

震源机制解的谱聚类计算软件

用户操作手册

V1.0

目录

1. 简介	1
2. 运行环境.....	1
3. 安装设置.....	2
4. 软件架构.....	2
5. 界面与操作.....	4
5.1 GUI	4
5.2 命令行操作.....	14
6. 说明	15

1. 简介

震源机制解的谱聚类（SC4FMS）是一套专为震源机制解的聚类而设计的软件。针对震源机制解数据的特点，采用最小旋转角为相似度矩阵，利用 NCut 切图准则完成类型判别，以 Gap 统计确定聚类数的最优解，从而对海量的震源机制解数据进行快速准确的类型划分，进而获得一系列与震源机制一致性的相关结论。SC4FMS 不依赖地球物理等相关理论解释，更有利于从数据层面分析震源机制特征。同时，方便用户从事各类数据分析，以期获得更有价值的结论。

SC4FMS 有独立的图形用户界面（GUI），可方便用户交互操作，易于上手。同时，也可以使用 python 语言进行批处理，最大限度保证用户方便快捷使用。SC4FMS 使用第三方模块较少，安装简便，使用简洁；支持多种操作系统，对电脑性能要求较低。适合各类地震相关从业的用户。

该软件包含的主要功能如下：

- 震源机制解的谱聚类；
- 求取每类震源机制解的平均解；
- 利用 Gap-Statistics 方法计算最佳聚类数；
- 相关的图件绘制。

2. 运行环境

SC4FMS 支持多种操作平台，包括 macOS，Linux 和 Windows 操作系统。运行该程序需要安装 Python 3.x（3.7 和 3.8 已测试可以完美运行，其它版本可自行测试）。

Python 安装方式请参照官方网站：<https://www.python.org/>。

建议使用 anaconda 进行配置安装。Anaconda 安装请参照官方网站：<https://www.anaconda.com/>。

安装完成后，请运行 python 程序，确保安装成功。

```
(base) C:\Users\Lin>python --version
Python 3.8.8

(base) C:\Users\Lin>python
Python 3.8.8 (default, Apr 13 2021, 15:08:03) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

3. 安装设置

通过 GitHub 获得源程序：`git clone git://github.com/forkiter/SC4FMS`。或直接通过 <https://github.com/forkiter/SC4FMS> 下载源程序包，并解压至安装目录。

建议采用以下安装方法：

(1) 安装虚拟环境

```
conda create -n SC4FMS python=3.8
```

```
conda activate SC4FMS
```

(2) 安装运行该程序所需的模块

```
pip install pandas
```

```
pip install scipy
```

```
pip install matplotlib
```

或者在主程序目录中输入：

```
pip install -r requirements.txt
```

(3) 在主程序目录运行 `run_gui.py`

```
python run_gui.py
```

若弹出程序主界面，则表示安装成功；否则，请根据 `Trackback` 提示进行其它模块安装，或直接联系作者。

4. 软件架构

软件包含 4 个文件夹：`core`，`data`，`docs` 和 `img`，主程序目录下包含 2 个 `python` 主运行程序：`example.py`，`run_gui.py`。分别介绍如下：

(1) `core`：核心程序包，包含运行该软件的所有程序。

(2) `data`：输入和输出文件存储区，文件格式均为 `csv`。

(3) `docs`：说明文档和相关软件图件存储区。

- (4) img: 输出图像结果存储区。
- (5) example.py: 使用命令行运行软件的示例。
- (6) run_gui.py: 运行 GUI 界面的程序。

Tree 图如下:

```

├── core
│   ├── ck_dist.py
│   ├── data_process.py
│   ├── gap_evaluation.py
│   ├── __init__.py
│   ├── kmeans.py
│   ├── lloyd.py
│   ├── sc4fms.py
│   ├── s_gui.py
│   ├── similarity.py
│   └── spectral_cluster.py
├── data
│   └── example_data.csv
├── docs
│   ├── fig
│   │   ├── head.ico
│   │   └── main_window.jpg
│   ├── LinQingxi.pdf
│   └── UserGuide.pdf
├── img
├── example.py
├── __init__.py
├── README.md
├── requirements.txt
└── run_gui.py

```

5. 界面与操作

该软件可使用图形用户界面（GUI）进行人机交互，也可以通过命令进行批处理。

5.1 GUI

在主程序目录中执行以下命令：

```
python run_gui.py
```



程序界面主要分为左右两个半区。左半区为用户操作，实现计算和绘图功能；右半区为数据、图像以及日志的显示窗口。

5.1.1 数据导入

在数据导入功能区点“导入”按钮，将原始震源机制解数据导入，并点击“确认”按钮。文件路径窗口会显示导入数据的绝对路径。

或者直接在“文件路径”中录入需要导入数据的绝对路径，然后点击“确定”按钮，同样可以导入数据。

导入后，“数据显示窗口”会显示该数据集的所有信息，并在“日志显示窗口”

给出文件地址路径，以及数据总条数。



导入数据需注意数据格式，文件必须为 csv 格式，数据需包含走向、滑动角和倾角信息。Label 为聚类结果，如果导入数据中包含了聚类结果，Label 列会自动列出相关聚类结果。csv 文件请参考/data/example_data.csv 的文件结构。

5.1.2 谱聚类计算

在谱聚类功能操作区，输入希望聚成 k 类的 k 值。k 为整数值，例如输入“8”。由于构建相似度矩阵时，采用了 KNN 近邻法，因此需要输入近邻数和核比例因子。根据震源机制解的数据特点，本软件给定了默认值，可不予修改。用户也可自行选择合适的参数进行调试。相关理论请自行查阅文献。

点击“开始计算”，程序会启动谱聚类计算，并在日志窗口显示出当前计算进度。计算完成后，日志窗口会显示“Spectral cluster complete”的字样，数据显示窗口的 Label 值会更新完成，并展示给用户。

震源机制解的谱聚类计算软件 V1.0



计算完成后，在 data 文件夹下会自动生成结果文件，命名为“result_sc_”+导入文件的文件名+“.csv”。例如上述示例中，导入的文件名为“example_data.csv”，生成的结果文件为“result_sc_example_data.csv”。

本软件所生成的结果文件均为 csv 格式，且均为“结果名+文件名”的命名方式，以下将不再重申。

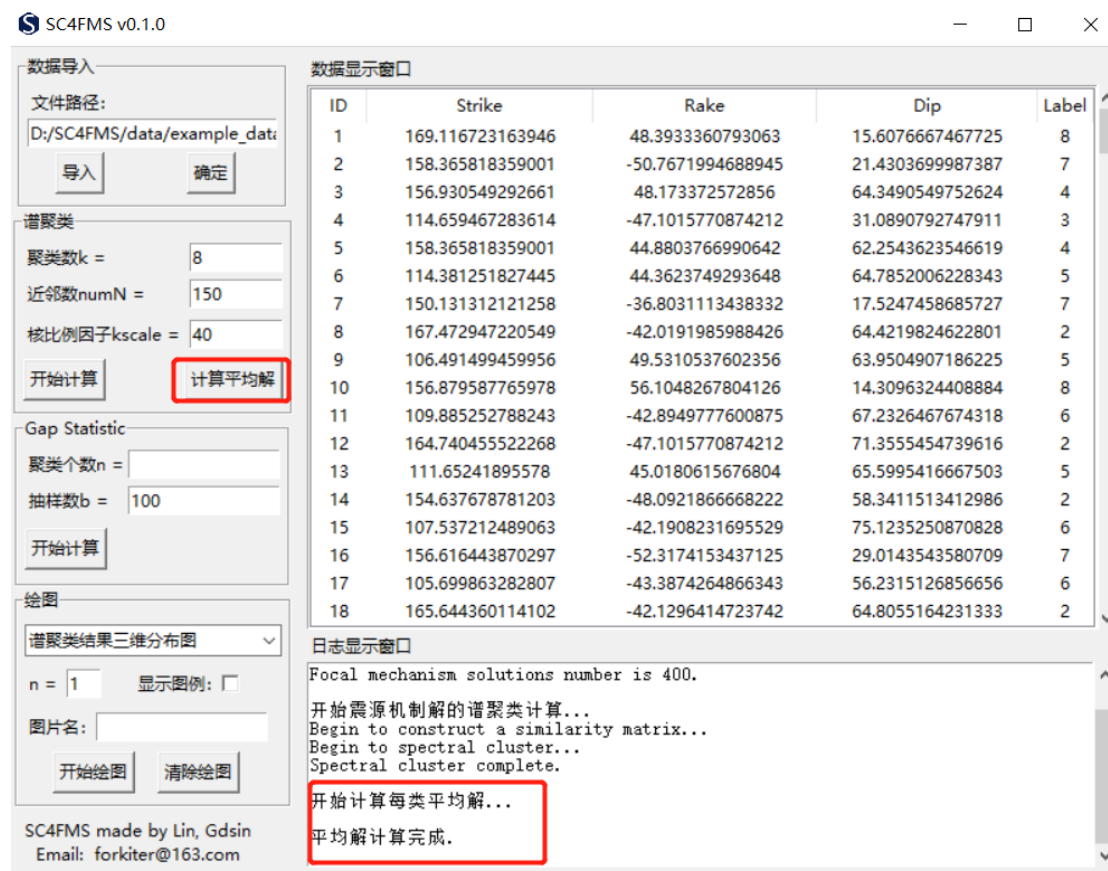
聚类结果文件是在原有数据基础上，增加了“label”列，格式如下：

strike	rake	dip	label
169.1167	48.39334	15.60767	4
158.3658	-50.7672	21.43037	5
156.9305	48.17337	64.34905	1
114.6595	-47.1016	31.08908	7
158.3658	44.88038	62.25436	1
114.3813	44.36237	64.7852	6
150.1313	-36.8031	17.52475	5
167.4729	-42.0192	64.42198	8
106.4915	49.53105	63.95049	6

5.1.3 平均解计算

在谱聚类功能操作区，点击“计算平均解”，启动平均解的计算。平均解的

计算采用的是刁桂苓的 P 轴和 T 轴和矢量算法。日志显示窗口会给出“平均解计算完成”的提示。



计算完成后，在 data 文件夹下自动生成 2 个结果文件，分别以“result_pt”和“result_average”开头。

result_pt: 将所有样本数据的走向、滑动角和倾角值转换为 P 轴和 T 轴的方位角（trend）和俯仰角（plunge）。生成的数据结构如下：

ptr	ppl	ttr	tpl	label
112.9888	32.60078	314.9263	55.41493	1
94.82007	32.69037	291.9069	56.12315	1
104.4397	22.97379	331.7402	57.98921	1
95.71243	27.99746	311.5907	56.72923	1
106.9094	32.23941	311.0109	55.35834	1
107.0535	31.07543	313.4198	56.07425	1
106.8483	21.29115	338.4638	57.88721	1
103.8695	18.69149	335.1045	61.61629	1
90.81211	20.69559	316.0001	61.80763	1

result_average: 计算得到的每一类平均解结果。结果中不仅包含平均解两

个节面的走向、滑动角、倾角数据，还包含 P 轴和 T 轴数据。第 n 行表示第 n 类的平均解结果。生成的数据结构如下：

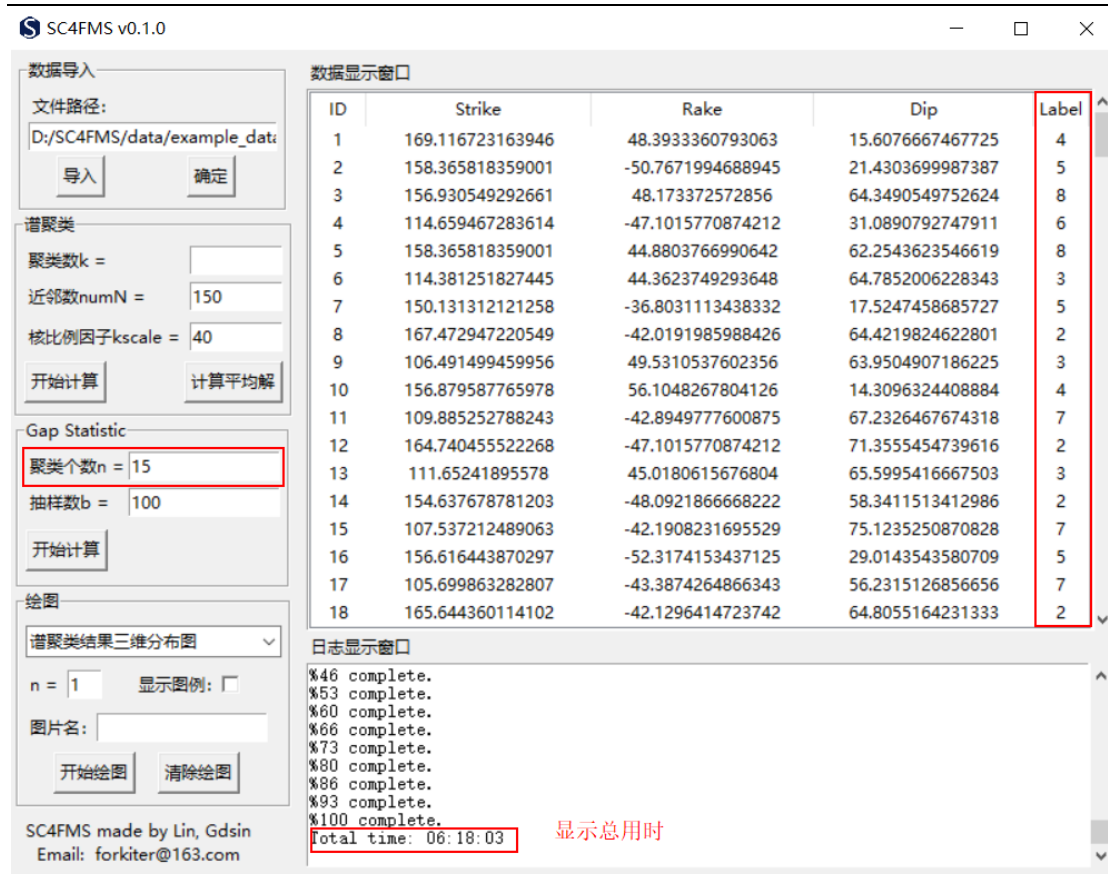
strike1	rake1	dip1	strike2	rake2	dip2	ptr	ppl	ttr	tpl
27.67022	108.4896	72.34826	159.8723	45.09749	25.34735	103.3297	25.1264	323.1549	58.58884
338.5492	108.5434	72.39115	110.5952	44.94281	25.35665	54.16413	25.15521	274.0632	58.52804
159.6848	-45.1404	64.51594	273.0672	-145.952	50.21776	119.4074	49.17389	219.397	8.523657
242.7345	-108.435	72.33227	110.4178	-45.2061	25.31998	127.2984	58.62542	347.0333	25.12365
159.7636	45.04974	64.50877	46.44172	145.9859	50.29577	280.1004	8.476146	20.01253	49.11605
110.6566	45.09352	64.48839	357.2874	145.9413	50.26756	230.968	8.478507	330.8983	49.15938
291.8473	-108.507	72.3658	159.7017	-45.0409	25.34797	176.3556	58.56581	36.2027	25.13924
110.5829	-45.1119	64.52161	223.9402	-145.972	50.23915	70.30745	49.15058	170.2791	8.515565

5.1.4 Gap-Statistics 计算

在 Gap Statistics 功能操作区，输入整数 n 值。n 值表示在利用 Gap Statistics 计算时的最大聚类数。例如：n=15 时，Gap Statistics 会按照聚类数 $k=1,2,3,\dots,15$ 的方式，依次聚类计算，得到每个 k 时的 Gap 值。抽样数 b 是指对原样本进行 bootstrap 抽样的个数，b 值越大，结果越准确稳定。抽样数 b 已给出默认值，可不予修改。具体原理请自行查阅相关文献。

用户也可以修改“谱聚类”功能区中的近邻数和核比例因子参数，以便在 Gap-Statistics 计算中调试合适的参数。

点击“开始计算”，程序会启动 Gap Statistics 计算，并在日志窗口显示出当前计算进度。计算完成后，日志窗口会显示“”的字样，并给出总的运算时间。数据显示窗口的 Label 值会更新成最优 k 值下的最优聚类结果，并展示给用户。



请注意，Gap Statistics运算时间较长，这和用户计算机配置有关。例如示例中的运行时间为6小时18分。长时间运行时，请勿触动窗口，可能会导致程序崩溃。

计算完成后，在data文件夹下会自动生成3个结果文件，分别以“result_gap”、“result_op_sc”和“result_labels”开头，介绍如下：

result_gap: 为计算得到的gap值和se值文件。gap值越大，表示聚类效果越好；se值越小，表示标准误差越小，数据越可靠。示例中n取值15，得到的结果文件结构如下：

label_1	label_2	label_3	label_4	label_5	label_6	label_7	label_8	label_9	label_10	label_11	label_12	label_13	label_14	label_15
-0.21493	-0.15351	-0.19454	0.034349	0.032875	0.111456	0.310099	0.537904	0.524719	0.518803	0.533085	0.416109	0.422363	0.3815	0.516362
0.021667	0.022376	0.019842	0.02489	0.026421	0.028451	0.025923	0.028264	0.030762	0.03206	0.033492	0.036683	0.041268	0.039276	0.050634

第一行header为聚类k，第二行为gap值，第三行为se值。

result_op_sc: 为计算得到的最佳聚类结果。数据结构同谱聚类计算后得到的结果文件“result_sc”。

最佳聚类数op_k选取标准为：寻找大于“最大gap值减去se值”的所有“gap值减去se值”的k，再搜索结果中最小的k即为最佳聚类数op_k。op_sc即为k=op_k时的聚类结果。

result_labels: 为 $k=1,2,3,\dots,n$ 时的所有聚类结果。示例中的 $n=15$ ，数据结构如下：

label_1	label_2	label_3	label_4	label_5	label_6	label_7	label_8	label_9	label_10	label_11	label_12	label_13	label_14	label_15
1	2	1	2	5	2	3	3	8	1	5	10	5	13	3
1	1	2	1	3	1	7	1	9	9	3	11	12	2	11
1	2	1	2	4	2	1	5	4	3	4	1	3	1	7
1	1	2	4	5	1	2	2	3	8	6	3	1	9	4
1	2	1	2	4	2	1	5	4	3	4	1	3	1	7
1	2	3	3	2	6	6	4	5	4	7	6	9	14	9
1	1	2	1	3	1	7	1	2	2	11	5	4	12	2
1	1	2	1	3	3	4	8	6	7	9	7	13	6	6
1	2	3	3	2	6	6	4	5	4	7	6	9	14	9
1	2	1	2	5	2	3	3	8	10	5	8	10	13	10
1	1	2	4	1	4	2	7	1	5	2	2	11	10	13
1	1	2	1	3	3	4	8	6	7	9	7	13	6	6
1	2	3	3	2	6	6	4	5	4	7	6	9	14	9

5.1.5 绘图操作

本程序提供了多种输出图件供用户选择，同时提供了多种输出图片格式。Matplotlib 支持 PNG, SVG, SVGZ, PDF, PS, EPS 等栅格或矢量图格式，生成时请在“图片名”一栏直接填入图片名，并加上想要的格式后缀，可直接生成想要的图片类型，并保存在 img 文件夹中。如果图片名为空，软件默认生成以“result_”开头的 png 文件。



点击下拉框，选择需要输出的图件类型。 n 值只有“ $k=n$ 时谱聚类结果三维分布图”和“震源机制平均解图”绘制过程中需要填入，其它图件生成可以省略。“显示图例”点击打勾后会生成图例，不点击将不会显示图例。完成选择后，点击开始绘图，在“数据显示窗口”会自动展示出所需要的图片。点击“清除绘图”按钮，将会清除展示的图片。

绘图调用的是用户生成的 csv 文件。如果用户未保存相关数据，需在前述的功能操作中进行相应的计算后，才能使用绘图功能。

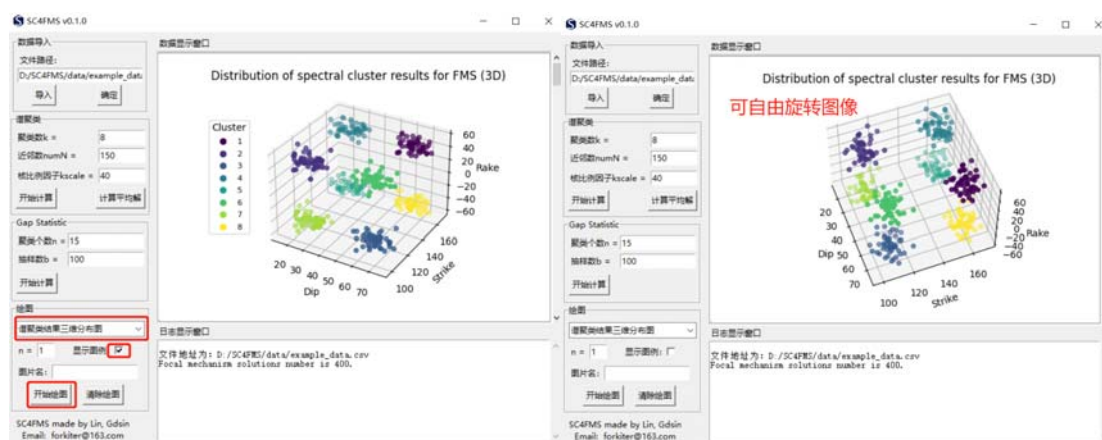
本程序提供输出的图件有 5 种类型：谱聚类结果三维分布图，最佳谱聚类结果三维分布图， $k=n$ 时谱聚类结果三维分布图，Gap-Statistic 曲线图，震源机

制平均解图。分别介绍如下：

(1) 谱聚类结果三维分布图

选择该绘图类型可显示“5.1.2 谱聚类计算”的三维分布结果，横坐标为走向和倾角，纵坐标为滑动角。由于震源机制解数据的多维度特点，该数据展示并不能给出相似性的结果，且不能作为聚类优劣的证据，但仍可以从该三维图中窥见一二。

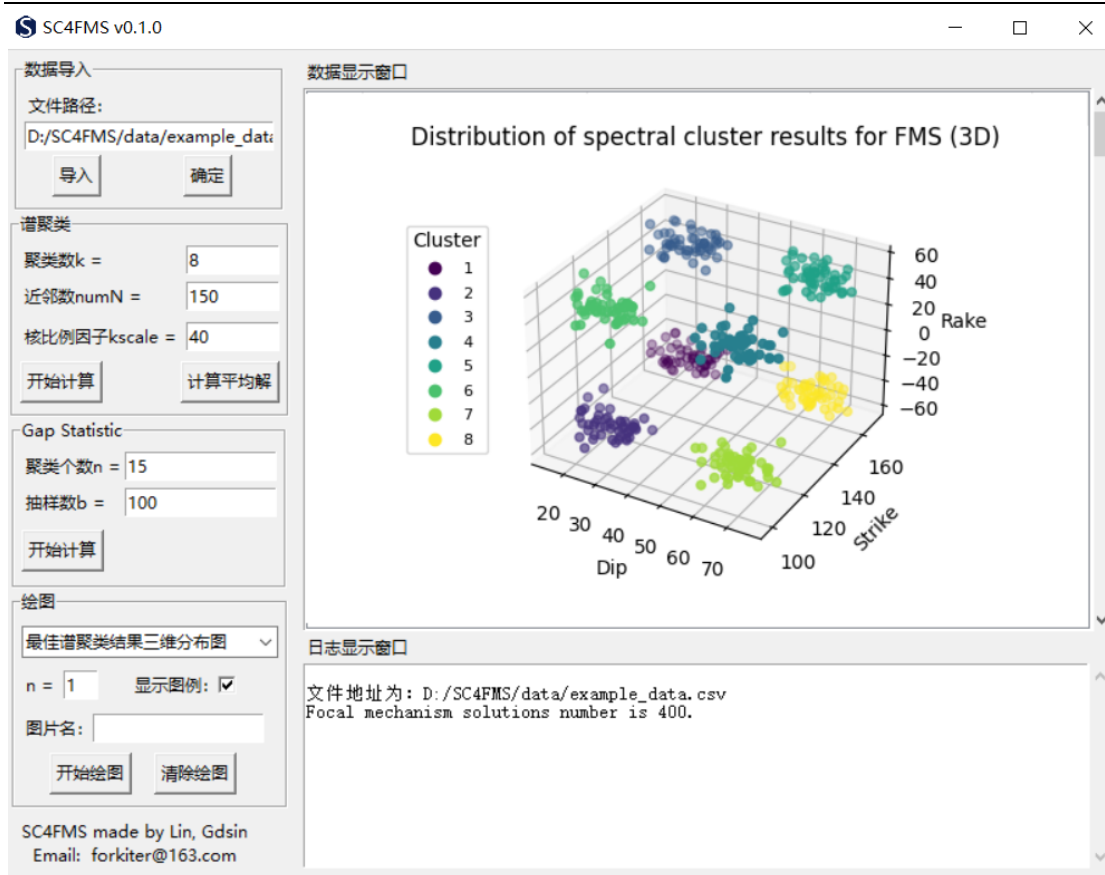
鼠标箭头放置在图片上，左键按住并滑动，可对图片进行三维旋转，可从各个角度观察聚类情况。



(2) 最佳谱聚类结果三维分布图

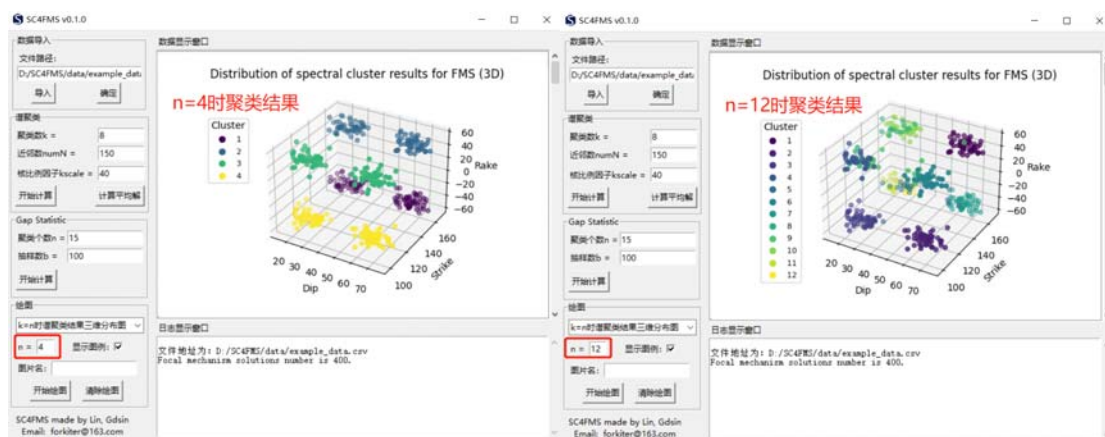
选择该绘图类型可显示“5.1.4 Gap-Statistics 计算”中的最优聚类三维分布结果。示例中的最优 op_k 值为 8，因此该图显示的为分为 8 类的结果。

因为谱聚类方法采用的是随机选择质心的方法，因此每次聚类结果对应的 label 值会发生变动，但聚类结果基本相似。下图中的聚类 label 和上图中的 label 有明显不同，但聚类效果基本相同。属正常现象。



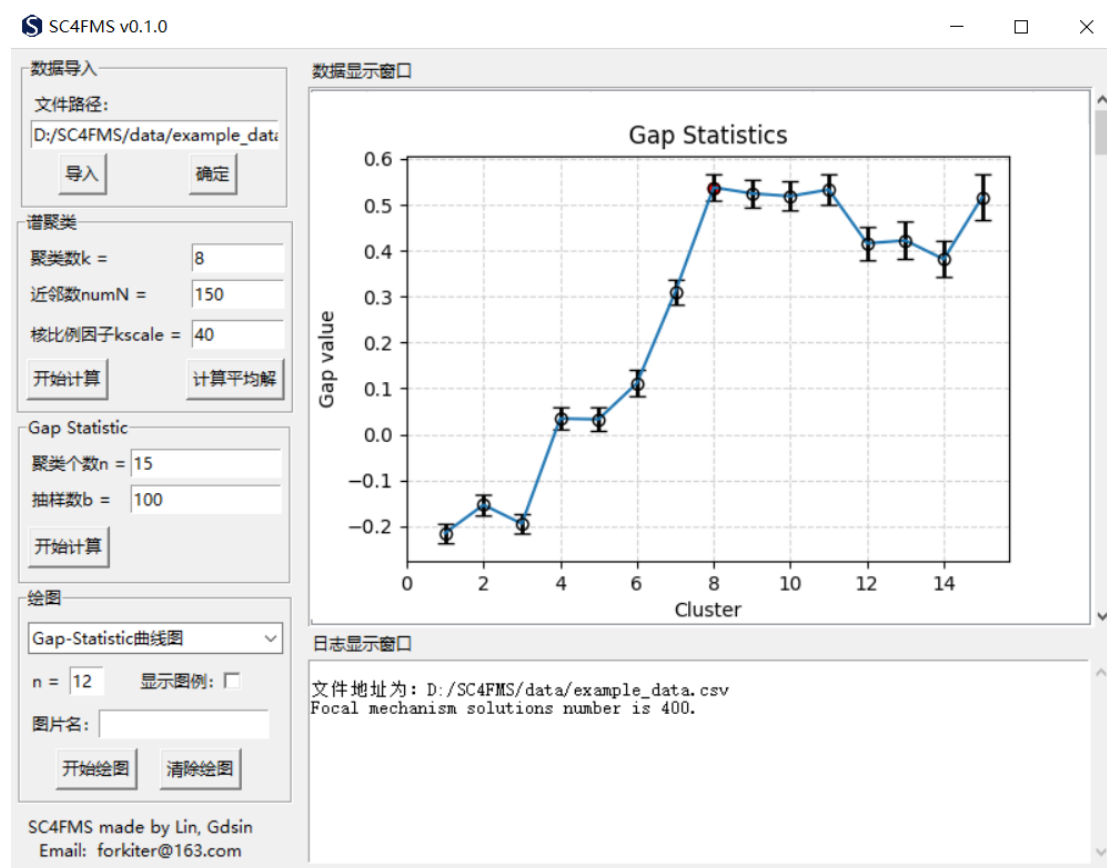
(3) $k=n$ 时谱聚类结果三维分布图

“5.1.4 Gap-Statistics 计算”完成后，会给出所有聚类结果的 label 值，对应不同的 $k=n$ 时，可显示不同的聚类结果。绘制此图像需在 n 中填入整数值，该值不能大于“5.1.4 Gap-Statistics 计算”中输入的“聚类个数 n ”值，否则程序会给出相应的错误警告。



(4) Gap-Statistic 曲线图

选择该绘图类型可显示“5.1.4 Gap-Statistics 计算”中的 gap 值和 se 值结果。横坐标为聚类数，纵坐标为 gap 值，误差条为 se 的标准误。最优 k 解会给出红色标记。该图可以帮助用户直观获取最优的聚类数目，以便对数据集选取合适的聚类数 k。

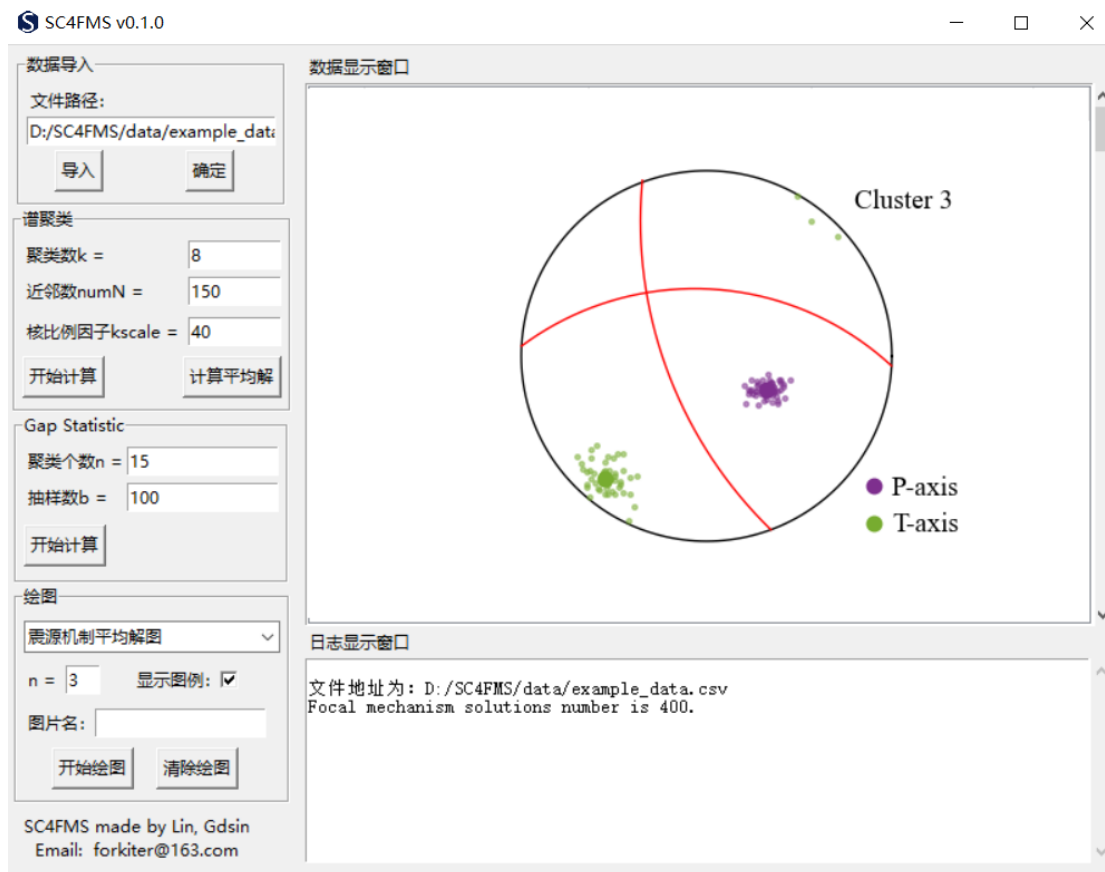


从图中可以看到，示例的最优解为 $k=8$ ，以红色标记。用户也可对自己的数据进行分析，尝试不同的聚类结果，结合构造背景等，给出相应的地球物理方面的解释。

(5) 震源机制平均解图

选择该绘图类型可显示“5.1.3 平均解计算”后的平均震源机制解图。绘制此图像需在 n 中填入整数值，该值不能大于“5.1.2 谱聚类计算”中输入的“聚类数 k”值，否则程序会给出相应的错误警告。

该图会绘制出平均解的两个节面和 P、T 轴，同时会绘制出该类所有震源机制数据的 P、T 轴。效果如下图所示：



5.2 命令行操作

该软件可通过 python 语言对数据进行批量处理，主要操作命令如下：

```
from core.sc4fms import ScFms

''' 创建实例 '''
sc = ScFms('data/example_data.csv')
print(sc.data) # 查看实例是否创建成功，查看 data 属性

''' 谱聚类 '''
sc.spectral_cluster_com(k=8) # 谱聚类计算
print(sc.labels) # 查看谱聚类结果
print(sc.result_sc)

''' 计算平均解 '''
sc.pt_average() # 平均解计算
print(sc.result_average) # 查看平均解结果
```



```
''' Gap-statistics (计算量大, 请选择合适时间运行) '''
sc.gap_eva(n_inspect=15, b_num=100) # Gap-statistics 计算
print(sc.gap_values) # 查看 Gap-statistics 计算结果
print(sc.se)
print(sc.op_k)

''' 绘制 3d 分布图 '''
sc.plot_3d(sc.user_labels('sc'), show_legend=1) # 绘制谱聚类结果 3d 图
sc.plot_3d(sc.user_labels('user', n=4)) # 绘制用户所需的聚类结果图

''' 绘制 gap 曲线图 '''
sc.plot_gap()

''' 绘制平均解图 '''
sc.plot_average(n=6, show_legend=1)
```

用户可通过主程序目录下的 example.py 进行测试运行。

6. 说明

震源机制谱聚类计算软件由作者本人独立开发完成, 目前版本号为 V1.0, 在测试过程中运行良好, 但不排除使用过程中出现 BUG 等问题, 如发现相关问题, 请联系作者本人, 联系方式如下:

林庆西, Email: forkiter@163.com

欢迎相关开发经验者和作者继续合作开发完善。在使用本软件时, 可引用作者发表的相关文献。

林庆西, 姜喜姣. 2023. 震源机制解谱聚类方法的研究及应用. 中国地震, 2023, 39(1):64~77.

本软件中震源机制解距离计算方法采用的是 Kagan 三维旋转方法^[1]; 谱聚类算法采用的是 Ng 的算法^[2]; 谱聚类中的 Ncut 算法采用的是 Shi 和 Malik 的算法^[3]; 谱聚类中的 kmeans++ 算法采用的是 Lloyd 的算法^[4]; 震源机制平均解的算法采用的是刁桂苓的 P 轴和 T 轴和矢量算法^[5]。在此对以上作者表示感谢。

参考文献

- [1] Kagan Y Y. 1991. 3-D rotation of double-couple earthquake sources. *Geophys J Int*, 106(3): 709~716.
- [2] Ng A Y, Jordan M I, Weiss Y. 2002. On spectral clustering: analysis and an algorithm. In: *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. Vancouver: ACM, 849~856.
- [3] Shi J and Malik J M. 2000. Normalized Cuts and Image Segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [4] Lloyd S P. 1982. Least squares quantization in PCM[J]. *IEEE Trans*, 28(2): 129~137.
- [5] 刁桂苓, 于利民, 李钦祖. 1992. 震源机制解的系统聚类分析——以海城地震序列为例. *中国地震*, 8(3): 86~92.