178 samples.
- Calculated and displayed key statistics for depth_sample, d18O, and d13C.

ARIMA Model

- The model captured the overall trend in the training data but faced limitations in accurately predicting variability in the testing set.
- Performance metrics:
    - MAE: 2.28
    - MSE: 8.68
    - RMSE: 2.95

PCA Analysis

- The first two principal components explained a significant portion of the variance in the data.
- The scatter plot revealed patterns indicating changes in isotope composition with depth.

Conclusion

The analysis provided insights into the isotopic composition of speleothems and their relationship with depth. The ARIMA model highlighted the challenges in time series prediction for speleothem data, while the PCA revealed underlying patterns. These findings enhance our understanding of past climate conditions recorded in speleothems.