

# DisperNet(py) Readme

---

## 安装环境

---

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

## Examples

---

DisperNet(py) 内置了提供全自动拾取、半自动拾取与手动拾取多功能的应用程序，于使用前可通过调用 `dispernet.save2h5()` 函数将需要拾取的频散谱矩阵转换保存为 `.h5` 文件供应用程序使用。

```
save2h5(spectrum, frequency_range, velocity_range, fileName='')
```

(详见函数说明)。

Example包含两个文件：`Example-App.py` 与 `Example-Code.py`，分别对应使用App配合GUI进行频散曲线自动/半自动/全手动拾取的例子与使用代码对频散曲线进行拾取的例子。

## 引入

---

将 `dispernet.py` 文件添加到工作目录后

```
import dispernet
```

## 函数说明

---

以下说明中, "[ ]"中为可选参数

函数 `dispernet.extract(spec,[threshold, freq, velo, net, mode, leapLimit, freqLimits, freqSeries,errorbar, flipUp,`

# searchStep, searchBorder, returnSpec, ind, url])

通过输入的二维频散谱矩阵 `spec` 拾取其中的频散曲线，并进行自动模式分离。

## `spec` - 频散谱 - 必要参数

`spec` 必须为二维矩阵，分辨率不限，频率/速度的递增方向为：

$$\begin{bmatrix} f_{\min} c_{\min} & \cdots & f_{\max} \\ \vdots & \ddots & \vdots \\ c_{\max} & \cdots & f_{\max} c_{\max} \end{bmatrix}$$

也可使用 `flipUp` 可选参数进行上下翻转。

## `threshold` - 拾取阈值 - 可选参数

拾取阈值应为  $0 \sim 1$  之间的浮点数，阈值高低与拾取灵敏度呈反比，即阈值越低拾取敏感度越高，阈值越高，拾取敏感度越低。默认值为 `0.5`。

## `freq` - 频率范围 - 可选参数

给定拾取的频散谱包含的频率范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[0.0,0.3]`

## `velo` - 相速度范围 - 可选参数

给定拾取的频散谱包含的相速度范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[2000, 6000]`

## `net` - 使用网络名 - 可选参数

DisperNet 提供使用不同数据训练的神经网络进行拾取的选项，使用网络名称时要求数据格式为字符串；默认为 `noise`；可选参数有：

1. `noise`: 背景噪声数据，来源：吴高雄 & 詹望
2. `event`: 地震事件数据，来源：李正波
3. `noise2`: Long Beach 浅层背景噪声数据，来源：傅磊
4. `noise3`:
5. `toLB`: 由 `noise` 网络迁移到 Long Beach 浅层背景噪声数据
6. `toLB2`: 由 `toLB` 网络迁移到由 `ccf` 提取的 Long Beach 数据
7. `hyper`:

## `mode` - 指定频散曲线分阶数 - 可选参数

通常情况下DisperNet无需手动给定频散曲线阶数。也可手动对DisperNet的频散曲线自动分阶部分的阶数进行修正。`mode` 为分类数。如需将频散曲线基阶与一阶分离，则可指定 `mode=2` 。

默认不指定。

### `leapLimit` - 指定检测频散曲线分阶准确性控制参数 - 可选参数

范围在0~1之间，默认为0.1，值越高着对频散曲线分阶中存在的跳跃（可能由分阶错误造成）容忍度越高，值越低则如果单阶频散曲线存在明显跳变则会自动进行修正。

### `freqLimits` - 拾取的频率上下限 - 可选参数

可以通过该参数对频散曲线拾取的频率范围进行限定，例如需要拾取0.1~0.4Hz频率范围内频散曲线，可设定 `freqLimits=[0.1, 0.4]` 。默认不限定。

### `freqSeries` - 对频散曲线根据输入的频率序列进行插值 - 可选参数

可以通过该参数对拾取到的频散曲线进行插值，插值为指定频率序列上的点。例如需要插值到0.1~0.4Hz 以0.05Hz为步长的频率序列，可指定 `freqSeries=np.arange(0.1, 0.4, 0.05)` 。默认为不插值。

### `errorbar` - 是否返回误差棒 - 可选参数

如需返回误差棒数据，则可以设定该选项为 `True` ，默认为 `False` 。

如返回误差棒，则返回的频散曲线数组为 频率-相速度-相速度误差上限-相速度误差下限的四列列表。

### `flipUp` - 是否翻转频散谱 - 可选参数

如输入频散谱的拾取出现上下颠倒的情况，大概率可能是输入频散谱的数据排列方向与DisperNet 不符，可通过将该选项设为 `True` 对频散谱矩阵进行上下颠倒后进行拾取。如果出现旋转，可配合矩阵转制（`.T`）对频散谱矩阵进行处理。该选项默认为 `False` 。

### `searchStep` - 搜索步长 - 可选参数

可通过该参数设定频散曲线拾取的步长，因DisperNet的默认输出分辨率  $512 \times 512$ ，实际步长为  $\text{searchStep} \cdot (f_{\max} - f_{\min})/512$ ，可通过  $f_{\text{Step}}/(f_{\max} - f_{\min}) \cdot 512$  由频率步长 $f_{\text{step}}$  计算搜索补偿步长。

该参数要求为整数输入，默认为 10。

需要注意的是，搜索步长有时也会影响到后面的频散曲线分阶，故建议使用较小的搜索步长以得到更好的频散曲线分阶结果。

## searchBorder - 边缘省略宽度 - 可选参数

在实际数据测试中，因输入频散谱数据类型与分辨率的不同，已知有一定概率会在边缘出现错误拾取，可使用该选项对边缘进行截断并不在该区域内拾取频散曲线。该参数为  $0 \sim 0.5$  之间的浮点数，按忽略宽度占全频散谱宽度的比例进行计算，最大为0.5，对上下及右侧的频散谱拾取生效。默认值为 0。

## searchSpec - 返回拾取到的频散谱 - 可选参数

该选项设为 `True` 时，上文仅 `spec` 参数生效，输出为DisperNet的原始输出概率矩阵，为  $512 \times 512$  的二维矩阵，可用作程序调试与频散谱降噪使用。默认为 `False`

## ind - 任务ID - 可选参数

在进行并行任务时，可将改选项设定为并行编号 `rank` 可避免在并行任务中出现文件冲突而导致拾取错误的情况。默认不设定。

## url - 指定 DisperNet 服务器位置

供调试程序用。

---

## 函数 save2h5(spectrum, freq, velo, [fileName]):

将频散谱与频率、速度范围保存到指定的文件。

### spectrum - 频散谱 - 必要参数

频散谱（能量图）矩阵，`np.array`或`list`格式。

### freq - 频率范围 - 必要参数

给定拾取的频散谱包含的频率范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[0.0,0.3]`

### velo - 相速度范围 - 必要参数

给定拾取的频散谱包含的相速度范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[2000, 6000]`

### fileName - 文件路径/文件名 - 可选参数

可以通过该参数指定保存到不同文件，文件路径支持绝对路径与相对路径。用作不同的多张频散谱时的区分，如不使用该参数，则保存为 `demoSpectra.h5` 文件。

## 函数 `readh5(fileName)`:

读取 `save2h5` 函数保存的频散谱文件。

`fileName` - 文件名 - 必要参数

读取该文件后返回三个np矩阵/数组：

`spec`, `freq`, `velo` 分别对应频散谱、频率范围、和速度范围。

## 类初始化 `App([filePath, curveFilePath, freqSeries, cmap, vmin, vmax, url])`:

通过调用该类初始化函数 `App()` 可直接启动 DiperNet App。

`filePath` - 频散谱文件目录 - 可选参数

默认为当前目录，也可另外指定。

`curveFilePath` - 频散曲线保存位置 - 可选参数

默认为当前目录，也可另外指定。

`freqSeries` - 频散曲线插值序列 - 可选参数

可以通过该参数对拾取到的频散曲线进行线性插值，插值为指定频率序列上的点。例如需要插值到 0.1~0.4Hz 以0.05Hz为步长的频率序列，可指定

`freqSeries=np.arange(0.1, 0.4, 0.05)`。

默认为不插值，可通过 `np.arrange()` 函数或 `np.linspace()` 函数对频率序列进行指定，以匹配反演所需的频率数值。

`cmap` - 绘制频散谱时使用的colormap - 可选参数

选择不同的colormap在App中绘制频散谱，可采用其他常用的colormap，比如 `jet` 与

`seismic` 注意格式要求字符串。默认为 `viridis`，可选 colormap可见参考: [matplotlib: color map](#)

`vmin` - 色表下限 - 可选参数

在绘制频散谱时设定值下限参数，默认为不设定。

### vmax - 色表上限 - 可选参数

在绘制频散谱时设定值上限参数，默认为不设定。

### url - 指定 DisperNet 服务器位置

供调试程序用。

## 函数 `dispernet.pick(spec,[threshold, freq, velo, net, errorbar, flipUp, searchStep, searchBorder,returnSpec, ind, url])`

通过输入的二维频散谱矩阵 `spec` 拾取其中的频散曲线，拾取到的频散曲线格式为 频率-相速度的两列列表。

### spec - 频散谱 - 必要参数

`spec` 必须为二维矩阵，分辨率不限，频率/速度的递增方向为：

$$\begin{bmatrix} f_{\min} c_{\min} & \cdots & f_{\max} \\ \vdots & \ddots & \vdots \\ c_{\max} & \cdots & f_{\max} c_{\max} \end{bmatrix}$$

也可使用 `flipUp` 可选参数进行上下翻转。

### threshold - 拾取阈值 - 可选参数

拾取阈值应为  $0 \sim 1$  之间的浮点数，阈值高低与拾取灵敏度呈反比，即阈值越低拾取敏感度越高，阈值越高，拾取敏感度越低。默认值为 `0.5`。

### freq - 频率范围 - 可选参数

给定拾取的频散谱包含的频率范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[0.0,0.3]`

### velo - 相速度范围 - 可选参数

给定拾取的频散谱包含的相速度范围，顺序任意，长度应大于2，即应当是一个范围。  
默认值为 `[2000, 6000]`

### net - 使用网络名 - 可选参数

DisperNet 提供使用不同数据训练的神经网络进行拾取的选项，使用网络名称时要求数据格式为字符串；可选参数有：

1. noise: 背景噪声数据，来源：吴高雄 & 詹望
2. event: 地震事件数据，来源：李正波
3. noise2: Long Beach 浅层背景噪声数据，来源：傅磊
4. noise3:
5. toLB: 由 noise 网络迁移到 Long Beach 浅层背景噪声数据

默认为 noise

### errorbar - 是否返回误差棒 - 可选参数

如需返回误差棒数据，则可以设定该选项为 True，默认为 False。

如返回误差棒，则返回的频散曲线数组为 频率-相速度-相速度误差上限-相速度误差下限的四列列表。

### flipUp - 是否翻转频散谱 - 可选参数

如输入频散谱的拾取出现上下颠倒的情况，大概率可能是输入频散谱的数据排列方向与 DisperNet 不符，可通过将该选项设为 True 对频散谱矩阵进行上下颠倒后进行拾取。如果出现旋转，可配合矩阵转制（.T）对频散谱矩阵进行处理。该选项默认为 False

### searchStep - 搜索步长 - 可选参数

可通过该参数设定频散曲线拾取的步长，因DisperNet的默认输出分辨率  $512 \times 512$ ，实际步长为  $\text{searchStep} \cdot (f_{\max} - f_{\min})/512$ ，可通过  $f_{\text{Step}}/(f_{\max} - f_{\min}) \cdot 512$  由频率步长  $f_{\text{step}}$  计算搜索补偿步长。

该参数要求为整数输入，默认为 10。

需要注意的是，搜索步长有时也会影响到后面的频散曲线分阶，故建议使用较小的搜索步长以得到更好的频散曲线分阶结果。

### searchBorder - 边缘省略宽度 - 可选参数

在实际数据测试中，因输入频散谱数据类型与分辨率的不同，已知有一定概率会在边缘出现错误拾取，可使用该选项对边缘进行截断并不在该区域内拾取频散曲线。该参数为 0 ~ 0.5之间的浮点数，按忽略宽度占全频散谱宽度的比例进行计算，最大为0.5，对上下及右侧的频散谱拾取生效。默认值为 0。

### searchSpec - 返回拾取到的频散谱 - 可选参数

该选项设为 `True` 时，上文仅 `spec` 参数生效，输出为DisperNet的原始输出概率矩阵，为  $512 \times 512$  的二维矩阵，可用作程序调试与频散谱降噪使用。默认为 `False`

### `ind` - 任务ID - 可选参数

在进行并行任务时，可将改选项设定为并行编号 `rank` 可避免在并行任务中出现文件冲突而导致拾取错误的情况。默认不设定。

### `url` - 指定 DisperNet 服务器位置

供调试程序用。

## 函数 `dispernet.modeSeparation(curves, [modes])`

输入拾取到的频散曲线，并通过聚类分析将频散曲线分离到不同阶。返回值为频散曲线列表，为输入参数 `curve` 增加一列数据标记当前频散曲线点所在的阶数。

### `curves` - 频散曲线 - 必要参数

即前文 `pick` 函数返回值的格式，支持频率-相速度两列格式与频率-相速度-相速度误差上限-相速度误差下限的四列格式。

### `modes` - 阶数 - 可选参数

需要分离的阶数，默认为 2 阶，即基阶与一阶频散曲线进行分离。

## 函数 `dispernet.autoSeparation(curves, [to])`

输入拾取到的频散曲线，并通过聚类分析将频散曲线分离到不同阶。返回值为频散曲线列表，为输入参数 `curve` 增加一列数据标记当前频散曲线点所在的阶数。

### `curves` - 频散曲线 - 必要参数

即前文 `pick` 函数返回值的格式，支持频率-相速度两列格式与频率-相速度-相速度误差上限-相速度误差下限的四列格式。

### `to` - 灵敏度 - 可选参数

范围在0~1之间，默认为0.1，值越高着对频散曲线分阶中存在的跳跃（可能由分阶错误造成）容忍度越高，值越低则如果单阶频散曲线存在明显跳变则会自动进行修正。



# 函数 `dispernet.show(spec, curve, [freq, velo, unit, s, ax, holdon, cmap, vmin, vmax])`:

通过该函数可快速绘制频散谱与频散曲线，以检查拾取结果。

## `spec` - 输入频散谱 - 必要参数

要求与 `pick` 函数相同

## `curve` - 频散曲线 - 必要参数

要求与 `modeSeraration` 函数相同，可以是分阶后的频散曲线，也可不分阶。如不需要绘制，可以给定 `[]`。

## `freq` - 频率范围 - 可选参数

与前文一致。

## `velo` - 相速度范围 - 可选参数

与前文一致。

## `unit` - 相速度单位 - 可选参数

相速度单位，默认为 `m/s`，需要时也可设定为 `km/s`。

## `s` - 频散曲线绘图点大小 - 可选参数

可以通过该选项改变绘制的频散曲线点大小，与 `plt.scatter` 中的 `s` 选项相对应，默认为 10。

## `ax` - 绘制位置 - 可选参数

可以通过指定 `ax` 的方式将频散曲线与频散谱绘制到指定位置，默认为当前`ax`，即 `plt.gca()`

## `holdon` - 是否等待 - 可选参数

提供类似于matlab 中的 `hold on` 功能，设定为 `True` 后可进一步进行绘图，在之后再进行 `plt.show()`。默认为 `False` 即直接展示。

## `cmap` - 绘制频散谱时使用的`colormap` - 可选参数

选择不同的colormap绘制频散谱，可采用其他常用的colormap，比如 `jet` 与 `seismic` 注意格式要求字符串。默认为 `viridis`，可选 colormap可见参考: [matplotlib: colormap](#)

### `vmin` - 色表下限 - 可选参数

在绘制频散谱时设定值下限参数，默认为不设定。

### `vmax` - 色表上限 - 可选参数

在绘制频散谱时设定值上限参数，默认为不设定。

## 函数 `curveInterp(curve, [freqSeries])`

对拾取到的频散曲线进行插值，以匹配反演程序的需要。

### `curve` - 频散曲线 - 必要参数

拾取到的频散曲线，且必须已经分阶。未分阶的频散曲线无法进行插值，如果该频散曲线仅含有基阶，可使用 `modeSeparation(curve, 1)` 为基阶频散曲线添加阶标记。该函数支持2列仅含有频散曲线与4列带有误差区间的频散曲线同时对误差上下限进行插值。

### `freqSeries` - 频率插值序列 - 可选参数

可通过 `np.arange()` 函数或 `np.linspace()` 函数对频率序列进行指定，以匹配反演所需的频率数值，默认为0到10Hz，步长0.1Hz。