



---

# 数据预处理

## 从RefTek130数采到SAC

---

作者：汪晟

版本：0.01

July 7, 2015

# Contents

<b>1</b>	<b>Ref格式数据的转换</b>	<b>2</b>
<b>2</b>	<b>SAC格式头段信息简介</b>	<b>3</b>
<b>3</b>	<b>补全SAC文件头段台站信息变量</b>	<b>4</b>
<b>4</b>	<b>补全SAC文件头段事件信息变量</b>	<b>5</b>
<b>5</b>	<b>时间序列预处理</b>	<b>6</b>
<b>6</b>	<b>合成理论接受函数</b>	<b>8</b>
<b>7</b>	<b>反演</b>	<b>10</b>
<b>8</b>	<b>基于KMI台的实验</b>	<b>12</b>
8.1	数据下载 . . . . .	12
8.2	预处理数据 . . . . .	12
8.3	提取接收函数 . . . . .	12
8.4	反演 . . . . .	13

## 1 Ref格式数据的转换

野外操作中，自130数采拷贝的原始数据是Ref格式，其单文件包含了多分量的数字记录以及部分台站与时间信息，可以使用PQL查看。此外，RefTek130数采亦生成log文件，实时记录了台站的状态信息，包括有：温度、GPS卫星数、电压等，可以使用newlogpeek.py查看。这些不同功能的软件均由PASSCAL () 提供。

根据需求，可以将Ref格式文件转换成其他格式的文件，相关程序均由PASSCAL提供。本文档依托项目为腾冲火山地区接受函数研究，需要数据转换成SAC格式。由Ref到SAC，需要借助中间文件格式，一般有两种方法较为常用：

Ref → miniSeed → SAC

Ref → Segy → SAC

本文档使用第二种方法，其命令行操作如下：

```
bash> ref2segy -f in.ref -l cas.file
bash> segy2sac in.ref
```

其中，ref2segy将Ref格式转换为Segy格式。-f参数声明Ref文件名或目录。-l参数声明130数采部分配置信息。一般情况下，原始Ref文件携有全部的数采配置信息，但是BUG时有发生，此时-l参数开启就很有必要。关于-l的详细介绍以及对应的130数采文件请参看ref2segy的manual，在本文档的后续章节中，亦有提及，可供参考。

转换得到的SAC格式文件头信息缺失，这里引出下一节。

## 2 SAC格式头段信息简介

除了时间序列记录之外，地震数据中的台站、事件等信息十分重要。本节承接上一节文件格式转换步骤，将简介SAC格式文件中的头段信息变量。

SAC头段信息变量包含了所有的数据参数信息，罗列接分类如下：

1. 基本变量 nvhdr nzyear nzjday nzhour nzmin nzsec nzmsec iztype iftype idep
2. 数据相关变量 npts delta odelta b e leven depmin depmax depmen scale xminimum xmaximum yminimum ymaximum nxsize nysize ical isynth
3. 事件相关变量 kevnv evla evlo evel evdp ievreg ievtyp mag imagsrc imagtyp gcarc dist az bz o ko khole
4. 台站相关变量 kstnm knetwk istreg stla stlo stel stdp cmpaz cmpinc kcmpnm kstcmp lpmpol
5. 震相相关变量 a ka f kf tn kn
6. 仪器相关变量 kinst iinst respn
7. 其他变量 usern kusern lovrok lcalda kdatrd

以上罗列中，×××变量为转换生产的SAC文件携有的有效头段信息。亦存在其他变量有值，但非有效，所以需要补全。

第一步需要补全的是台站相关变量。对任一数据记录，台站信息都是存在的且必须的。并非只有事件相关数据才可以用于研究，噪声数据的价值亦不可估量。

第二步需要补全的是事件相关变量。此操作只针对那些包含了时间记录的文件，所以需要先抽取出这些文件，再补全头信息。

第三步操作，需要补全部分震相等其他信息。

需要注意的是SAC头段中的时间类变量。由Ref转换得到的SAC文件携有的有效信息中有“参考时间”、“起始相对时间”、“结束相对时间”、“采样间隔”、“采样点数”。显然，这些时间信息倘若缺失，那么整个预处理都将失去意义。转换得到的SAC文件“参考时间”为数据首个采样点的绝对时间（伦敦时间），那么“起始相对时间”自然为0。在更为靠后的操作中，需要写入事件信息，并截断有效长度，此时“参考时间”会发生改变，相对类时间将会由SAC软件自动计算变化。总之，时间类变量的写入与更改，应当细心。

此外，SAC头段信息变量冗余，一些变量可由其他变量推导计算获得，这一特性减轻了补充头段信息变量的工作。

### 3 补全SAC文件头段台站信息变量

上一节提及了对所有的文件记录，台站信息都是存在且必需的。本节将描述补全这一类信息。

SAC程序提供了更改任意头段信息变量的命令，如下：

```
SAC> r eg1.sacS
SAC> ch ARGV1 VAL1 ARGV2 VAL2 ...
SAC> w over
```

第一行读SAC文件到内存中，第二行更改了内存中数据的头段变量值，第三行将内存中数据覆盖写入文件。本文档所有的SAC头段变量信息的赋值与更改均为以上操作。

本节中赋值更改的ARG范围列表如下： 其中，通道名与通道角度之间的关系为：

Table 1: 台站信息列表

变量名	解释	是否需要赋值
kstnm	台站名	是
knetwk	台网名	是
istreg	台站地理区域	
stla	台纬度	是
stlo	台经度	是
stel	台高度	是
stdp	台深度	是
cmpaz	通道方位角	是
cmpinc	通道倾角	是
kmpnm	通道名	是
kstcmp	辅助名	
lpspol	通道极性正负	是

Table 2: 通道信息表

通道名	cmpaz	cmpinc
N	0	90
E	90	90
U	0	0

显然NEU为左手坐标系，lpspol变量为TRUE表明通道的正极性与NEU方向相同，为FALSE则相反。

## 4 补全SAC文件头段事件信息变量

补全所有记录文件的台站信息之后，可挑选出包含有时间记录的文件。此时，当务之急是如何挑选出这类文件，并为这些文件补全事件类信息。

一种简单且可批量处理的方法是，给出台站与时间信息列表，组合成不同的“事件-台站”对，计算理论P波初至震相到时。由这些“事件-台站-到时”组合，挑选出包含了P波初至震相的记录文件。更为复杂的情况是，需要提取出P波初至震相前后一段时长内的记录。那么这些时间记录片可能会跨不同文件，甚至缺失某些时段的记录文件。对此，可以事先给出P波初至震相前后时间偏移长度，计算出有效时间片的“起始绝对时间”和“终止绝对时间”，随后提取出所有与此时间区间存在交集的记录文件，合并即可。此步得到的“起始绝对时间”和“终止绝对时间”亦将用到后续的数据截断中。

获取了包含有事件记录的文件后，需要补全对应事件信息，以及理论计算的P波初至震相到时。变量列表如下：

Table 3: 事件信息列表

变量名	解释	是否需要赋值
kevnrm	事件	是
evla	事件纬度	是
evlo	事件经度	是
evel	事件高程	是
evdp	事件深度	是
ievreg	事件地理区域	
ievtyp	事件类型	选
mag	震级	是
imagsrc	震级信息来源	是
imagtyp	震级类型	是
gcarc, dist	大圆路径长	当lcalda为TRUE时，将自动计算
az,bz	前后方位角	当lcalda为TRUE时，将自动计算
t1	理论P波初至相对时间	是

## 5 时间序列预处理

补全SAC文件头段信息之后，接下来的预处理操作将针对时间序列展开。与头段信息不同的是，时间序列的更改不可逆，且不同的处理手段及先后顺序将直接影响到后续的数据研究分析。随意随性的操作将会带来假象，所以先于预处理，应有数据备份。

时间序列的预处理包括有：1.去毛刺、2去仪器相应、3.时间截断、4.坐标旋转、5.去均值与线性趋势、6.尖灭、7.滤波。简介：

以上操作中，“去仪器相应”、“去均值与线性趋势”、“尖灭”三项均从根本上改变了采样点的数值，其执行需要严格遵守一定的先后顺序。在确定先后顺序中，需要深刻理解时间-频谱关系及仪器工作原理，否则将产生不可预知的假象和错误。一般预处理中，按照“去均值与线性趋势”→“尖灭”→“去仪器相应”的先后顺序进行。而“时间截断”和“坐标旋转”可以在后续操作中任意进行。“滤波”则针对研究目的的需要而进行。将预处理完的数据应用到研究中，如果涉及到频谱操作或（反）卷积操作，则需要再做一次“去均值与线性趋势”与“尖灭”。

当然，以上的步骤与顺序介绍并非放置四海而皆准。微小的预处理差异会带来截然不同的结果，所以需要备份数据，反复操作对比，并针对研究目的确定最合适的预处理流程和参数选择。

SAC软件预处理命令范例如下：

```
SAC> r eg.sac           #读入SAC文件
SAC> rmean              #去均值
SAC> rtr                #去线性趋势
SAC> taper              #尖灭
SAC> transfer           #去仪器响应
SAC> rotate to gcp      #旋转至大圆路径，即RTZ分量
SAC> cut t1 -50 200     #截取t1时间前50s到后200s
SAC> bandpass BU CORNERS 0.01 1 N 4
      #带通滤波，选择Butterworth滤波器，拐点频率为0.01Hz 1Hz，极点数为4
SAC> lowpass BU CORNER 1 N 4
      #低通滤波，选择Butterworth滤波器，拐点频率为1Hz，极点数为4
SAC> highpass BU CORNER 1 N 4
      #高通滤波，选择Butterworth滤波器，拐点频率为1Hz，极点数为4
SAC> bandrej BU CORNERS 0.01 1 N 4
      #带阻滤波，选择Butterworth滤波器，拐点频率为0.01Hz 1Hz，极点数为4
```

SAC软件数据查看命令简介:

```
SAC> r *.ri          #读入当前目录下所有R分量到内存中
                        #每读入一次，内存就会刷新一次，内存之前读入的数据全部清除
SAC> p1              #绘制内存所有的数据
SAC> lowpass BU CORNER 1 N 2 #简单低通滤波
SAC> p1              #更新绘图
SAC> ed x            #关闭图形窗
SAC> lh default columns 2 #显示所有的SAC头段信息，两列显示
SAC> lh kstnm stla stlo kevnm evla evlo columns 2
                        #指定SAC头段变量，并两列显示
SAC> q              #退出
```



## 6 合成理论接受函数

接受函数正演是反演的基础。人工合成的结果也应用于反演，以判断一种方法的正误。在此介绍李志伟的合成理论接受函数程序。synrflzw来源于DERFmulti\_submit20090526\_v2包，在terminal中运行即可获得以下帮助：

```
bash> synrflzw
*****
*Given Model and some necessary information, compute
* synthetic receiver function waveforms in ZRT coordinate.
*
**** Help information *****
*synrflzw MOD= FM= BAZ= RP= O=
*
*MOD = Model
*FM  = Frequency Maximum, used for forward calculation.
*BAZ = Back Azimuth, (in degree) NOT work for isotropic
*      inversion.
*RP  = Ray Parameter(Horizontal Slowness ), (in s/km.)
*O   = Out put receiver function waveforms,
*      (H T L - components )
*
*****
bash>
```

例子如下：

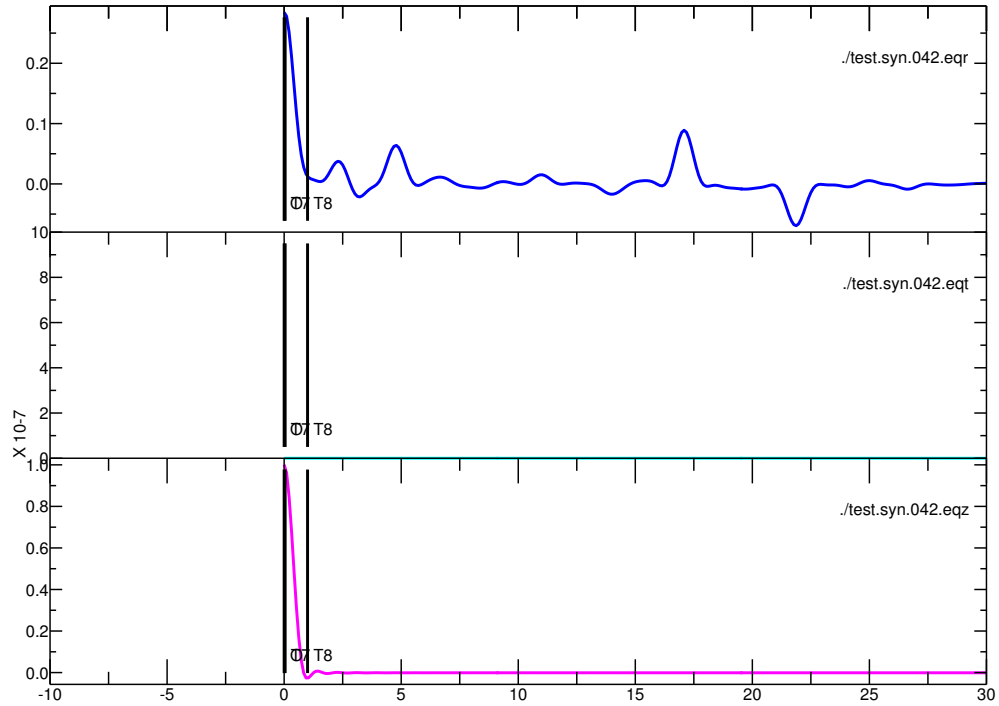
```
bash> synrflzw MOD=essmod.syn FM=1.2 BAZ=0.0 RP=0.042 \
O=./rfr/test.syn.042
```

其中，essmod.syn描述了理论模型，其内容如下所示。文件的规则是：

```
syn
Theta Phig
DepthofLayerBottom,Vp(km/s),B,C,Vp/Vs,E,density(g/cm^3),NSL
4
0 0
5.0      5.50      0.0      0.00      1.70      0.0      2.665      0
0 0
15.0     6.20      0.0      0.00      1.73      0.0      2.782      0
0 0
25.0     6.30      0.0      0.00      1.73      0.0      2.811      0
0 0
40.0     6.60      0.0      0.00      1.75      0.0      2.897      0
0 0
100.0    8.00      0.0      0.00      1.80      0.0      3.302      0
```

- 1.前3行是注释行;
  - 2.第4行是半空间上的覆盖层数, essmod.syn中是4;
  - 3.后续每2行为一层介质的参数, 第1行是各向异性的theta和fai值, 第2行列出了层厚, P波速, B, C, 波速比, E, 密度, NSL;
- FM指定了合成地震记录最大频率值, BAZ参数指定了大圆弧路径反方位角, RP参数指定了射线参数值(s/km), O指定了输出文件名。

试运行如下:



## 7 反演

这里介绍DERFmulti\_submit20090526\_v2包中的derfmod。其使用了差异演化算法()反演获得速度结构。terminal中的命令给出帮助信息:

```
bash> derfmod
*****
*Using DE method to search the best 1-D Model
*  (isotropic or anisotropic).
*Using Receiver Function waveforms in Z R T system.
*Only using R component in isotropic media.
*Using R and T components in anisotropic media.
*****
* derfmod  -Mmod -Bmod.scale -Drfs.list -Ggauss -Tbeg/end \
*          [-Sst/seed/F/CR/NP/genmax] \
*          [-Ianiso -Hrfs.listT -WRweight/Tweight ]
*
* -M mod: 1-D model used to determine the searching scales
* -B mod.scale 1-D searching scales relative to 1-D model
* -D rfs.list list of stacked RFs according ray parameter,
*       with stack number
* -G Gauss: Gauss filter for distrilling receiver functions
* -T beg/end: begin and end time relative to P onset for
*       inversion
*
* -I aniso: 0, Inverse for Isotropic Vp, Thickness,
*           and Vp/Vs [0]
*           -1, Inverse for Isotropic Vp, Thickness,
*           with fixed Vp/Vs
*           -2, Inverse for Isotropic Vp, with fixed
*           Thickness and Vp/Vs
*           1, Inverse for Anisotropic Model.
* -W weight: Rweight, weight for Radial RFs; 1>Rweight>0
*           Tweight, weight for tangential RFs;1>Tweight>0
*           Rweight + Tweight = 1.0
* -H rfs.listT list of stacked tangential RFs according ray
*           parameter, with stack number
* -S st: strategy for DE, can be 1 to 10.[3] integer
*       seed: random number seed, [5] integer
*       F: weight factor, 0<F<1, [0.85] float
*       CR: crossing over factor, 0<CR<1 [0.98] float
*       NP: population size, about 5-10 times of unknow para-
*           meters
*           [10] integer
*       genmax: maximum number of generations, [40] integer
```

例子如下:

```
bash> derfmod -Messmod -Bessmod.DEB -Drfr.list \  
-G2.5 -T0.0/32.0 -I-1 -S3/5/0.85/0.98/10/50
```

命令参数中, -M指定了搜索中心模型文件, 其文件格式与synrflzw中的essmod.syn相同; -B参数指定了一维搜索模型的范围, 其文件内容格式是:

- 1.第1行是注释行;
- 2.第2行是半空间上的覆盖层数, 这与搜索中心模型对应;
- 3.后续内容每2行指定了每一参数的搜索范围, 参数顺序是: 厚度, P波速, B, C, 波速比, E, 密度;

```
ESS forward model  
4  
0      0  
5.0    0.6    0.0    0.0    0.0    0.0    0.0  
0      0  
5.0    0.6    0.0    0.0    0.0    0.0    0.0  
0      0  
5.0    0.6    0.0    0.0    0.0    0.0    0.0  
0      0  
10.0   0.6    0.0    0.0    0.0    0.0    0.0  
0      0  
0.0    0.0    0.0    0.0    0.0    0.0    0.0
```

-D参数指定了接受函数文件名list, 本例子中, 其内容如下。其中, 第一列指定了文件名, 第二列指定了此接收函数的叠加次数。

```
rfr/test.syn.056.eqr 1
```

-G参数指定了高斯滤波参数值; -T指定了时间序列截取的起始和终止相对时间; -I指定需要反演哪些参数, 本例子中, 反演各向同性介质中的层厚与P波速, 这也与-B参数指定的模型参数搜索范围相对应; -W指定了径向与切向接收函数的权重, 仅适用于各向异性反演; -H指定了切向接收函数的文件列表和叠加次数, 其内容格式与-D参数相同, 仅适用于各向异性反演; -S参数指定了差异演化算法的策略与对应的参数值。在例子中, 选择策略3, 随机数为5, 权重因子为0.85, 差异系数为0.98, 每一代的人口数为10, 演化代数为10。

## 8 基于KMI台的实验

本小节以KMI台为例，从下载数据到接收函数提取，最后反演获得合理的速度模型。

### 8.1 数据下载

本例使用JWEED软件挑选台站与事件，下载事件文件。JWEED由IRIS提供，可以在<http://ds.iris.edu/ds/nodes>下载。此软件有JAVA开发，可交互式挑选数据并下载。

本例选择KMI(昆明台)自2013-12-31-00:00:00 2015-07-02-00:00:00，震中距30度~90度的内震级大于5.5的所有事件记录。时间截取长度选择P波出动前20s 后150s的数据。数据放在/home/wsh/JWEED.dir/Data\_KMI目录下。

### 8.2 预处理数据

使用JWEED下载的数据已经是SAC格式数据，其文件名格式是IC.KMI.00.BH1.M\_at\_2015-03-22T06.03.54.019Z.SAC，不可以直接用于X X X软件。更改后的文件格式是IC.KMI.00.M\_at\_2015-03-22T06.03.54.019Z.SAC.BH[NEZ]，其中BH1对应BHN，BH2对应BHE。根据derfmod的需求，sac文件T7头段给定射线参数值。不需要其他预处理操作。

### 8.3 提取接收函数

使用pburg\_lzw做反褶积提取接收函数，pburg\_lzw的使用格式是：

```
bash> ../src/pburg_lzw <<EOF
2008.187.02.12.04
y
30
10
80
n
2.5
10
EOF
bash>
```

其中，2008.187.02.12.04制定了文件名中除了后缀.BH[NEZ]的其他部分，y指定了地理坐标，30指定了B到A的时间差(s)，10指定了截取A之前10s，80指定截取A之后80s，n指定不Taper，2.5是高斯参数，10指定接收函数结果延后10s。

使用脚本运行如下：

```
bash> decon.sh -D KMI/Data_KMI -L KMI/KMI.list.info \
-R KMI/RFS -V
bash> #!Note:不要在-D参数后加"/"
bash>
```

此脚本运行后，会在每一层子目录下生成相应接收函数，并在工作目录下生成KMI.list.info文件，其格式如下：

```
#Counts RF-PreName      Dir-Name
1 1.KMI.2015.03.2206 : 2015-03-22T05.56.22.3700_5.5
2 2.KMI.2015.03.2822 : 2015-03-28T22.28.50.7800_5.9
3 3.KMI.2015.03.2907 : 2015-03-29T07.50.54.2900_5.6
...
```

## 8.4 反演

按照derfmod程序，要求，需要给定中心模型，及对应的搜索范围模型，以非线性搜索最合适的模型，在第7小节中给反演需要的不同的文件及其对应格式。本范例中，将由最简单的半无线空间单覆盖层开始，逐次增加模型的复杂度。

第一次反演，模型文件与模型搜索范围如下：

```
***** 1st Model KMI.1.mod *****
Theta Phig
Thickness,Vp(km/s),B,C,Vp/Vs,E,density(g/cm^3),NSL
1
0 0
40.0 6.80 0.00 0.00 1.75 0.00 2.80 0
0 0
100.0 8.00 0.00 0.00 1.80 0.00 3.30 0
```

```
***** 1st Model KMI.1.mod.deb *****
1
0 0
10.0 0.5 0.0 0.0 1.0 0.0 0.0
0 0
0.0 0.0 0.0 0.0 0.0 0.0 0.0
```