VaporFit User Guide

Przemysław Pastwa and Piotr Bruździak
Department of Physical Chemistry, Gdańsk University of Technology,
Narutowicza 11-12, 80-233 Gdańsk, Poland
piotr.bruzdziak@pg.edu.pl

Contents

1 Introduction

VaporFit is an open-source software tool designed for automated atmospheric correction in FTIR spectroscopy. It employs a **least squares fitting approach** using multiple reference atmospheric/vapor spectra, dynamically adjusting subtraction coefficients to improve accuracy. You are welcome to send us any comments or suggestions.

2 License

This software is distributed under the **GNU GPL v.3.0** license + citation requirement (see LICENSE file). By using this software, you agree to its terms and conditions.

3 Citations

If you use VaporFit in your research, you must cite the following publications:

- Bruzdziak P., Vapor correction of FTIR spectra – A simple automatic least squares approach, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2019, **223**, 1–4, doi:10.1016/j.saa.2019.117373
- Pastwa P., Bruzdziak P., VaporFit: An Open-Source Software for Accurate Atmospheric Correction of FTIR Spectra, *in preparation*

4 Data sets

Three sample data sets are available to download from the project site:

- `test_data_betaine_set.zip` — a set of 29 spectra of betaine aqueous solutions with two atmospheric spectra (2 cm^{-1}, 128 scans, single background measurement),
- `test_data_D2O-H2O_set.zip` — a set of 11 spectra of D_2O/H_2O mixtures with three atmospheric spectra (4 cm^{-1}, 128 scans, single background measurement).
- `test_data_urea_evaporation_set.zip` – a set of 20 spectra of a 20 µL urea solution (initial concentration 1 M), freely evaporated on the ATR crystal for 20 minutes, with two atmospheric spectra (resolution: 4 cm^{-1}, 64 scans, single background measurement).

All spectra were measured on a Bruker Invenio-R FTIR spectrometer equipped with a single reflection diamond ATR accessory. The spectrometer was purged with dry nitrogen (Claind Brezza NiGen LCMS 40.1) for 30 minutes before and during the measurement.

5 User Guide

5.1 VaporFit window overview

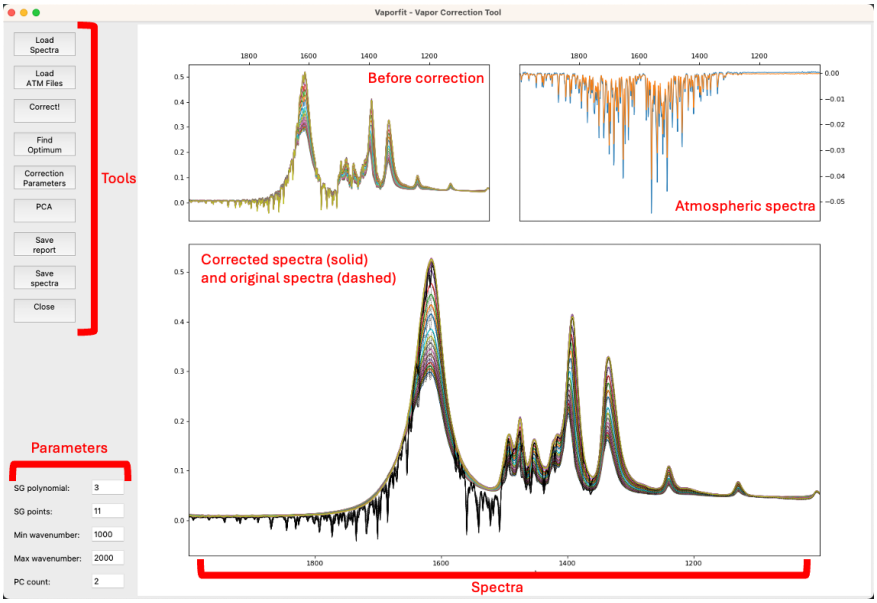


Figure 1: **Left**: tools and parameters. **Right**: original spectra before correction (upper left panel), atmospheric spectra (upper right panel), correction results (bottom panel; solid lines–corrected spectra, dashed lines–spectra before corrections).

Use the **left mouse button** to select and zoom a fragment of the spectrum (each panel is independent). Use the **right mouse button** to reset the view. Zooming in or out does not affect the wavenumber range for correction indicated in the section of the VaporFit window.

5.2 Load Spectral Files

Upload spectra that require correction.

Supported formats:

- CSV
- DPT

Important: Ensure that the spectral resolution, wavenumber range, number of points, and units are **consistent across all files** to avoid errors.

5.3 Load Atmospheric Spectra

Upload atmospheric spectra.

- The format and spectral parameters must match those of the spectra being corrected.

5.4 Set SG Parameters and Wavenumber Range

Adjust the **desired wavenumber range** and the **Savitzky-Golay (SG) parameters** in the Parameters Section of the main window.

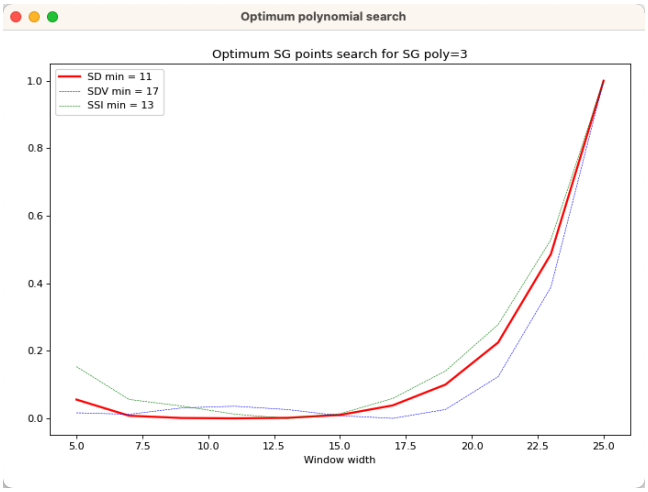
- Default settings are predefined for standard atmospheric corrections for the amide I region of the spectra. Adjust as needed.
- If values exceed the range of loaded spectra, they will be limited to the closest minimum or maximum wavenumber.
- It is **not recommended** to perform correction over the entire spectrum range (typically 4000–400 cm^{-1}). Choosing a narrower region improves result quality.
- Modifying SG parameters may improve results, but incorrect settings can introduce artifacts. A good rule of thumb is to use the lowest possible SG parameter values for optimal correction.

5.5 Run the Correction Algorithm

Press **"Correct!"** to apply the atmospheric correction algorithm.

- The resulting graph will display:
  - **Original spectra** (black dashed lines)
  - **Corrected spectra** (color-coded)
- The correction process **automatically optimizes subtraction coefficients** based on least squares fitting.

5.6 (Optional) Run "Optimum Search" Tool



This tool generates a **graph showing the dependence of four spectral smoothness metrics on the SG window size** (+/-10 points) for a given **SG polynomial order**. Values represent the sum of these metrics for all corrected spectra.

- The minima of these functions (shown in the legend) suggest **potential optimal SG window sizes**.
- However, these values **should not be used blindly**. Verify with **PCA analysis** to ensure meaningful correction.
- The four available smoothness metrics may have different applications depending on the spectral characteristics. For amide I band spectra, the SD parameter works best.
- Depending on computational power, this step may take some time.

5.6.1 Smoothness Functions

1. **Standard Deviation of the Residue After Smoothing (SD)**:

$$\sigma_{\text{residue}} = \text{std}(\text{spectrum} - \text{smoothed spectrum})$$

where smoothed spectrum is the result of applying the Savitzky-Golay filter to the spectrum Spectrum (with default values of polynomial order 3 and window width 11).

2. **Spectral Smoothness Index (SSI)**:

$$\text{SSI} = \frac{\sum \left(\frac{d}{dx}\text{spectrum}\right)^2}{\sum(\text{spectrum})^2}$$

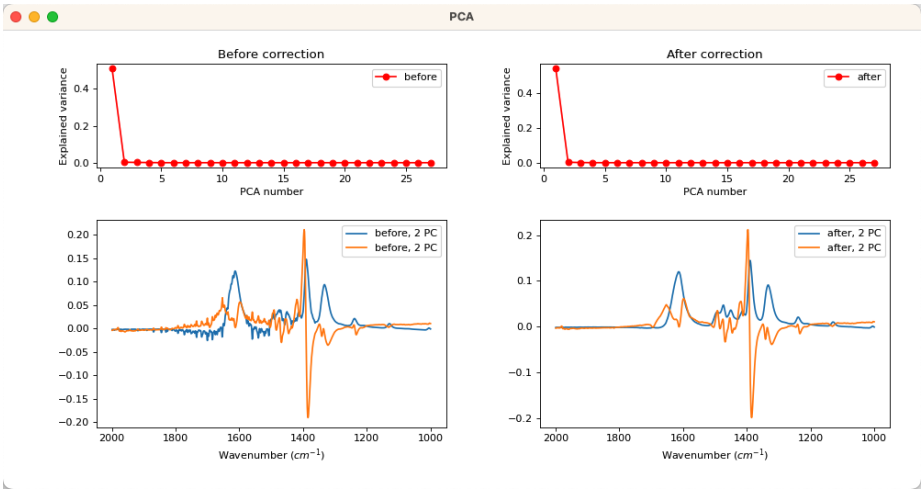
where $\sum$ denotes the sum of all data points of the spectrum or its function, and $\frac{d}{dx}$spectrum is the first derivative of the spectrum, i.e., the differences between consecutive values of the spectrum.

3. **Second Derivative Variance (SDV)**:

$$\text{Variance of second derivative} = \text{Var}\left(\frac{d^2}{dx^2}\text{spectrum}\right)$$

where $\frac{d}{dx}$spectrum is the second derivative of the spectrum, i.e., the differences between consecutive values of the $\frac{d}{dx}$spectrum.

5.7 (Optional) Perform PCA Analysis



After correction, press **"PCA"** to analyze the **principal components (PCs)** of the spectral series **before** (left) and **after** (right) correction. You can set the number of extracted PC spectra in the main window. Use the **left mouse button** to select and zoom a fragment of the spectrum. Use the **right mouse button** to reset the view. Zooming in or out in one spectrum graph will affect the other.

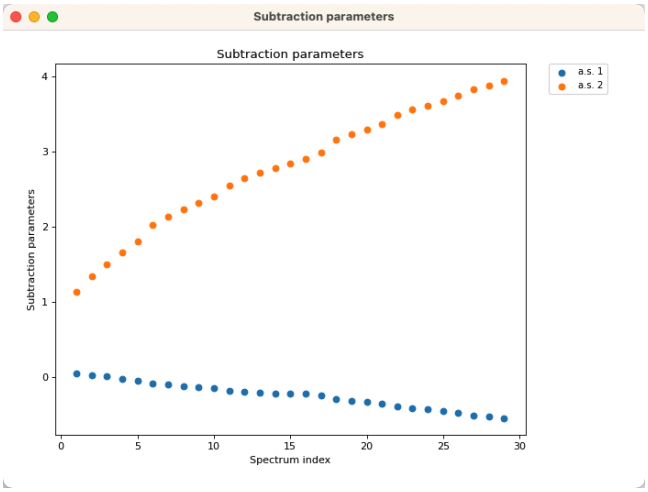
PCA window includes:

- **Explained Variance (EV) plots** (top)—A sharp drop suggests the **minimum number of significant PCs**.
- **Principal component shapes** (bottom)—Before correction, atmospheric components should be recognizable in PCs. After proper correction, they should no longer be present.

Important Notes:

- PCA **will not run** if the correction is incomplete or if fewer than **two spectra** are available.
- Misinterpretation of PCA can lead to incorrect assumptions about correction quality.
- PCA in this software is for visual assessment only. The actual number and shape of principal components should be determined using specialized software or appropriate libraries.

5.8 (Optional) Check Correction Coefficients



Use the **"Parameters Graph"** tool to inspect the correction coefficients applied to each atmospheric reference spectrum.

6 Save Results

6.1 Save Report

Press **“Save report”** to export **multiple CSV files**, which include:

- General information about the correction (date, number of spectra, file names, and paths)
- Spectra **before and after correction** (as a matrix; for the selected wavenumber range)
- **Atmospheric reference spectra** within the same range (as a matrix)
- **Correction parameters** for each spectrum
- **PCA results**, if performed

You will be asked for the report filename. The given name will serve as a prefix to all these files.

6.2 Save Corrected Spectra

Press **“Export Spectra”** to save all corrected spectra, each in a separate CSV file, in the same folder as the original spectra. All filenames will start with the ”corr_” prefix.

7 Exit the Program

Press **“Close”** to exit VaporFit.

8 Final Notes

- VaporFit is designed for atmospheric corrections in **FTIR spectroscopy** but can process any CSV spectral files.
- If you experience issues with correction quality, double-check **input spectra consistency** and consider adjusting **wavenumber ranges** or **reference spectra selection**.
- If correction is unsatisfactory, ensure you have enough atmospheric spectra covering varied cconditions; set the SG polynomial order to 3 and theSG window size to 5, or try correction with a single atmospheric spectrum.
- The software is actively developed, and new features may be added in future releases.
- We tested and compiled the software on MacOS v.15.3.1 (Mac Mni M4) and Windows 11 (Inter Core i5-8365).
- Try running the Python code in your Python environment if you are experiencing issues with compiled versions.