Repository guide

This repository contains datasets supporting the evaluation of voltammetric signal-separation procedures for capsaicin and dihydrocapsaicin. The deposited files include datasets, an example baseline-corrected and smoothed voltammetric signal provided for workflow demonstration, and example output files generated after each signal-separation method.

Contents

.zip files

This repository contains the data used to verify the accuracy of the models developed for separating CPS and DHC signals. It contains three groups (*.zip* files), named according to the terminology used in the manuscript:

CPS-G1.zip – A folder containing a series of individual voltammograms for CPS

DHC-G2.zip – A folder containing a series of individual voltammograms for DHC

MIX-G3.zip – A folder containing a series of individual voltammograms for a mixture of CPS and DHC. Among these, three data series can be distinguished, in which:

- Voltammograms with a variable CPS concentration and a constant DHC concentration
- Voltammograms with a constant CPS concentration and a variable DHC concentration
- Voltammograms with a variable CPS and DHC concentrations

The repository also contains a file named *concentrations.xlsx*, which lists the CPS and DHC concentrations for each *.xlsx* file. The file has three columns:

- *file* – lists the name of the file
- *cps* – describes the CPS concentration [µmol · L^{-1}]
- *dhc* – describes the DHC concentration [µmol · L^{-1}]

example_processed_voltammograms.xlsx

The file contains a sample measurement series for the standard addition method, compiled during testing of one of the Bhut Jolokia pepper extracts under study. The voltammograms have been baseline-corrected and smoothed, making them suitable for direct analysis using the previously developed models for separating CPS and DHC signals.

Data format

The first column of files with the *.xlsx* extension containing voltammetric traces lists the potential values E [V]. The next column contains the current response expressed in µA. Rows correspond to potential points, while columns correspond to individual voltammetric signals or separate signal components, depending on the file.

Code usage instructions

The scripts used for signal preprocessing, independent component analysis, Gaussian curve fitting and Lorentzian curve fitting were prepared in Python and run in the Google Colab environment. The code is organized into consecutive sections corresponding to data import, preprocessing, model parameter definition, signal separation, visualization and export of the obtained results.

Before running the code, the user should upload the appropriate Excel file containing the voltammetric data to the Colab working environment. The input file should contain potential values in the first column and the corresponding current responses in the following columns.

For curve-fitting procedures, the potential window used for modelling should be adjusted individually in the section defining the model parameters. This is necessary because the position and width of the analytical signal may vary depending on the relative contribution and concentration of the analysed compounds. After defining the input file and, where necessary, the fitting window, the code cells should be executed sequentially from the beginning to the end of the notebook.

The generated outputs include separated CPS and DHC signal components, the reconstructed total signal and files containing the calculated fitting or separation results. The scripts are available from the corresponding author upon reasonable request.

The scripts were written in Python 3.12.12 Required packages: NumPy (v. 2.0.2), Matplotlib (v. 3.10.0), Pandas (v. 2.2.2), SciPy (v. 1.16.3), Scikit-learn (v. 1.6.1), Seaborn (v. 0.13.2).

Notes

The complete numerical results, including voltammetric determinations, HPLC-DAD reference values and relative error values, are reported in the manuscript. The files deposited in this repository are intended to support the evaluation and demonstration of the signal-separation workflow rather than to duplicate all results presented in the article.

Code availability

The scripts used for signal preprocessing, independent component analysis, Gaussian curve fitting and Lorentzian curve fitting are available from the corresponding author upon reasonable request.