

炭質物ラマンスペクトルの自動ピーク分離コードの使用マニュアル

金木 俊也

産業技術総合研究所 地質調査総合センター 地圏資源環境研究部門 地圏メカニクス研究グループ

kaneki.shunya@aist.go.jp

2024 年 3 月 22 日

はじめに

本マニュアルでは、Kaneki et al. (submitted to *Progress in Earth and Planetary Science*) が開発した、炭質物ラマンスペクトルの自動ピーク分離コードを動かす手順を解説する。プログラミング初学者にも分かりやすい文章を心がけているため、勝手知ったる方々にとって丁寧すぎるくらいがあると思われるが、ご容赦頂きたい。本マニュアルおよび本コードの使用に起因する全ての損害や不利益について、所属機関および筆者はいかなる責任も負いかねることをご承知願いたい。

1 Python環境の構築

コードは Python で書かれているため、まずは解析用 PC に Python 環境を構築する必要がある。環境構築は初学者にとって敷居が高いため、挫折する方も結構いるのではなかろうか（筆者の失敗バイアスかもしれないが）。そのような事態を避けるため、ここでは Anaconda というソフトウェアを利用する前提で話を進める。Anaconda は Windows、Macintosh、Linux に対応しており、Python 環境の構築に付随する面倒事に対処してくれる。以下、Anaconda を用いた Python 環境の構築手順を示す。筆者は Macintosh 以外の OS で Anaconda をインストールしたことはないが、手順に大きな変化はないだろう（あったらすみません）。なお、**所属機関の性質や規模によっては、Anaconda の使用に際して有償のライセンス契約が必要な場合があるので要注意**。無償使用にこだわるなら Miniforge が代替候補となるが、環境構築手順は少し難解になる。なお、Anaconda と Miniforge の基本機能はほぼ同じである。

1. 「anaconda install」などで検索して、OS に対応した Anaconda installer をダウンロードする。特にこだわりが無ければ、最新バージョンの Python の 64-Bit の Graphical Installer が良いだろう。なお、コードの開発環境と PC での動作環境を厳密に一致させたい場合、インストール後に Python 自体やライブラリのバージョンを調整する必要がある（このページの最後の補足を参照）。
2. ダウンロードした installer を実行し、Anaconda を PC にインストールする。インストールが終わった段階で、コードを走らせるために必要な Python 環境が構築されている。とても便利。ただし、PC のユーザー名に全角文字が含まれているとインストールが失敗する、という報告を誰かのブログで見た。
3. Anaconda-Navigator を起動し、Spyder を起動する。（.py が動けば何でも良いが、コード開発を Spyder で行ったのでここではそれに倣う。）この段階では Spyder がインストールされていないかもしれないので、その場合はまず Spyder をインストールする（Install ボタンを押すだけで良かったと思う）。Spyder を起動して図 1 のようなウィンドウが開いたら、Python 環境の構築は完了。
4. （補足）本コードでは有名なライブラリのみを使用しているので、バージョンの違いが結果に及ぼす影響は無視できると思うが、可能なら解析用 PC の Python 環境とコードの開発環境は一致することが望ましい（コードの開発環境の詳細は Readme.txt を参照）。例えば、解析用の仮想環境を作って、そこに必要なライブラリをインストールしていく、とか。

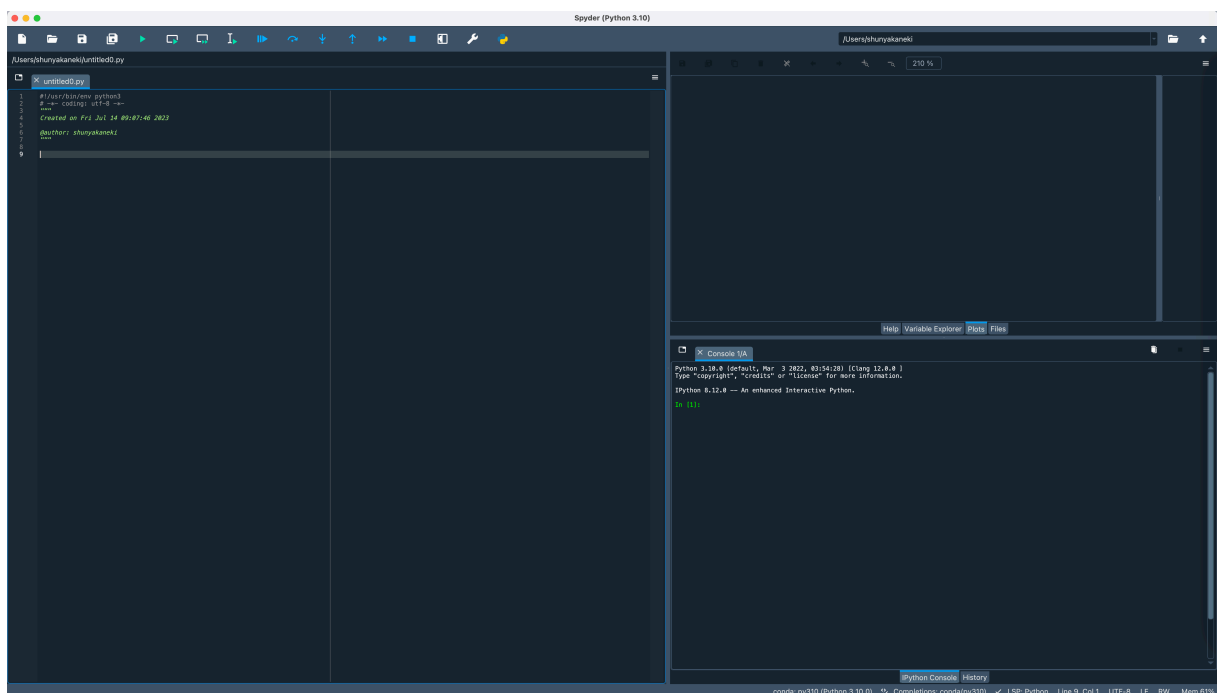
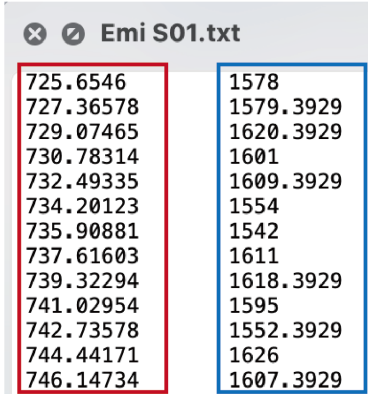


図 1: Spyder の起動画面。

2 解析可能な炭質物ラマンスペクトルの生データ

解析する炭質物ラマンスペクトルの生データファイルは、以下の二つの事項を満たしている必要がある。また、これらの必要事項を満たしているファイル例を図2に示す。

- データの拡張子は、.txt、.csv、.CSV の三種類のいずれかである。
- データの一列目にラマン波長（単位は cm^{-1} ）、二列目に強度が記録されており、数値データのみで構成されている。時折、強度の欄に"#NaN"と入力されているデータがあるが、これではコードは動かないので、削除する必要がある。手作業でもできるが数が多いと非常に面倒なので、金木まで問い合わせていただければ、自動で削除するコードをお渡しすることもできる。



Raman shift (/cm) (1st column)	Intensity (2nd column)
725.6546	1578
727.36578	1579.3929
729.07465	1620.3929
730.78314	1601
732.49335	1609.3929
734.20123	1554
735.90881	1542
737.61603	1611
739.32294	1618.3929
741.02954	1595
742.73578	1552.3929
744.44171	1626
746.14734	1607.3929

図2: 解析可能なラマンスペクトルのデータファイル例。

3 解析用のデータフォルダの構築

1. AutoRaman_K2024 フォルダを開いて、中身が図 3 のようになっていることを確認する。（もし__pycache__があっても気にしなくて良い。）
2. AutoRaman_K2024 フォルダをコピーし（コピー先はどこでも良い）、コピーしたフォルダ名を解析する試料名に変更する。以下、これを解析フォルダと呼ぶ。
3. 準備した生データを解析フォルダ中の raw フォルダにコピーする（図 4）。

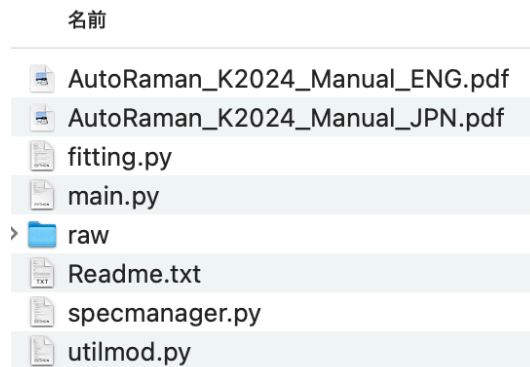


図 3: AutoRaman_K2024 フォルダの中身。

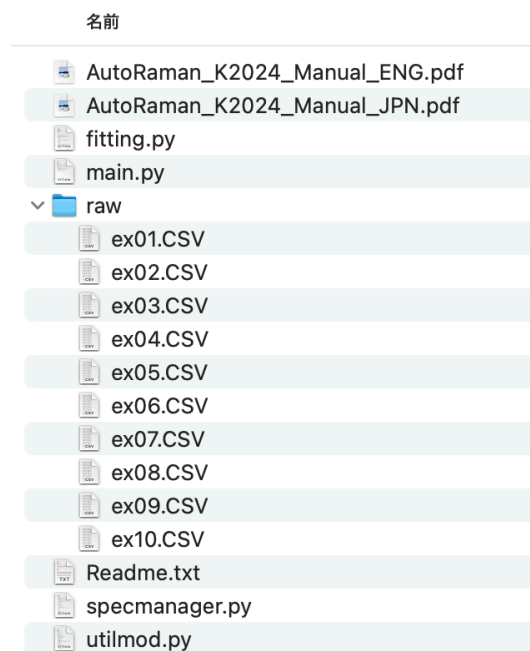


図 4: 生データを raw フォルダにコピーした後の解析フォルダ。

4 自動ピーク分離の実行

1. Spyder を起動し、画面右上の小窓の File タブで解析フォルダ中の main.py を開く（図 5）。
2. 解析データおよび出力データに合わせて INPUT 項目を設定する。dtype は生データファイルの拡張子で、.csv or .CSV or .txt のいずれか。**datadir は基本的に変更しない**。figurePDF や figurePNG を True にすると、各拡張子でピーク分離画像が保存される。specdat を True にすると、ピーク分離結果の csv ファイルが保存される。
3. F5 を押して実行する（上部の再生マークでも実行できる）。解析中のデータ名が右下に追加されていく。
4. 解析が終わると、右下に推定温度の平均値が表示される（図 6）。マシンパワーにもよるが、図を保存する場合は一ファイルあたり数秒、しない場合は一ファイルあたり一秒未満で解析が終わる。
5. 解析フォルダ中に”result”というフォルダが自動生成されている。その中の”image”にはピーク分離画像およびラマンパラメータのヒストグラムが、”param”には初期および最適化後のラマンパラメータの値が、”spec”には最適化したスペクトル強度が保存されている。INPUT 項目を全て False にした場合でも、ヒストグラムのみは必ず出力される。”param”中の”all_fin.csv”には、各スペクトルで計算されたラマンパラメータが保存されている。論文投稿時にフィッティングデータの提出を求められた時はこれを出せば一発。”image”中の図は、後に解析データを見直すときや論文の図を清書する時などに有用。”param.csv”には、自動ピーク分離で計算された R1 比、R2 比、D および G バンドの位置と半値幅について、平均値 mean（平均から 2SD 以上外れたデータは除外済）、標準偏差 SD、標準誤差 SE、データ数 No が記録されている。”temp.csv”には、R2 比の平均値および標準誤差から計算された温度 T、温度の標準誤差 Tse、温度の 95%予測区間 TerPI95 が記録されている。
6. （補足）本コードでは、フィッティング前に線形ベースライン補正を実施している。その際に参照するデータ範囲は、 $1100\text{--}1150\text{ cm}^{-1}$ と $1700\text{--}1750\text{ cm}^{-1}$ もしくは $1200\text{--}1250\text{ cm}^{-1}$ と $1700\text{--}1750\text{ cm}^{-1}$ のどちらかから自動で決定される。これらの範囲を手動で変更したい場合は、”specmanager.py”中の 15-16 行目もしくは 25-26 行目の該当する数字を変更すれば良い。（ベースライン補正に解析者の主観性が入ることには留意する。）

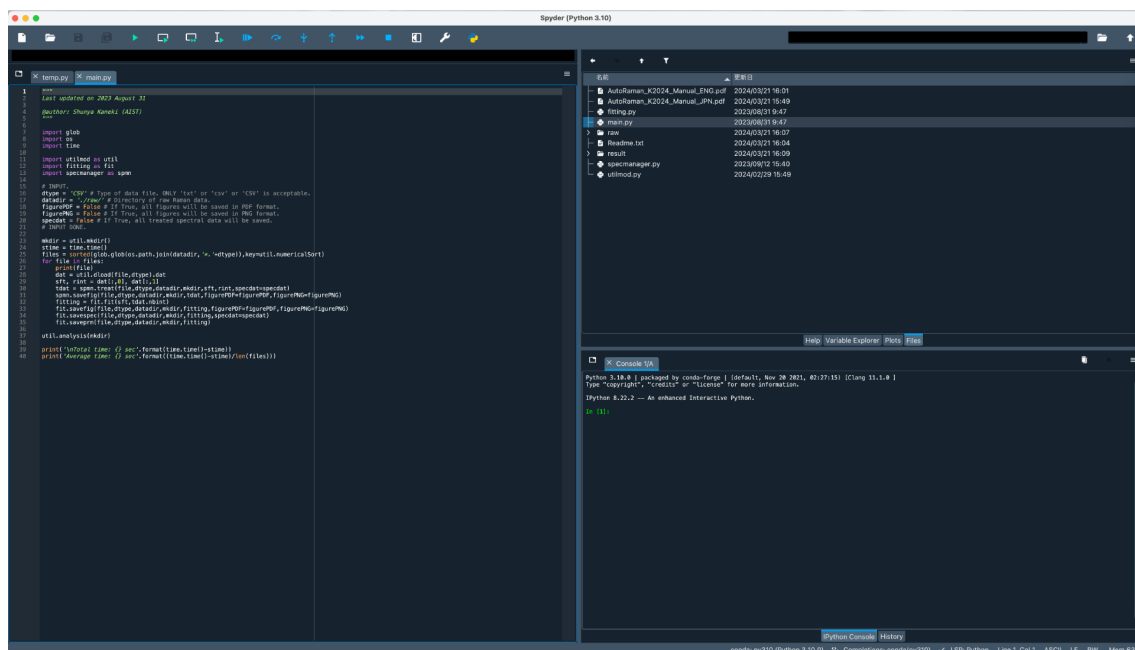


図 5: main.py を開いた画面。

