Python Matching Phase Algorithm

Alessandro Vuan Monica Sugan

Created on : April, 2019

Last updated : April, 2019

More documents are freely available at PyMPA

Table of contents

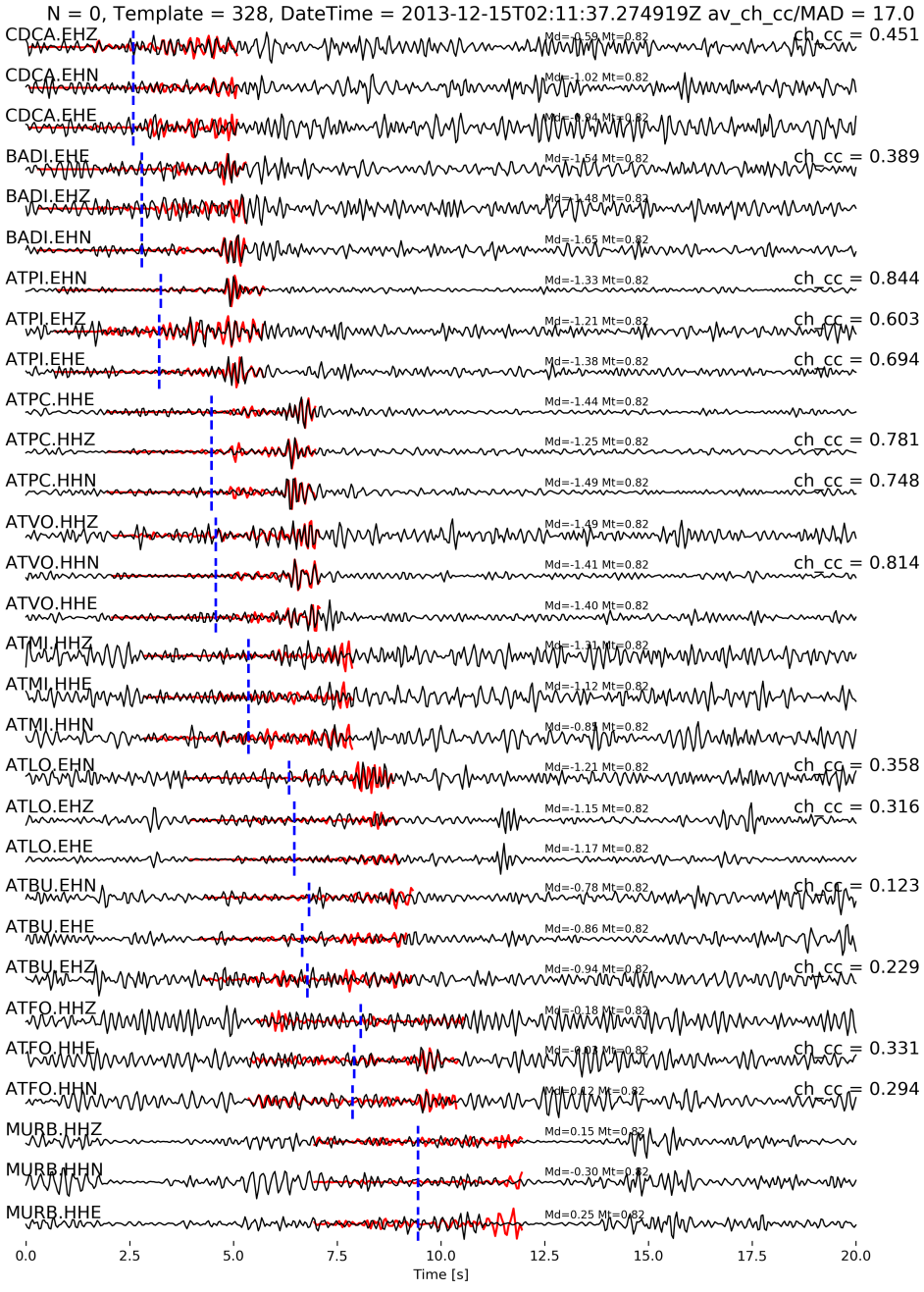
List of figures

List of tables

PyMPA is designed to detect microseismicity from the cross-correlation of continuous data and templates. PyMPA is an open source seismological software and consists of some separate utilities for input preparation, the main program, and output post-processing tools to obtain a catalogue and verify events. The code repository with some examples is stored on , and is free to be cloned on your Mac, Windows or Linux platforms. It supports Python 2.7, 3.5, 3.6 and 3.7 releases and uses ObsPy for reading and writing seismic data, and for handling most of the event-station metadata. A flowchart of the main program is shown in Figure 1. Matched-filter correlations are calculated using a vectorised python function or using ObsPy v.1.2.0 correlate_template function allowing to select time or frequency domain. Together with the manual, we also provide some practical examples. Python scripts, updates and modifications are automatically verified using TRAVIS, for testing and deployment (https://travis-ci.org/). This package is distributed under the LGPL GNU Licence, Copyright PyMPA developers 2019.

The algorithm, which exploits routines (Krischer et al., 2015), is versatile and supports most commonly used seismic data and earthquake catalogue formats. In addition to PyMPA, we develop other tools external to the main code to manage the input-output preparation and validation for (1) downloading data from Observatories and Research Facilities for European Seismology–European Integrated Data Archive (ORFEUS-EIDA) servers, (2) evaluating data quality, (3) selecting earthquakes as templates from a reference catalog, (4) trimming and filtering them from continuous waveforms, (5) avoiding redundant detections in the output, and (6) validating new events. The repository will continue to grow and develop, and any modification will be promptly reported.

Important: we recommend to use an updated version of .



Example of detection using PyMPA. Templates (red waveforms) overlapped on continuous data (black) filtered from 3 to 8 Hz are shown. On the left for the used channels the corresponding cross-correlation value.

This package contains:

- *Routines for downloading data from eida servers*;

- *Routines for creating and trimming templates*;
- *Routines for calculating moveout time for synchronization*;
- *Kurtosis based template verification*;
- *Template matching by using daily estimation of MAD and all the available channels*;
- *Postprocessing routines*;
- *Visual verification of detections*;

Acknowledgements

The software development was partially funded by a joint research project within the executive program of scientific and technological cooperation between Italy and Japan for the period 2013–2015. Additional funds for software development come from the project "Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe" (SERA), responding to the priorities identified in the call INFRAIA-01-2016-2017 Research Infrastructure for Earthquake Hazard. We thank Monica Sugan for the extensive testing of the codes and Aitaro Kato at the Earthquake Research Institute (ERI) in Tokyo for fruitful discussions. The authors also wish to thank the ObsPy community for the continuous support and constant development of related libraries.

Citation

If you use this package in your work, please cite the following papers:

Vuan A., Sugan M., Amati G., Kato A., 2017 - Improving the Detection of Low-Magnitude Seismicity Preceding the Mw 6.3 L'Aquila Earthquake: Development of a Scalable Code Based on the Cross-Correlation of Template Earthquakes, BSSA https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/525813/improving-the-detection-of-low-magnitude?redirectedFrom=fulltext

Vuan A., Sugan M., Chiaraluce L., Di Stefano R., 2017 - Loading rate variations along a mid-crustal shear zone preceding the MW6.0 earthquake of the 24th of August 2016 in Central Italy, Geophysical Research Letters http://onlinelibrary.wiley.com/doi/10.1002/2017GL076223/full

Sugan, M., Vuan, A., Kato, A., Massa, M., & Amati, G. (2019). Seismic evidence of an early afterslip during the 2012 sequence in Emilia (Italy). Geophysical Research Letters, 46, 625–635. https://doi.org/10.1029/2018GL079617

Contents

Chapter 1

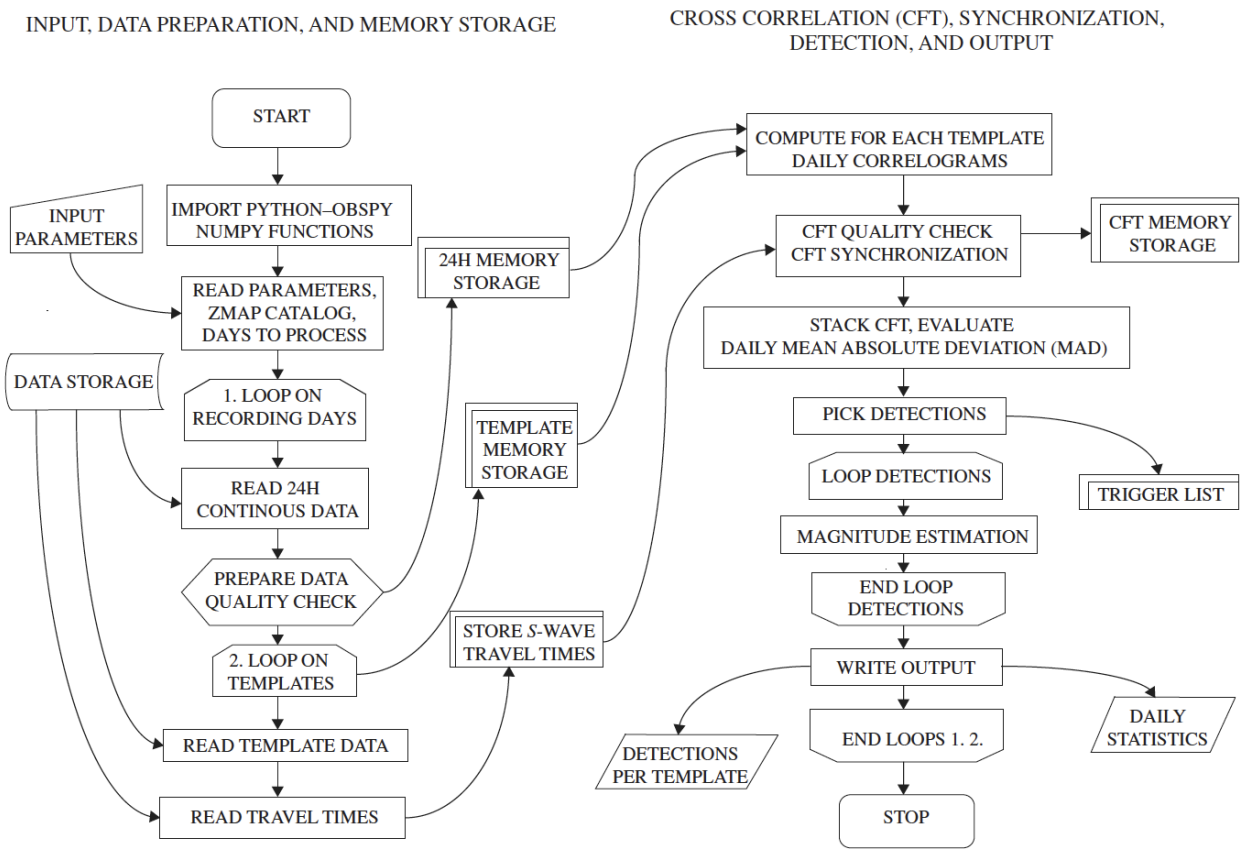
Introduction to the PyMPA package

This document is designed to give you an overview of the capabilities and implementation of the PyMPA Python package.

1.1 Motivation

PyMPA is designed to augment the detection capability of earthquakes, or any seismic signal (explosions or low frequency tremors) by using advanced routines different from the faster standard amplitude-ratio STA/LTA methods.

The technique allows to improve seismic catalogs decreasing the completeness magnitude and is particularly useful in detecting seismicity below the background noise level, and during a strong aftershocks sequence when the network sensitivity is lower. PyMPA is based on MFT search for earthquakes that resemble well-located events, termed templates (e.g., Shelly et al., 2007; Peng and Zhao, 2009; Yang et al., 2009; Kato et al., 2012; Zhang and Wen, 2015). The algorithm, which exploits ObsPy routines (Krischer et al., 2015), is versatile and supports most commonly used seismic data and earthquake catalog formats. A PyMPA flowchart is also shown below.



In addition to PyMPA, we develop other tools external to the main code to manage the input–output preparation and validation for (1) downloading data from Observatories and Research Facilities for European Seismology–European Integrated Data Archive (ORFEUS-EIDA) servers, (2) evaluating data quality, (3) selecting earthquakes as templates from a reference catalog, (4) trimming and filtering them from continuous waveforms, (5) avoiding redundant detections in the output, and (6) validating new events. This repository will continue to grow and develop and any modification will be reported in the github repository.

1.2 Supported environments

Linux, OSX and Windows environments running Python 2.7 and 3.x. We will stop support for Python 2.7 in a forthcoming release.

1.3 Functionality

Within *input* you will find the routines to generate templates, (*create_template*) select good templates (*template_check*), calculate travel times (*calculate_ttimes*), compute cross-channel correlations from these templates (*pympa*), process_detections (*process_detections*), and apply for visual inspection (*verify_detection*)

1.4 Running tests

For running tests examples are provided in the github subdirectories, tests are recalled when modifications are performed to the codes and a TRAVIS CI report is released.

You can also run these tests by yourself locally to ensure that everything runs as you would expect in your environment.

Although every effort has been made to ensure these tests run smoothly on all supported environments , if you do find any issues, please let us know on the page.

1.5 References

Shelly, D. R., G. C. Beroza, and S. Ide (2007). Non-volcanic tremor and low frequency earthquake swarms, Nature 446, 305–307.

Peng, Z., and P. Zhao (2009). Migration of early aftershocks following the 2004 Parkfield earthquake, Nature Geosci. 2, 877–881.

Yang, H., L. Zhu, and R. Chu (2009). Fault-plane determination of the 18 April 2008 Mount Carmel, Illinois, earthquake by detecting and relocating aftershocks, Bull. Seismol. Soc. Am. 99, 3413–3420.

Kato, A., K. Obara, T. Igarashi, H. Tsuruoka, S. Nakagawa, and N. Hirata (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake, Science 335, 705–708.

Zhang, M., and L. Wen (2015). An effective method for small event detection: Match and locate (M&L), Geophys. J. Int. 200, 1523–1537.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem, Comput. Sci. Discov. 8, no. 1, 014003, doi: 10.1088/1749-4699/8/1/014003.

Chapter 2

PyMPA installation

PyMPA is a pure Python package. It runs after the installation of a virtual environment with Numpy, Scipy, Matplotlib and Obspy libraries. Some C extensions of ObsPy toolkit are also used and Bottleneck libraries. Bottleneck is a set of functions inspired by NumPy and SciPy, but written in Cython with high performance in mind. Bottleneck provides separate Cython functions for each combination of array dimensions, axis, and data type.

We heavily recommend installing ObsPy using conda because:

- separate your install from your system default Python, avoiding to have problems with your OS;
- correct compilation is more probable

If you do not have either a miniconda or anaconda installation you can follow the instructions.

If you do not already have a conda environment we recommend creating one with the following:

```
conda create -n obspy python=3.6
conda activate obspy
conda install obspy
conda install bottleneck
```

For installing PyMPA you can simply clone the git repository and try it within your obspy env:

```
git clone git@github.com:avuan/PyMPA37.git
```

On a Linux system for installing conda, obspy, bottleneck and mirror the PyMPA code follow this commands:

```
curl -O https://repo.continuum.io/archive/Anaconda3-5.0.1-Linux-x86_64.sh
chmod +x Anaconda3-5.0.1-Linux-x86_64.sh
./Anaconda3-5.0.1-Linux-x86_64.sh
source ~.bashrc
conda config --add channels conda-forge
conda create -n obspy37 python=3.7
source activate obspy37
conda install obspy
conda install bottleneck
```

(continues on next page)

(continued from previous page)

```
mkdir test_obspy1.2.0
cd test_obspy1.2.0
git clone https://github.com/avuan/PyMPA37
```

Chapter 3

Tutorial

This tutorial is designed to give you an overview of the capabilities and implementation of the PyMPA Python package.

3.1 Downloading Seismological Data

To download seismological data from EIDA (European Integrated Data Archive) servers: Data from broad band seismic stations are available from many European Institutions. To download seismological data from EIDA (European Integrated Data Archive) servers and inventory data in STATIONXML format many examples can also be found in ObsPy.

PyMPA requires, continuous data and stations inventories. EIDA servers can easily release data from permanent networks and the corresponding inventories. The examples in the subdirectory input.download_data.dir show the python scripts that allows the download.

In the case your data come from other sources, PyMPA through ObsPy libraries is able to manage most of the seismological data formats (MSEED, SAC, SEISAN, SEGY, etc..). An inventory data file including station information needs to be created by modifying an existing StationXML file.

PyMPA does not use databases and prefers to store single channel daily continuous data in archieves or subdirectories.

Executable files:

- download_data.py (https://github.com/avuan/PyMPA37/tree/master/input.download_data.dir/download_data.py)
- download_inventory.py (https://github.com/avuan/PyMPA37/tree/master/input.download_data.dir/download_inventory.py)

(*download_data*) download data

3.2 Create Templates

(*create_template*) create templates

A Python script create_templates.py is used to trim templates from continuous data and inventories stored in an archive. Generally, we use travel times to cut events before and after S-wave arrivals. Thus, a reference 1D velocity model is needed. Trimmed waveforms have to be carefully checked to evaluate the effectiveness of S-wave travel time calculations. Take care that a high sampling rate could result in memory consumption and prolonged times of execution. Input data should be decimated a priori accordingly with your needs and availability of cores. Check the example for running create_templates.py at https://github.com/avuan/PyMPA37/tree/master/input.create_templates.dir

Executable file:

- create_templates.py (https://github.com/avuan/PyMPA37/tree/master/input.create_templates.dir/create_templates.py)

Input parameters:

```
#Line 1 -- list of stations
#Line 2 -- list of channels
#Line 3 -- list of networks
#Line 4 -- Lowpass frequency
#Line 5 -- Highpass frequency
#Line 6 -- Trimmed Time before S-wave
#Line 7 -- Trimmed Time after S-wave
#Line 8 -- UTC precision
#Line 9 -- Continuous data dir
#Line 10 -- Template data dir
#Line 11 -- Processing days list
#Line 12 -- Zmap catalog
#Line 13 -- Starting template
#Line 14 -- Stopping template
#Line 15 -- Taup Model
AQU CAMP CERT FAGN FIAM GUAR INTR MNS NRCA TERO
BHE BHN BHZ
IV MN
2.0
8.0
2.5
2.5
6
24h
template
lista1
templates.zmap
26
27
aquila_kato
```

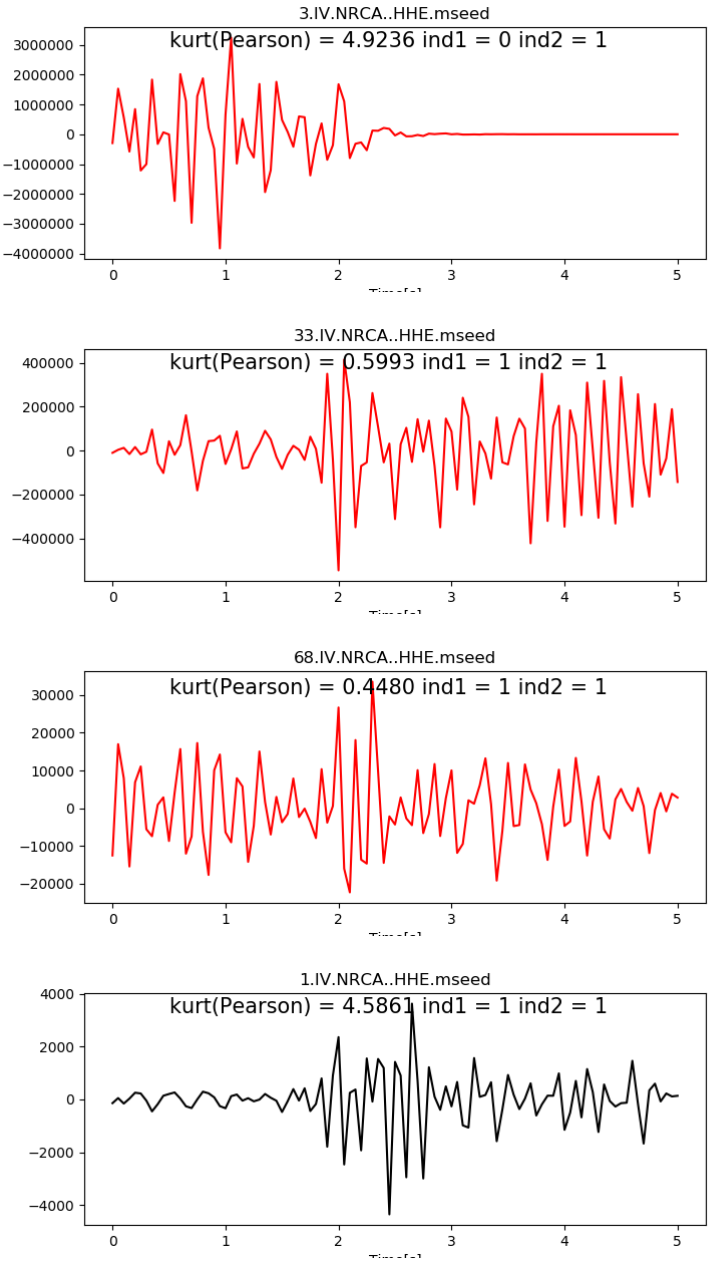
Note that input and output file names, inventories, template catalogs, velocity models are recalled also in the next steps.

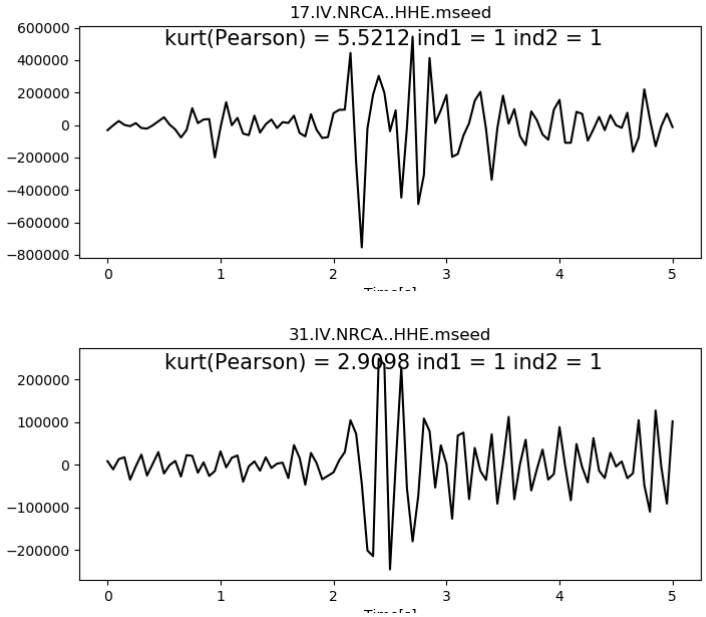
3.3 Check Template Quality

(*template_check*) select good templates

Evaluating template quality allows to input only a good signal to noise ratio avoiding artifacts resulting in unwanted detections. The selection is based on Kurtosis method (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html) supposing that the waveform is simmetrically trimmed at the first S-wave arrival. Kurtosis evaluates if the time distrbution of amplitudes is simmetric or not excluding data having a low signal to noise ratio. Peak amplitudes at the beginning or in the signal coda are unwanted and a selection is also made to exclude them.

Check examples running test_kurtosis1.py at https://github.com/avuan/PyMPA37/tree/master/input.template_check.dir After running kurtosis based selection, templates are separated in two subdirectories "bad"(red waveforms see figure below) and "good" (black waveforms see figure below)





Executable python scripts:

- template_check1.py at https://github.com/avuan/PyMPA37/tree/master/input.template_check.dir/template_check1.py (performs on waveform)
- template_check2.py at https://github.com/avuan/PyMPA37/tree/master/input.template_check.dir/template_check2.py (performs on the absolute values of waveform)

3.4 Calculate Travel Times

(*calculate_ttimes*) calculate travel times

Travel time calculation is based on Java TauP Toolkit as implemented in ObsPy (https://docs.obspy.org/packages/obspy.taup.html) Travel times are needed for synchronization to obtain a stacked cross-correlation function. It is supposed that trimmed templates are stored in ./template directory. The same reference 1D velocity model used for trimming templates is needed.

Executable file:

- calculate_ttimes.py at https://github.com/avuan/PyMPA37/tree/master/input.calculate_ttimes.dir/calculate_ttimes.py

Input parameters:

```
#Line 1 -- list of stations
#Line 2 -- list of channels
#Line 3 -- list of networks
#Line 4 -- Lowpass frequency
#Line 5 -- Highpass frequency
#Line 6 -- Trimmed Time before S-wave
#Line 7 -- Trimmed Time after S-wave
#Line 8 -- UTC precision
#Line 9 -- Continuous data dir
```

(continues on next page)

(continued from previous page)

```
#Line 10 -- Template data dir
#Line 11 -- Processing days list
#Line 12 -- Zmap catalog
#Line 13 -- Starting template
#Line 14 -- Stopping template
#Line 15 -- Taup Model
AQU CAMP CERT FAGN FIAM GUAR INTR MNS NRCA TERO
BHE BHN BHZ
IV MN
2.0
8.0
2.5
2.5
6
template
ttimes1
lista1
templates.zmap
26
27
aquila_kato
```

Note that input and output file names, inventories, template catalogs, velocity models are recalled also in the next steps.

3.5 Running PyMPA

Template matching code, using cross-correlation based on well located events. The code is embarassingly parallel and different templates/days can be run on different cores. We do not provide the scripts to parallelize jobs preferring to leave to the user to find the best strategy to accomplish the task. We generally prefer to distribute the workload by using Slurm.

Executable files:

- pympa.py (working on daily chunks and with a reduced number of channels). Chunking daily data results in MAD calculated on the working time window.

Input parameters:

Input parameters are described line by line in parameters24 file

```
# input parameters for runnig 27 version
# line1 = stations available,
# line2 = channels available,
# line3 = networks available,
# line4 = low bandpass filter frequency,
# line5 = high bandpass filter frequency,
# line6 = sample tolerance in detecting maximum cft amplitude for each channel,
```

(continues on next page)

(continued from previous page)

```
# line7 = cross-correlation threshold to be overcome by cft,
# line8 = min number of channels overcoming the cross-correlation threshold,
# line9 = template duration(s),
# line10 = UTCDateTime.DEFAULT_PRECISION (number of digits considered in fractions of␣
↪seconds),
# line11 = string variable containing the directory name for continuous 24h waveforms,
# line12 = string variable for templates' directory,
# line13 = string variable for travel times directory,
# line14 = filename for day list to process,
# line15 = filename for zmap catalog,
# line16 = template start number,
# line17 = template stop number(if line16==0  and line17==0 all templates are␣
↪processed,
# line18 = multiplying factor for MAD to determine threshold
# line 19 = multiplying factor to remove daily cft channels with std greater than␣
↪(average std from all channels * factor at at line 19)
# line 20 = multiplying factor to remove daily cft channels with std smaller than␣
↪(average std from all channels * factor at at line 20)
# line 21 = maximum number of templates to be used in template matching (choice is␣
↪made preferring the closest channels)
# line 22 = number of chunks per day (1=86400s, 2=43200, 3=28800, 4=21600, 6=14400 etc.
↪.. icreasing the factor allows reducing memory consumption)
APEC ATBU ATCA ATCC ATFO ATLO ATPC ATPI ATSC ATVO BADI FOSV FRON MURB NARO PARC PIEI␣
↪PE3 SSFR
EHE EHN EHZ HHE HHN HHZ
IV
3.0
8.0
6
0.35
6
5
6
24h
template
ttimes
lista1
templates.zmap
200
203
8
1.5
0.25
12
6
```

3.6 Output Processing

(*output.process_detections*) controls multiple detections in short time windows

A bash script calling python code performs the catalog sythesis. Some templates could concur to the same detection. The detection showing the highest threshold value is preferred in a fix time window (e.g. 6 seconds).

Executable file:

- bash script postproc37.sh
- process_detections.py at https://github.com/avuan/PyMPA37/tree/master/output.process_detections.dir/process_detections.py

Input parameters:

```
# - Line 1 - FlagDateTime (Enter 1 for timestamp format (2009-03-30T21:59:43.616111Z)␣
↪or Enter 0 for having seconds in a day (for plotting))
# - Line 2 - Flag_cc (1 to have returned the best results in terms of average cross-
↪corr using the sample tolerance, 0 results at sample tolerance 0)
# - Line 3 - Input file
# - Line 4 - UTC precison used in time
# - Line 5 - window_length (half time window for searching unique events)
# - Line 6 - min_threshold (threshold used in parameters24 for detecting events, ␣
↪threshold * MAD)
# - Line 7 - min_nch (minimum number of channels to overcome the setup cc value in␣
↪parameters24 e.g. 0.6)
1
1
./dcat
6
3.0
8.0
7
```

3.7 Verify Detections

(*output.verify_detection*) for visual verification of events

Produce graphics windows showing the continuous data overlapped by templates events at the detection time.

Executable file:

- verify_detection.py at https://github.com/avuan/PyMPA37/tree/master/output.verify_detection.dir/verify_detection.py

Input parameters:

```
# Line 1 -- list of stations
# Line 2 -- list of channels
# Line 3 -- list of networks
# Line 4 -- Lowpass frequency
# Line 5 -- Highpass frequency
# Line 6 -- Trimmed Time before S-wave
# Line 7 -- Trimmed Time after S-wave
# Line 8 -- UTC precision
# Line 9 -- Continuous data dir
# Line 10 -- Template data dir
# Line 11 -- Ttimes data dir
# Line 12 -- Zmap catalog
# Line 13 -- Starting detection
# Line 14 -- Stopping detection
# Line 15 -- Half of the visualized time window in sec
# Line 16 -- Taup Model
# Line 17 -- Flag_Save_Fig (0 = show, 1 = save figure)
# Line 18 -- Flag_Read_Stats (0 = no stats, 1 = read stats)
# Line 19 -- Tolerance in time between outcat origin time and stats
ATBU ATFO ATLO ATMI ATPC ATPI ATSC ATVO BADI CDCA MURB
EHE EHN EHZ HHE HHN HHZ
IV MN
3.0
8.0
2.5
2.5
6
24h
template
ttimes
templates.zmap
0
5
10
ato
1
1
0.05
```

3.8 References

Shelly, D. R., G. C. Beroza, and S. Ide (2007). Non-volcanic tremor and low frequency earthquake swarms, Nature 446, 305–307.

Peng, Z., and P. Zhao (2009). Migration of early aftershocks following the 2004 Parkfield earthquake, Nature Geosci. 2, 877–881.

Yang, H., L. Zhu, and R. Chu (2009). Fault-plane determination of the 18 April 2008 Mount Carmel, Illinois, earthquake by detecting and relocating aftershocks, Bull. Seismol. Soc. Am. 99,

3413–3420.

Kato, A., K. Obara, T. Igarashi, H. Tsuruoka, S. Nakagawa, and N. Hirata (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake, Science 335, 705–708.

Zhang, M., and L. Wen (2015). An effective method for small event detection: Match and locate (M&L), Geophys. J. Int. 200, 1523–1537.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem, Comput. Sci. Discov. 8, no. 1, 014003, doi: 10.1088/1749-4699/8/1/014003.

Chapter 4

Preprocessing Input

This package contains utilities for:

- *Routines for downloading data from eida servers*;
- *Routines for creating and trimming templates*;
- *Routines for calculating moveout time for synchronization*;
- *Kurtosis based template verification*;

These utilities are written by the PyMPA developers, and are distributed under the LGPL GNU Licence, Copyright PyMPA developers 2019.

4.1 Contents:

4.1.1 Download data

Downloading data from EIDA servers is easily performed by using ObsPy tools and routines. Please see the scripts at https://github.com/avuan/PyMPA37/blob/master/input.download_data.dir/

download_data.py and download_inventory.py allows downloading continuous data and inventories for the selected stations and time period.

For some networks or stations please check before data availability or if there are some restrictions and a token or userid is needed. Some EIDA servers when receiving from the same user many requests in a short time close the door to the user. That is the reason why we introduced a pause command in our scripts, this avoids the shutdown of the connection.

Here below a simple code to download continuous data,

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
#

import time
import os
```

(continues on next page)

(continued from previous page)

```
from obspy.clients.fdsn import Client
from obspy.core.utcdatetime import UTCDateTime

client = Client("INGV")
networks = ["IV"]
#
stations=["MURB", "NARO"]

channels = ["EH*", "HH*"]

start = "2012-02-05T00:00:00.000"
stop = "2012-02-06T00:00:00.000"

# 24h as seconds
chuncklength = 86400

# output directory
inp_dir = "./24h/"

# output directory
if not os.path.exists(inp_dir):
    os.makedirs(inp_dir)
# ---------------------------do not change below

for sta in stations:
    t1 = (UTCDateTime(start)).timestamp
    t3 = (UTCDateTime(stop)).timestamp
    t2 = t1 + chuncklength

    while t2 <= t3:
        print("station == ", sta)

        for net in networks:

            for chann in channels:

                try:
                    bulk = [(net, sta, "*", chann, UTCDateTime(
                        t1), UTCDateTime(t2))]
                    print("bulk == ", bulk)
                    yy = str(UTCDateTime(t1).year)
                    mm = str(UTCDateTime(t1).month).zfill(2)
                    dd = str(UTCDateTime(t1).day).zfill(2)

                    newfile = inp_dir + yy + mm + dd +\
                        "." + sta + "." + chann[0:3]
                    client.get_waveforms_bulk(bulk, filename=newfile)
                    time.sleep(2)
                except Exception:
```

(continues on next page)

(continued from previous page)

```
                    time.sleep(1)
                    pass

        t1 = t1 + chunclength
        t2 = t2 + chunclength
```

and for downloading station inventories

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
#
from obspy.clients.fdsn import Client
from obspy.core.utcdatetime import UTCDateTime


client = Client("INGV")

starttime = UTCDateTime("2011-01-01")
endtime = UTCDateTime("2017-06-30")

inventory = client.get_stations(
    starttime=starttime, endtime=endtime,
    network="MN", sta="AQU", channel="*", filename="inv.aqu", format='xml')
print(inventory)
```

4.1.2 Create templates

Actually templates are created by trimming a fixed time window focused on S-wave theoretical travel times. For details on the travel time calculations see file:///Users/vuan/PycharmProjects/PyMPA37/docs/build/html/tutorial.html#calculate-travel-times.

The length of the time window is established by the inter-statiion network distance and the frequency range used. The user should carefully check to exclude signal deriving from numerical artifacts (e.g. filtering applied to zero padding time windows having no data), or pre and coda signals not connected with the seismic perturbation investigated (e.g. LFEs, earthquakes, icequakes etc. . .) In the next versions the trimming will allow for selecting variable length P and S-waves.

Needed files:

- Events in a catalog: e.g. templates.zmap (quakeml or zmap format) see ObsPy for more details. An example of zmap file format is given here below (ZMAP is a simple 10 column CSV file (technically TSV) format for basic catalog data. It originates from ZMAP, a Matlab® based earthquake statistics package (see [Wiemer2001]).
- Columns represent: longitude, latitude, year, month, day, magnitude, depth, hour, minute, second - note that fields have to be separated by tabs.

```
13.38347 42.38842 2009 03 30 1.3 11.180 14 04 34.50
13.38981 42.33004 2009 03 30 1.5 9.170 14 15 4.32
13.39447 42.32878 2009 03 30 1.6 9.610 14 16 6.35
13.40577 42.31941 2009 03 30 1.4 5.370 14 17 1.74
13.36791 42.59639 2009 03 30 1.3 5.710 14 35 16.74
13.45651 42.25600 2009 03 30 1.3 5.620 14 42 23.29
13.38686 42.31381 2009 03 30 1.2 6.270 15 10 34.73
13.49527 42.30006 2009 03 30 0.9 6.810 15 21 2.28
13.32714 42.33182 2009 03 30 1.4 6.210 15 40 5.62
13.29606 42.27235 2009 03 30 1.2 8.690 15 53 5.35
```

- Suitable velocity model for computing travel times

```
aquila + ak135
depth P vel. S vel. density older density
    0.000      3.7500      2.1650      2.4500
    1.500      3.7500      2.1710      2.4500
    1.510      4.9400      2.8520      2.7800
    4.510      4.9400      2.8580      2.7800
    4.520      6.0100      3.2790      2.7600
   14.520      6.0100      3.2850      2.7600
   14.530      5.5500      3.3950      2.9100
   29.530      5.5500      3.4010      2.9100
   29.540      5.8800      4.0990      3.1000
   43.540      5.8800      4.1050      3.1000
   43.550      5.8800      4.5610      3.1000
   57.500      5.8800      3.3600      3.1000
   57.500      7.1100      4.0100      3.2300
   93.000      7.1100      4.0100      3.2300
   93.000      7.1000      3.9900      3.3000
  136.500      7.1000      3.9900      3.3000
  165.000      8.1750      4.5090      3.3487
  210.000      8.3000      4.5180      3.3960
```

- Station inventory (format consistent with ObsPy read_inventory routine see https://docs.obspy.org/packages/autogen/obspy.core.inventory.inventory.read_inventory.html)
- Days to process: one column file including days to process e.g. lista1

```
090330
```

- Set parameters: e.g. trim.par

```
#Line 1 -- list of stations
#Line 2 -- list of channels
#Line 3 -- list of networks
#Line 4 -- Lowpass frequency
#Line 5 -- Highpass frequency
#Line 6 -- Trimmed Time before S-wave
#Line 7 -- Trimmed Time after S-wave
#Line 8 -- UTC precision
```

(continues on next page)

(continued from previous page)

```
#Line 9 -- Continuous data dir
#Line 10 -- Template data dir
#Line 11 -- Processing days list
#Line 12 -- Zmap catalog
#Line 13 -- Starting template
#Line 14 -- Stopping template
#Line 15 -- Taup Model
AQU CAMP CERT FAGN FIAM GUAR INTR MNS NRCA TERO
BHE BHN BHZ
IV MN
2.0
8.0
2.5
2.5
6
24h
template
lista1
templates.zmap
26
27
aquila_kato
```

- Directory i.e. ./24h where 24h continuous data are stored
- Output dir i.e. ./template (find trimmed time series)

Note that input and output file names, inventories, template catalogs, velocity models are recalled also in the next steps and remain almost fixed. Parameters in files .par could change.

References Wiemer S. (2001), A software package to analyzeseismicity: ZMAP, Seismol Res Lett 92, 373-382

4.1.3 Calculate Travel Times

Theoretical travel-time arrivals are calculated using the ObsPy port of the Java TauP Toolkit routines, see https://docs.obspy.org/packages/obspy.taup.html (Crotwell et al., 1999). For using your own earth model see https://docs.obspy.org/packages/autogen/obspy.taup.taup_create.build_taup_model.html#obspy.taup.taup_create.build_taup_model Model initialization is an expensive operation. Thus, make sure to do it only if necessary. ObsPy include custom built models can be initialized by specifying an absolute path to a model in ObsPy's .npz model format instead of just a model name. See below for information on how to build a .npz model file. Building an ObsPy model file from a "tvel" or "nd" file is easy.

An example of tvel model to be compiled by build_taup_model Obspy function:

```
aquila + ak135
depth P vel. S vel. density older density
    0.000      3.7500      2.1650      2.4500
```

(continues on next page)

(continued from previous page)

```
    1.500      3.7500      2.1710      2.4500
    1.510      4.9400      2.8520      2.7800
    4.510      4.9400      2.8580      2.7800
    4.520      6.0100      3.2790      2.7600
   14.520      6.0100      3.2850      2.7600
   14.530      5.5500      3.3950      2.9100
   29.530      5.5500      3.4010      2.9100
```

Needed files:

- Events in a catalog: e.g. templates.zmap (quakeml or zmap format) see ObsPy for format
- Suitable velocity model for computing travel times
- Station inventory (format consistent with ObsPy read_inventory routine see https://docs.obspy.org/packages/autogen/obspy.core.inventory.inventory.read_inventory.html)
- Days to process: one column file including days to process e.g. lista1
- Set parameters: e.g. times.par

```
#Line 1 -- list of stations
#Line 2 -- list of channels
#Line 3 -- list of networks
#Line 4 -- Lowpass frequency
#Line 5 -- Highpass frequency
#Line 6 -- Trimmed Time before S-wave
#Line 7 -- Trimmed Time after S-wave
#Line 8 -- UTC precision
#Line 9 -- Continuous data dir
#Line 10 -- Template data dir
#Line 11 -- Processing days list
#Line 12 -- Zmap catalog
#Line 13 -- Starting template
#Line 14 -- Stopping template
#Line 15 -- Taup Model
AQU CAMP CERT FAGN FIAM GUAR INTR MNS NRCA TERO
BHE BHN BHZ
IV MN
2.0
8.0
2.5
2.5
6
template
ttimes1
lista1
templates.zmap
26
27
aquila_kato
```

- Input directory i.e. ./template where trimmed templates are found
- Output dir i.e. ./ttimes (find moveout times from different channels used to synchronize

cross-correlation functions)

References

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, Seismol. Res. Lett. 70, 154–160.

4.1.4 Template Kurtosis-based Waveform Check

The input data quality of continuous waveforms is verified by testing daily gaps and overlapping streams and ensuring that a certain percentage of the requested time span is available. To obtain accurate results, input preparation is critical, and template waveforms should be carefully checked before running PyMPA. Filter frequency band selection depends on the best signal-to-noise ratio of templates and is related to the structural model and epicentral distances involved. statistics is used to evaluate the simmetry of the time series neglecting from the pool of trimmed templates waveforms that show high simmetry in the S-wave selected time window. It is supposed that low signal to noise ratios have low values of index. This contributes to exclude the template/channel from the estimation of the cross-correlation. The routine avoids also signals or glitches that are located at the beginning and at the end of the signal (please check carefully the code before using it). Scipy is used as a standard routine.

(see https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html)

An example using the input_template_check is provided also showing accpted and removed waveforms. 

Chapter 5

Main Program

PyMPA procedure provides the detection of microseismicity starting from well located templates by using a cross-correlation function from a network. The code is stored on , and can be freely cloned on your platform. It supports Python 2.7, 3.4, 3.5, 3.6, 3.7 releases and uses for reading and writing seismic data, and for handling most of the event metadata. Matched-filter correlations are calculated using a python normalised cross-correlation function or the ObsPy v. 1.2.0 released on April 2019. Detections can be also obtained using a single station and three channels by modifying the input parameters (thresholds etc..). This version is running on a single core and does not include multiprocessing routines. However, in the need of massive calculations for years and thousands of templates, it could be easily implemented a script using SLURM or other schedulers to submit many jobs to the available processors.

Important: we recommend to use an updated version of ObsPy.

Main packages contains:

- *Template matching by using daily estimation of MAD and all the available channels*;

This package is written by the PyMPA developers, and is distributed under the LGPL GNU Licence, Copyright PyMPA developers 2019.

5.1 Contents:

5.1.1 Running PyMPA

The input files are those prepared in advance by using pre-processing tools. The only input file that changes is the parameters24 input file.

```
# input parameters for runnig 27 version
# line1 = stations available,
# line2 = channels available,
# line3 = networks available,
# line4 = low bandpass filter frequency,
# line5 = high bandpass filter frequency,
# line6 = sample tolerance in detecting maximum cft amplitude for each channel,
```

(continues on next page)

(continued from previous page)

```
# line7 = cross-correlation threshold to be overcome by cft,
# line8 = min number of channels overcoming the cross-correlation threshold,
# line9 = template duration(s),
# line10 = UTCDateTime.DEFAULT_PRECISION (number of digits considered in fractions of␣
↪seconds),
# line11 = string variable containing the directory name for continuous 24h waveforms,
# line12 = string variable for templates' directory,
# line13 = string variable for travel times directory,
# line14 = filename for day list to process,
# line15 = filename for zmap catalog,
# line16 = template start number,
# line17 = template stop number(if line16==0  and line17==0 all templates are␣
↪processed,
# line18 = multiplying factor for MAD to determine threshold
# line 19 = multiplying factor to remove daily cft channels with std greater than␣
↪(average std from all channels * factor at at line 19)
# line 20 = multiplying factor to remove daily cft channels with std smaller than␣
↪(average std from all channels * factor at at line 20)
# line 21 = maximum number of templates to be used in template matching (choice is␣
↪made preferring the closest channels)
# line 22 = number of chunks per day (1=86400s, 2=43200, 3=28800, 4=21600, 6=14400 etc.
↪.. icreasing the factor allows reducing memory consumption)
APEC ATBU ATCA ATCC ATFO ATLO ATPC ATPI ATSC ATVO BADI FOSV FRON MURB NARO PARC PIEI␣
↪PE3 SSFR
EHE EHN EHZ HHE HHN HHZ
IV
3.0
8.0
6
0.35
6
5
6
24h
template
ttimes
lista1
templates.zmap
200
203
8
1.5
0.25
12
6
```

Needed input files:

- Events in a catalog: e.g. templates.zmap (quakeml or zmap format) see ObsPy for format
- Suitable velocity model for computing travel times

- Station inventory (format consistent with ObsPy read_inventory routine see https://docs.obspy.org/packages/autogen/obspy.core.inventory.inventory.read_inventory.html)
- Days to process: one column file including days to process e.g. lista1
- Set parameters: e.g. parameters24
- Input directory ./template where trimmed templates are found
- Input directory ./24h where 24 hours continuous waveforms are stores
- Input directory /ttimes (find moveout times from different channels used to synchronize cross-correlation functions)

Output:

- Output files .cat, .stats, .stats.mag, .except (details on the output)

Detections (.cat)

Detections are listed in a .cat file (e.g. 200.100723.cat)

```
200 2010-07-23T02:20:29.833321Z 0.18 0.624 19.077 0.388 11.855 11
200 2010-07-23T06:22:08.273321Z 0.4 0.565 16.951 0.289 8.667 10
200 2010-07-23T06:46:09.553321Z 0.55 0.652 19.56 0.351 10.526 11
200 2010-07-23T22:20:57.553321Z 1.88 1.0 28.431 0.361 10.273 12
200 2010-07-23T22:32:10.713321Z -0.35 0.499 14.198 0.318 9.04 10
200 2010-07-23T22:42:34.113321Z -0.02 0.555 15.773 0.292 8.302 11
200 2010-07-23T22:53:24.393321Z 0.72 0.642 18.254 0.282 8.012 12
200 2010-07-23T23:09:24.953321Z 1.77 0.675 19.178 0.407 11.568 12
```

Columns in 200.100723.cat file are:

- 1.Template number corresponding to the python line index in file templates.zmap (event catalog)
- 2.UTC Date and Time (2010-07-23T02:20:29.833321Z) Time precision selection is possible in parameters24 input file
- 3.Magnitude estimated as in Peng and Zhao (2009). The magnitude of the detected event is calculated as the median value of the maximum amplitude ratios for all channels between the template and detected event, assuming that a 10-fold increase in amplitude corresponds to a one-unit increase in magnitude.
- 4.Average cross-correlation estimated from the channels that concurred to the detection. This value is estimated using a time shift between the channels that optimized the stacked CFT.
- 5.Threshold value (ratio between the amplitude of the CFT stacking and the daily MAD Median Absolute Deviation). The higher the threshold the most probable the detection. This value is estimated using a time shift between the channels that optimized the stacked CFT.
- 6.Average cross-correlation estimated from the channels that concurred to the detection at no time shift.
- 7.Threshold value (ratio between the amplitude of the CFT stacking and the daily MAD Median Absolute Deviation). No time shift of the signal cross-correlation functions is allowed.
- 8.Number of channels for which the cross-correlation is over a certain lower bound (e.g. 0.35)

Single Channel Statistics is listed in a .stats file (e.g. 200.100723.stats)

```
IV.ATFO HHE -0.010616075281 0.721025616044 -1.0
IV.ATFO HHZ 0.283285750909 0.66306439801 -1.0
IV.ATLO EHE 0.777449171081 0.777449171081 0.0
IV.ATLO EHN -0.0852349513925 0.650466054757 1.0
IV.ATLO EHZ 0.322139846713 0.366398836428 -4.0
IV.ATPI EHE 0.537447491069 0.650271304516 -1.0
IV.ATVO HHE 0.435372478132 0.435372478132 0.0
IV.ATVO HHN 0.573539172544 0.573539172544 0.0
IV.ATVO HHZ 0.133907687217 0.346238209304 -1.0
IV.MURB HHE 0.435748932184 0.792480883079 1.0
IV.MURB HHN 0.452987060234 0.713344389699 1.0
IV.MURB HHZ 0.794985414714 0.794985414714 0.0
100723 200 0 2010-07-23T02:20:29.833321Z 0.18 1.88 11.0 0.0326947876687 0.624 19.077 0.
↪388 11.855 12.0 9.0 5.0 0.0
IV.ATFO HHE 0.044988233346 0.231377904481 -5.0
IV.ATFO HHZ 0.149551602865 0.420184950211 -1.0
IV.ATLO EHE 0.224489982653 0.664636238755 1.0
IV.ATLO EHN 0.631160567096 0.631160567096 0.0
IV.ATLO EHZ 0.033085860272 0.262689503234 -3.0
IV.ATPI EHE 0.0362705954363 0.610553204714 -1.0
IV.ATVO HHE 0.297118217589 0.525933024914 -1.0
IV.ATVO HHN 0.36078302459 0.403840027808 -1.0
IV.ATVO HHZ 0.482921869799 0.482921869799 0.0
IV.MURB HHE 0.0567199403744 0.857500320272 1.0
IV.MURB HHN 0.287008419281 0.828772698562 1.0
IV.MURB HHZ 0.864438991339 0.864438991339 0.0
100723 200 0 2010-07-23T06:22:08.273321Z 0.4 1.88 10.0 0.0333501008197 0.565 16.951 0.
↪289 8.667 10.0 7.0 3.0 0.0
```

Columns in 200.100723.stats file are:

- 1.Network.Station
- 2.Channel
- 3.Cross-correlation value at no time shift
- 4.Cross-correlation value with time shift (nsamples) as in column 5
- 5.Time shift in nsamples (e.g. -1.0 means that the shift is equal to 0.05 at 20Hz sampling rate)

At the end of each trace id you find other parameters related to the detection in part repeating the detection parameters in .cat file and in part related to the cross-correlations values over some limits (0.3 - 0.5 - 0.7 - 0.9).

- date, template_num, detection_num, date&time, template_magnitude, detection_magnitude, threshold_fixed, MAD, ave_crosscc, threshold_record, ave_crosscc_0, threshold_record_0, num_channels_gt0.3, num_channels_gt0.5, num_channels_gt0.7, num_channels_gt0.9
- 100723 201 0 2010-07-23T22:20:57.712239Z 1.51 0.07 9.0 0.0342230009997 0.486 14.193 0.301 8.796 11.0 5.0 2.0 0.0

```
ATLO.EHZ 0.596105028863
```

(continues on next page)

(continued from previous page)

```
ATLO.EHE 0.170393549956
ATLO.EHN 0.373070244109
ATVO.HHE 0.329951522763
ATVO.HHZ 0.42550877913
ATVO.HHN 0.0483225655661
MURB.HHZ 0.191841464532
MURB.HHE 0.0947942692072
MURB.HHN 0.171838180932
ATPI.EHE 0.250359739024
ATFO.HHZ 0.145329568414
ATFO.HHE 0.201548385896
100723 200 0 2010-07-23T02:20:29.833321Z 0.18 1.88 11.0 0.0326947876687 0.624 19.077 0.
↪388 11.855 12.0 9.0 5.0 0.0
ATLO.EHZ 0.388076513533
ATLO.EHE 0.292051072606
ATLO.EHN 0.495577417282
ATVO.HHE 0.351604524548
ATVO.HHZ 0.379804503709
ATVO.HHN 0.144512967971
MURB.HHZ 0.495658912113
MURB.HHE 0.331121201612
MURB.HHN 0.497348207422
ATPI.EHE 0.151842475707
ATFO.HHZ 0.36823188478
ATFO.HHE -0.00476247218528
100723 200 0 2010-07-23T06:22:08.273321Z 0.4 1.88 10.0 0.0333501008197 0.565 16.951 0.
↪289 8.667 10.0 7.0 3.0 0.0
```

Columns in 200.100723.stats.mag file are:

- 1.Station.Channel Mag.
- 2.date, template_num, detection_num, date&time, template_magnitude, detection_magnitude, threshold_fixed, MAD, ave_crosscc, threshold_record, ave_crosscc_0, threshold_record_0, num_channels_gt0.3, num_channels_gt0.5, num_channels_gt0.7, num_channels_gt0.9
- 3.100723 201 0 2010-07-23T22:20:57.712239Z 1.51 0.07 9.0 0.0342230009997 0.486 14.193 0.301 8.796 11.0 5.0 2.0 0.0

Note that input and output file names, inventories, template catalogs, velocity models are recalled also in the next steps and remain almost fixed. Parameters in files .par could change.

Chapter 6

Creating an output catalog and verify detections

It consists of some separate utilities in , post-processing tools to obtain a catalog and verify events. Many events could be correlated to more than one template in a narrow time window. A fixed time window length can be selected, and within each, the template for which the normalized cross-correlation coefficient is the greates provides the event location and data to determine the magnitude. This process is run by <./sub/output.process_detections> in to two steps, by using the last event origin time as a reference to set the next time window scrutinized. If template matching performs well and input data are reliable we expect you are able to increase your catalog.

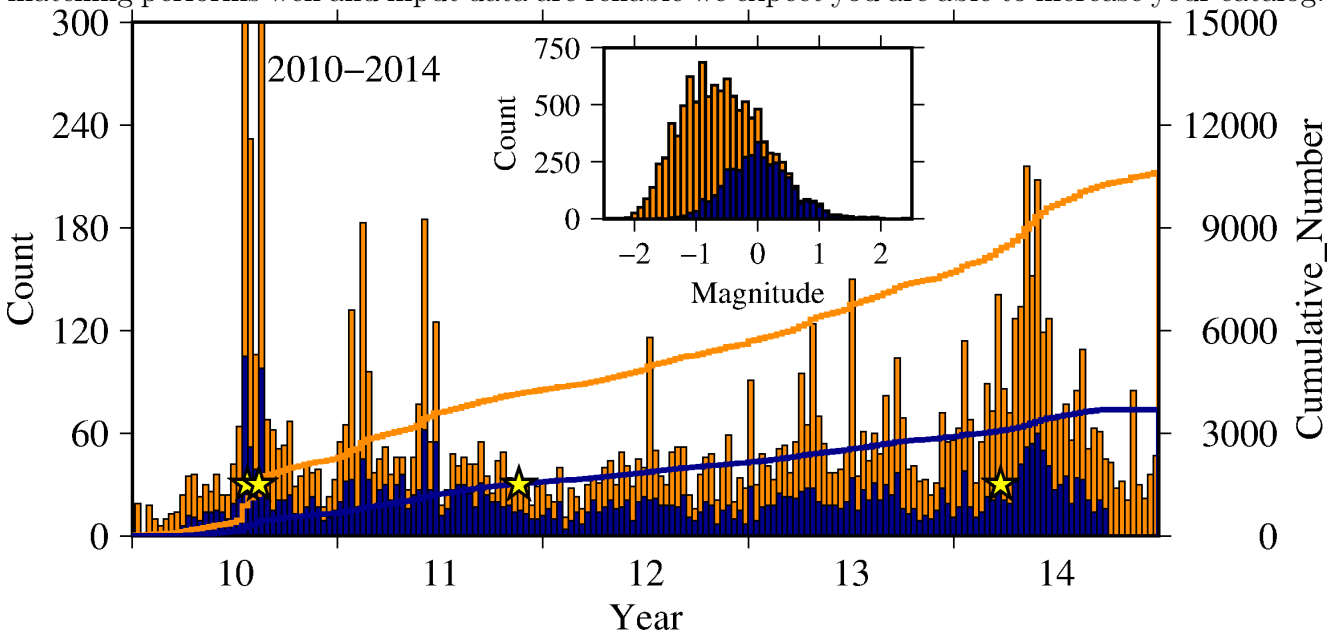


Figure - 5 years template events along the Alto-Tiberina fault (blue histograms) as recorded by ATF test bed and augmented detections by PyMPA (orange). Solid lines represent the cumulative number of earthquakes. In the inset the magnitude distribution of the templates in comparison with the augmented catalog. The completeness magnitude is decreased by 0.5. Stars indicate the timing of eartquakes between magnitude 2 and 2.8.

The final catalog should be verified by visual inspection for a number of sampled detections. Generally, we proceed by verification of events having low thresholds to understand a safe value to validate the catalog. The routine <./sub/output.verify_detection>

creates graphs of time windows where continuous data and trimmed templates are plotted with info grasped from channel by channel cross-correlation process.

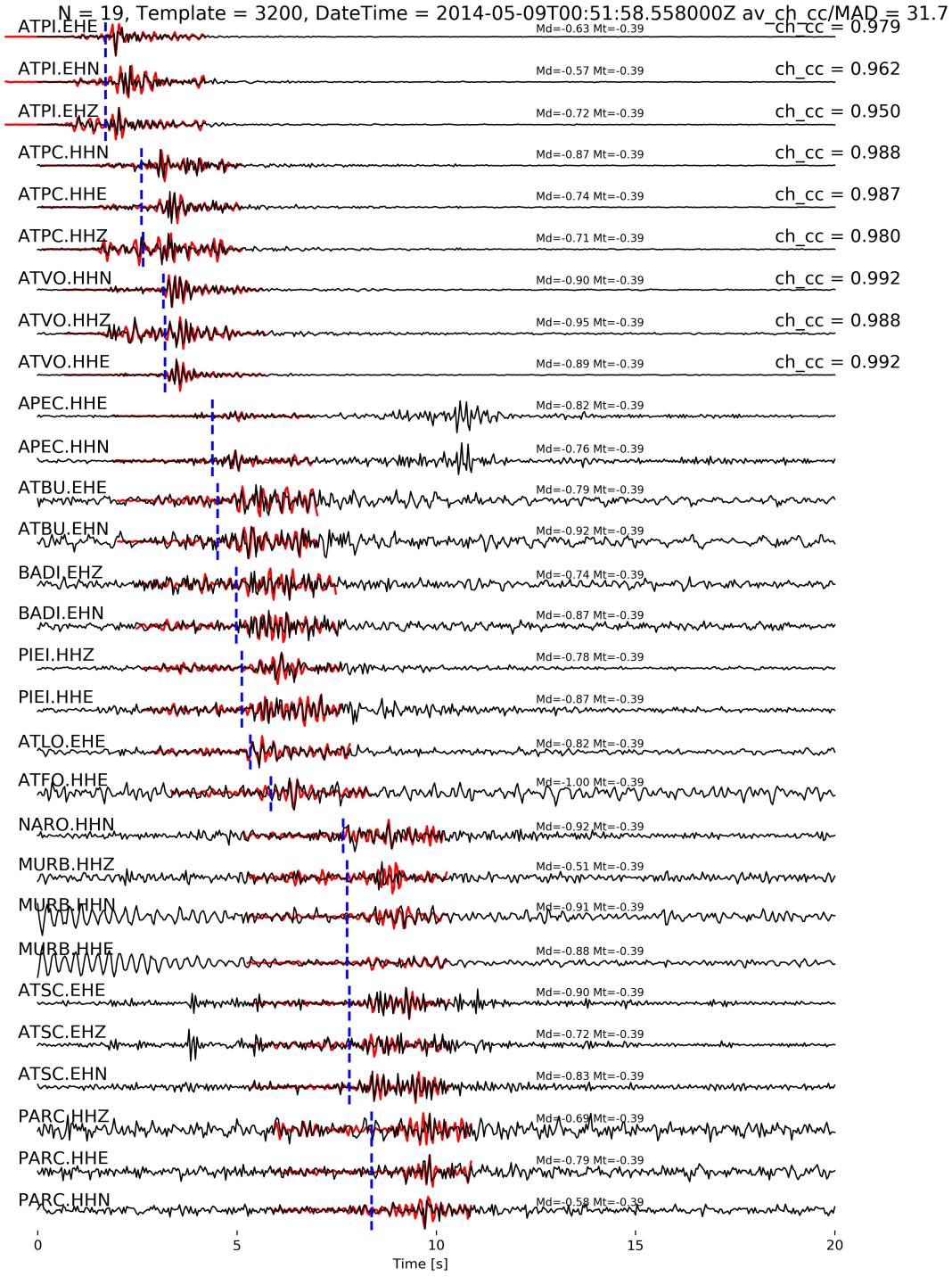


Figure - Visual inspection of a repeater with an average cross-correlation value of 0.98 along the Alto-Tiberina fault. Seismic data from ATF test bed (black waveforms) and a template (red) are overlapped. On the left of the panel, we list the station and channel codes, and on the right the corresponding value of cross-correlation between the two events. The template event (a magnitude M=-0.4) allows to detect a M=-1.0 repeater.

Important: we recommend to use an updated version of .

These utilities contains:

- *Postprocessing routines*;
- *Visual verification of detections*;

This package is written by the PyMPA developers, and is distributed under the LGPL GNU Licence, Copyright PyMPA developers 2019.

6.1 Contents:

6.1.1 Process PyMPA Output

From the defined template by cross-correlation for a selected day is obtained a list of possible detections. Since time overlapping detections could be found also by different templates, the procedure allows a search for the template able to detect the events with the highest threshold value that is also related with the highest average cross-correlation value for the used network. From the main program, many cat files are daily sorted and grouped for a scan of the events showing the best detections. A bash script is used to collect data and create the input for the procedure and a python script is used to filter the result on the basis of the fiter.par parameters used. The filter.par defines some additional filtering to detections to possibly overcome a visual validation of the new detections.

6.1.2 Visual verification of detections

Visual verification is needed to evaluate problematic cases or estimate a safe threshold to validate the catalog of detections. Visual verification needs the new detections catalog, the templates.zmap list, the daily template detections (e.g. 230.140913.cat file) and relative statistics (e.g. 230.140913.stats)cat file) and relative statistics (e.g. 230.140913.stats)cat file) and relative statistics (e.g. 230.140913.stats). The verify.par parameters allow for selecting the stations, frequency range, and window length for the visual verification.