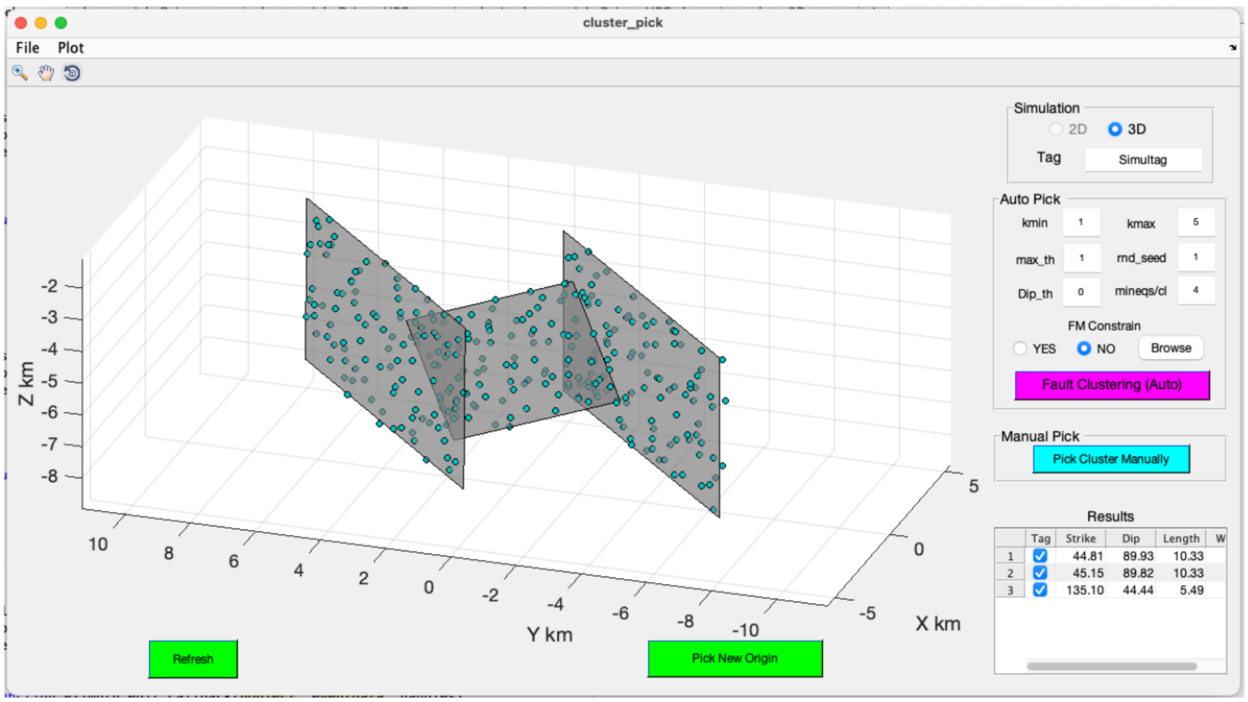
Cluster Pick User Manual

Oluwaseun I. Fadugba, Charles A. Langston and Christine A. Powell

September 15, 2022



Center for Earthquake Research and Information (CERI), University of Memphis, TN.

September 2022

Contents

1 Overview

Cluster Pick (written in MATLAB_R2021a) performs clustering analyses of hypocenter locations to constrain the realistic fault geometry in a seismic zone using a modified Optimal Anisotropic Dynamic Clustering (OADC) algorithm. The OADC method is a generalization of k-means clustering using principal component analysis of hypocenter locations by minimizing the cluster's smallest eigenvalue (Ouillon et al., 2008). We extend the OADC method by incorporating earthquake focal mechanisms to specify seed planes rather than randomly seeded planes, minimum dip angles to avoid subhorizontal faults, local splitting of faults within the 'thick' cluster to resolve complicated fault geometry involving systems of closely spaced, parallel, and listric faults, and avoid overfitting by combining coplanar faults. **Cluster Pick**'s fault models can be incorporated into seismic hazard models, geodynamic model for stress modeling and other applications. **Cluster Pick** is built on previous codes by Dr. Charles Langston.

Cluster Pick can be used as a Graphic User Interface (GUI) on a personal computer. It can also be run on the terminal and remotely on the High-Performance Computers (HPC) to determine a suite of fault geometries using several random-number generator seeds that fit the input hypocenters. The fault models can be delineated either by manually picking the clustered hypocenters in the input hypocenters or by automatic picking. The GUI version can only be used for one random number generator seed at a time. The suites of fault models help determine the distribution of the strike and dip angles of the faults in all the fault models to generate a unified fault model for the seismic zone.

Cluster Pick can also be used to generate 2D and 3D synthetic hypocenters with complicated geometries given the strike and dip, fault centers, and length and widths of the desired number of fault planes. **Cluster Pick** can also be used to perform a data clustering analysis on the original hypocenters using the cumulative tetrahedra volume method of Ouillon and Sornette (2011) to highlight the faults by separating clustered and diffuse seismicity. We refer readers to Fadugba et al. (2021) for a detailed description of the method and its application to real datasets.

2 Tutorial

2.1 Downloading and Setting up Cluster Pick

Download **Cluster Pick** software package into a working directory. The folder contains the following files and folder:

- cluster_pick.m ⇒ MATLAB script to run the graphic user-interface version of **Cluster Pick**

- cluster_pick_Driver.m ⇒ MATLAB script to run the terminal version **Cluster Pick**. It can be used to generate fault geometries using several random-seed generator values.

- cluster_pick_Drive_HPC.m ⇒ MATLAB script to run **Cluster Pick** on the HPC. This version can also be used to generate fault geometries using several random-seed generators.

- cluster_pick.fig ⇒ GUI figure file. You don't need to open this file.

- src ⇒ A directory containing the source codes for **Cluster Pick**

- submit_cluster_pick_Driver_HPC.sh ⇒ Submission shell script example to run **Cluster Pick** on the HPC.

- All_write_mechanisms.txt ⇒ Sample focal mechanisms file for **Cluster Pick**

- testdata.txt ⇒ Sample hypocenter location file for **Cluster Pick**

Run **cluster_pick.m** file in the working directory to open the user-interface version of **Cluster Pick**. The layout of **Cluster Pick** is shown in Figure 1. The left panel has a graph panel to display input hypocenters and the results, a refresh button, and the button to pick a new origin for the dataset. The right panel contains several editable fields to input the turnable parameters, buttons for running either manual fault picking or automatic fault picking, and a table panel to display the fault geometry results such as the strike, dip, length (L) and width (W) of the fault planes.

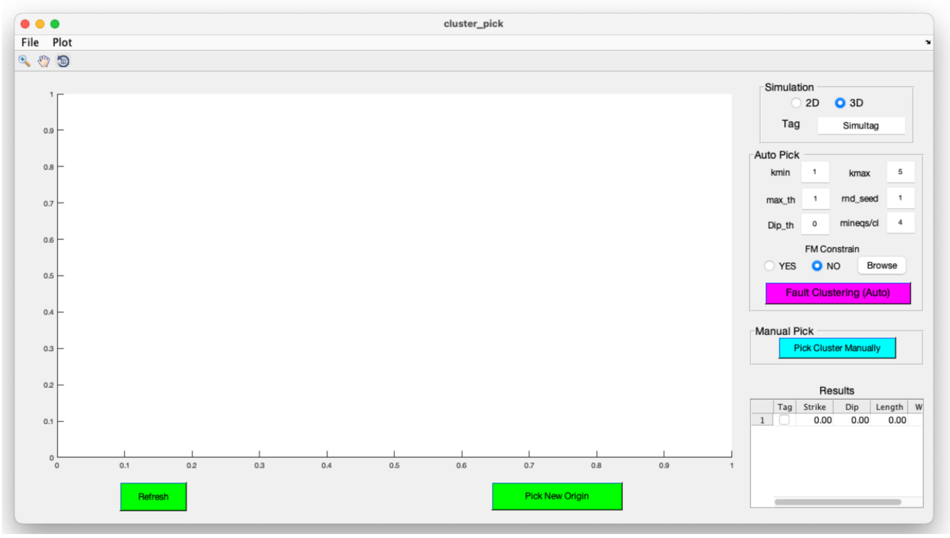


Figure 1: GUI of **Cluster Pick** at startup.

The input parameters and descriptions are as follows: The input parameters are:

- Tag ⇒ Simulation tag. The default tag is Simultag.

- kmin ⇒ minimum number of faults. The default value is 1.

- kmax ⇒ maximum number of faults. The default value is 5.
- max_th ⇒ the maximum cluster thickness. The default value is 1.
- rnd_seed ⇒ random-seed generator (integer). The default value is 1.
- Dip_th ⇒ minimum fault dip angle allowed in the algorithm. The default value is 0.
- mineqs/cl ⇒ minimum number of earthquakes in each cluster. The default value is 4.
- FM_Constrain ⇒ high confidence focal mechanisms to generate focal mechanism seeded planes instead of using the default randomly-seeded planes. The default is NO.

The **Files** and **Plot** menu bars (exapnded in Fig. 2) contain the following commands:

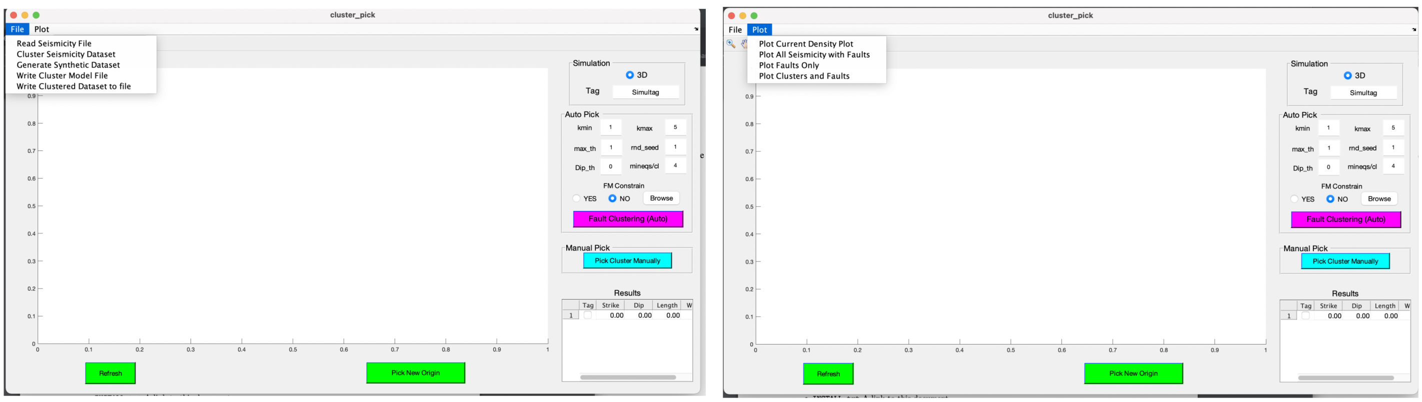


Figure 2: Menu bars of **Cluster Pick**.

- **File**
  - Read Seismicity File
  - Cluster Seismicity Dataset
  - Generate Synthetic Dataset
  - Write Cluster Model File
  - Write Clustered Dataset to file
- **Plot**
  - Plot current Density Plot
  - Plot All Seismicity with faults
  - Plot faults Only
  - Plot Clusters and Faults

Table 1: Fault Parameters for Generating Synthetic Datasets in Figure 3.

Parameter	Fault 1	Fault 2	Fault 3
Strike (°)	45	45	135
Dip (°)	90	90	45
Length (L, km)	10	10	5
Width (W, km)	5	5	4
Cloud thickness (Err, km)	1	1	1
No of hypocenters (nhypos)	100	100	100
Fault centers:			
x-coord (xb, km)	0	0	0
y-coord (yb, km)	4	-4	0
z-coord (zb, km)	-5	-5	-5

2.2 Loading Data

The input file is a text file with three columns for the hypocenter locations (x, y, z). A synthetic example of the input file is **testdata.txt** generated using the simple three fault geometry parameters in Table 1. The file format is the same for both synthetic and real datasets. The file can be loaded into **Cluster Pick** by clicking the **Read Seismicity File** in the **File** menu, and a dialog box will appear to navigate to the input file. At this point, the user can use the standard zoom, pan, and rotate buttons to explore the input hypocenters (Fig. 3).

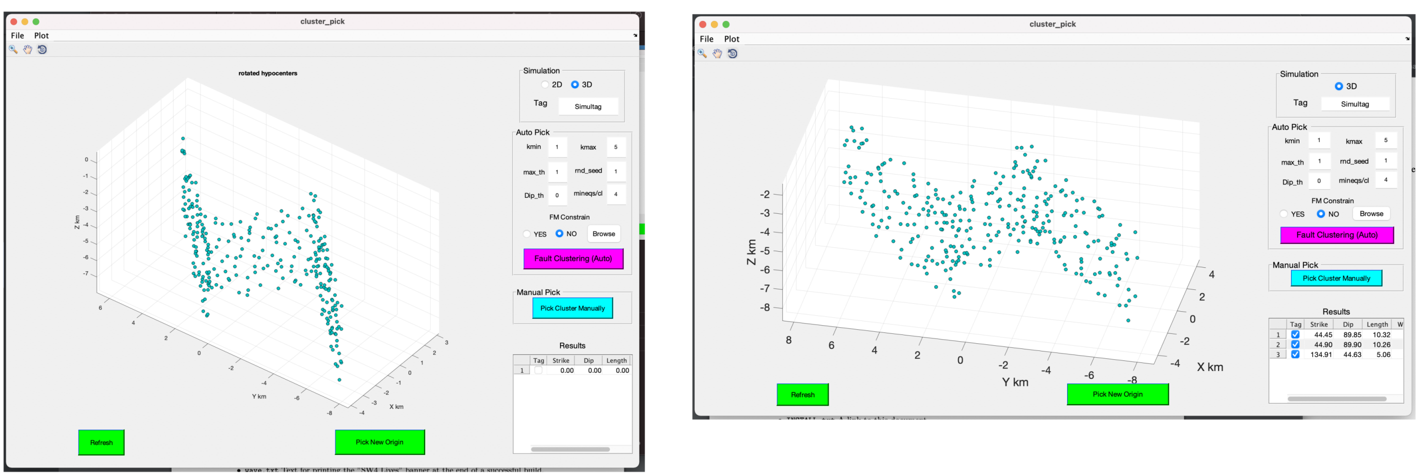


Figure 3: Loading hypocenter dataset to **Cluster Pick**.

Focal mechanism file is an optional file and should only be uploaded when the user wants to use focal-mechanism-seeded planes for the automatic fault picking instead of the default randomly-seeded planes. The file can be uploaded by changing the **FM Constrain** option to **Yes** from the default **No**, and using the **Browse** push button to navigate to the focal mechanism file (e.g., **All_write_mechanisms.txt**).

Cluster Pick can also be used to generate 2D and 3D synthetic hypocenters with complicated geometries given the strike, dip, length of the fault plane, etc. When the original hypocenters are diffused, **Cluster Pick** can also be used to perform clustering analysis to highlight the faults by separating clustered and diffuse seismicity. We will describe these two features in sections 2.4 and 2.5.

2.3 Fault Geometry Determination

The user can determine the fault geometries using manual or automatic picking.

2.3.1 Automatic Fault Picking Method

After setting all the required input parameters and files, the user can click the **Fault Clustering (Auto)** button to automatically determine the fault geometry that best fits the input hypocenters. Depending on the complexity of the input hypocenters and the input parameters, the algorithm may take some time at this stage. The fault parameters for each of the fault planes are displayed in the table panel, and a diary and saved variables files for other statistical analyses are also generated in the working directory. The results of the automatic fault picking using the default input parameters are in Figure 4 and Table 2.

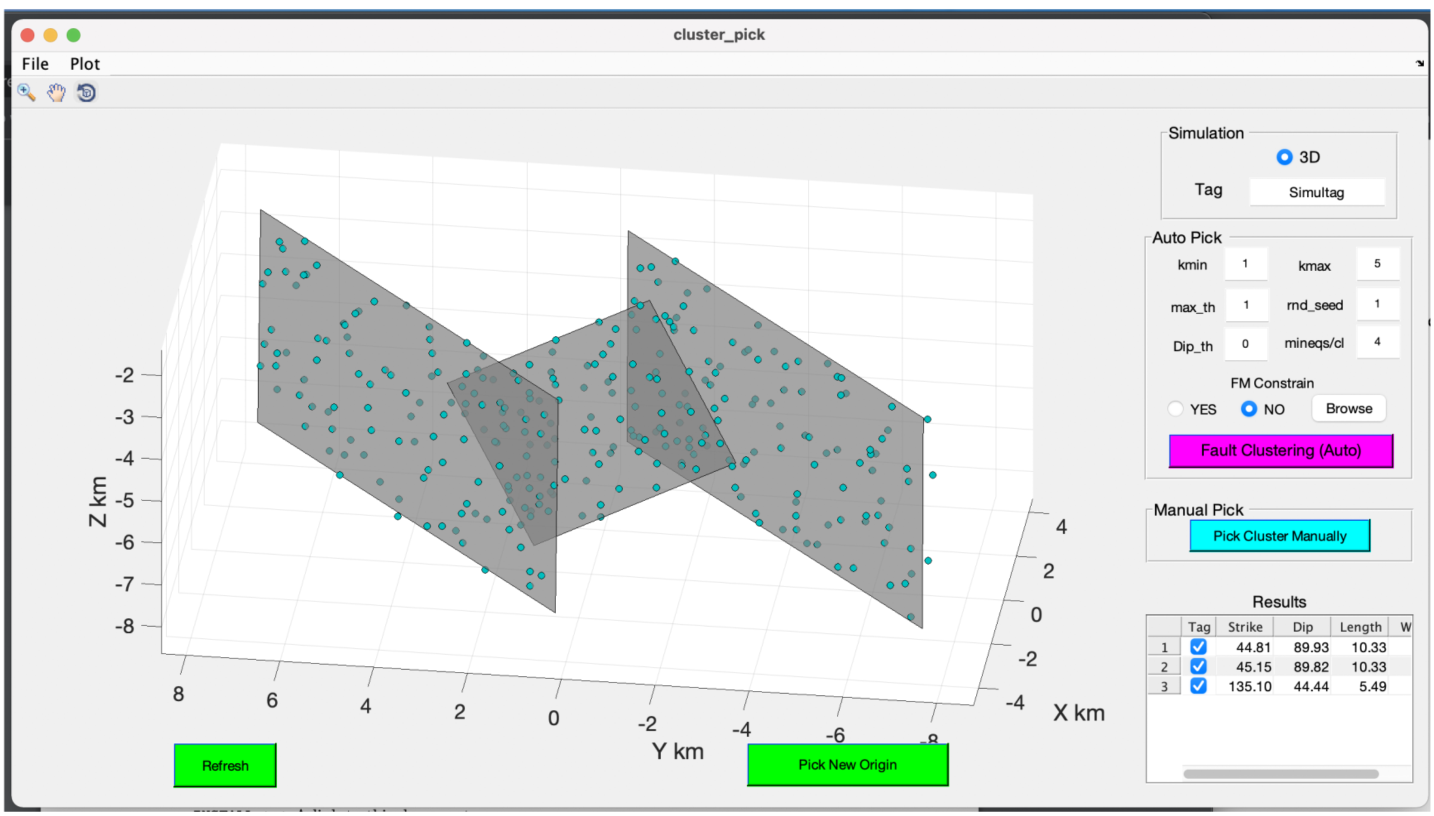


Figure 4: Automatic Fault Picking results.

Table 2: Results of the automatic and manual fault picking using the synthetic datasets in Table 1.

Parameter	Fault 1 (Auto/Manual)	Fault 2 (Auto/Manual)	Fault 3 (Auto/Manual)
Strike (°)	44.81/44.45	45.15/44.90	135.10/134.91
Dip (°)	89.93/89.85	89.82/89.90	44.44/44.63
Length (L, km)	10.33/10.32	10.33/10.26	5.49/5.06
Width (W, km)	5.14/5.02	5.07/4.98	3.93/3.93

2.3.2 Manual Fault Picking Method

The fault geometry can also be determined by manually picking with a polygon about hypocenter clusters that are visually recognizable using the **Pick Cluster Manually** button. **Cluster Pick** determines the fault plane that best fits the selected clusters. The user can click the **Pick Cluster Manually** button again to select another cluster. There is no limit to the number of clusters when using manual picking.

Pick Cluster Manually changes the plot view to a plane view for the manual picking, but the user can use the zoom, pan and rotate features after the clusters have been picked to explore the result (Fig. 5). The fault parameters for the fault planes that fit each cluster are displayed in the table panel. Figure 6 show three steps of the manually picking the faults in the synthetic dataset and their corresponding fault planes. The results of the manual fault picking are compared with the automatic picking in Table 2.

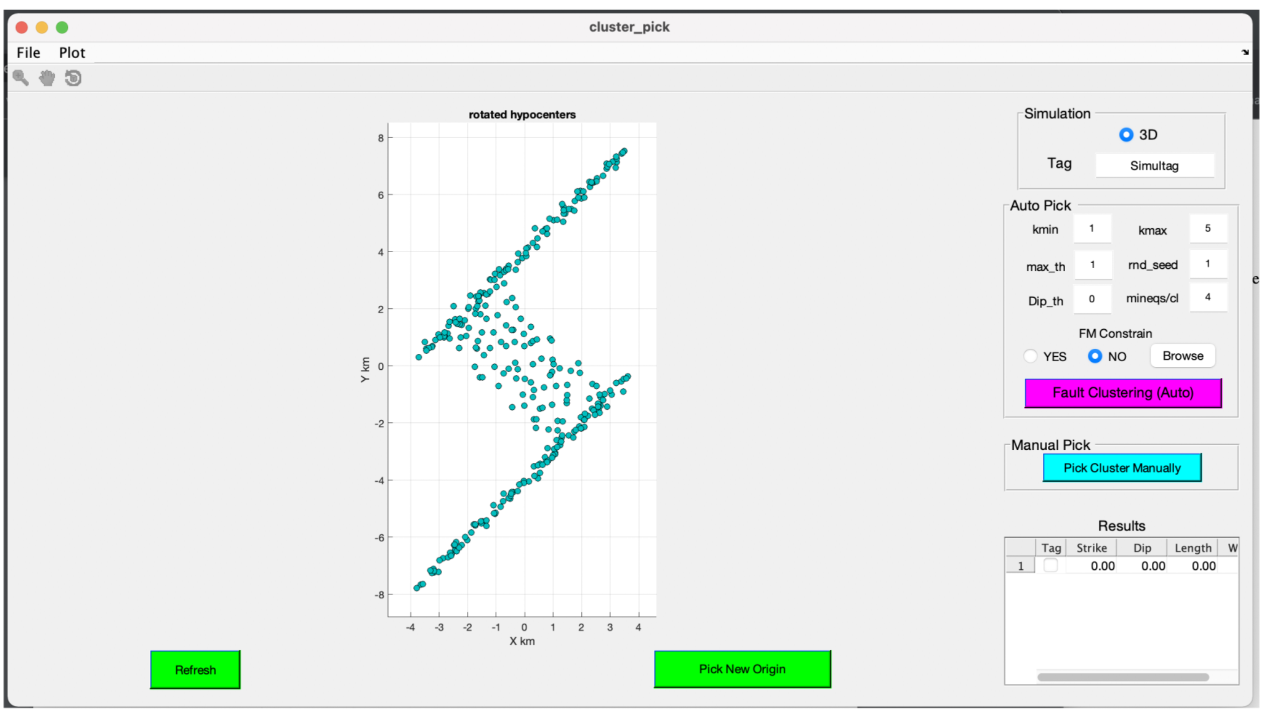


Figure 5: Starting Manual Fault Picking.

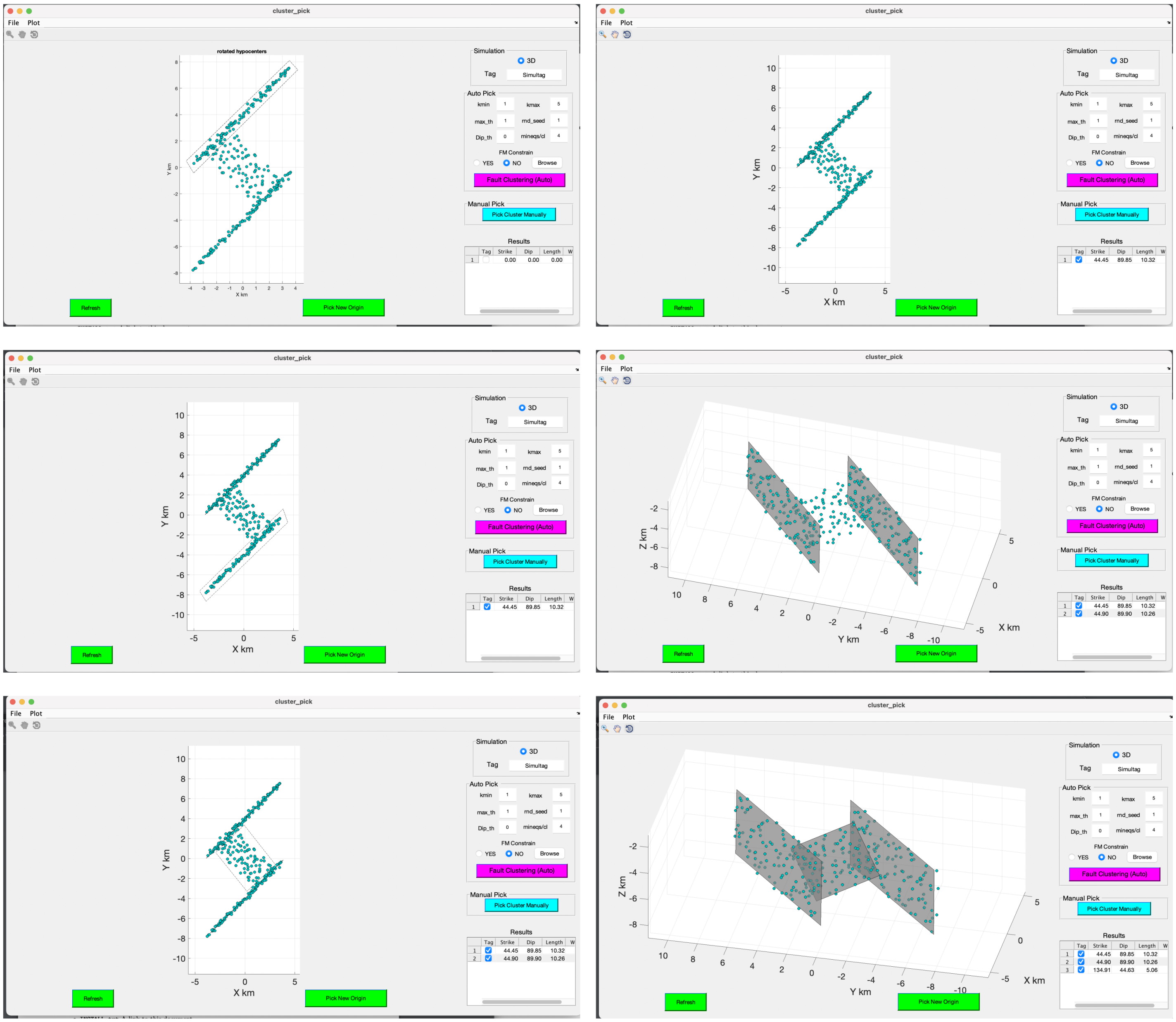


Figure 6: Manual Fault Picking procedures showing the three steps and the corresponding fault planes.

2.3.3 Plotting options and saving results

It may be challenging to associate each fault parameter in the table panel with a particular fault in the figure. **Cluster Pick** allows a turn on and off check button in front of each fault parameter so the user can interact with the fault geometry result in a more practical way. Figure 7 shows examples of this feature and how it enables the user to associate each fault parameter with a specific fault plane in the result.

The user can also plot the resulting fault geometry in four different ways using the **Plot** menu (Fig. 8). The options are **Plot current Density Plot**, **Plot All Seismicity with faults**, **Plot faults Only** and

Plot Clusters and Faults. The figure below shows examples of each option. The user can then save the fault geometry results by clicking the **Write Cluster Model File** under the **File** menu (Fig. 9).

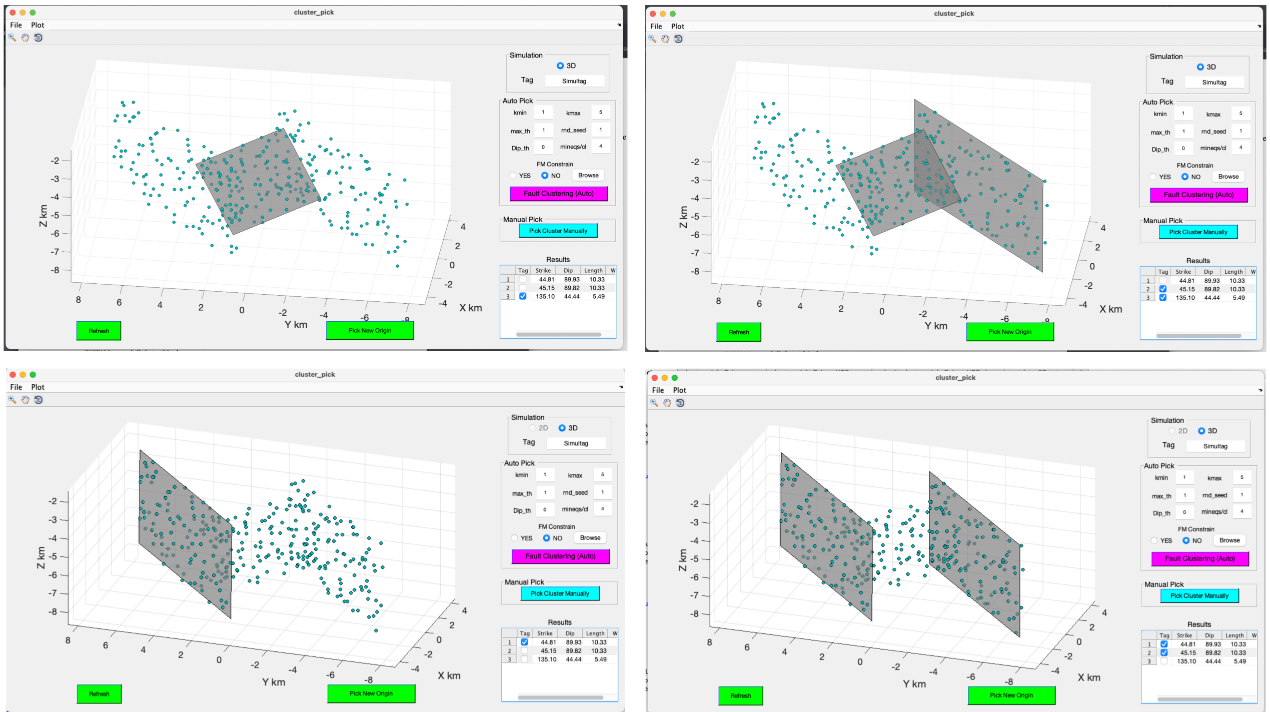


Figure 7: Interacting with the fault geometry

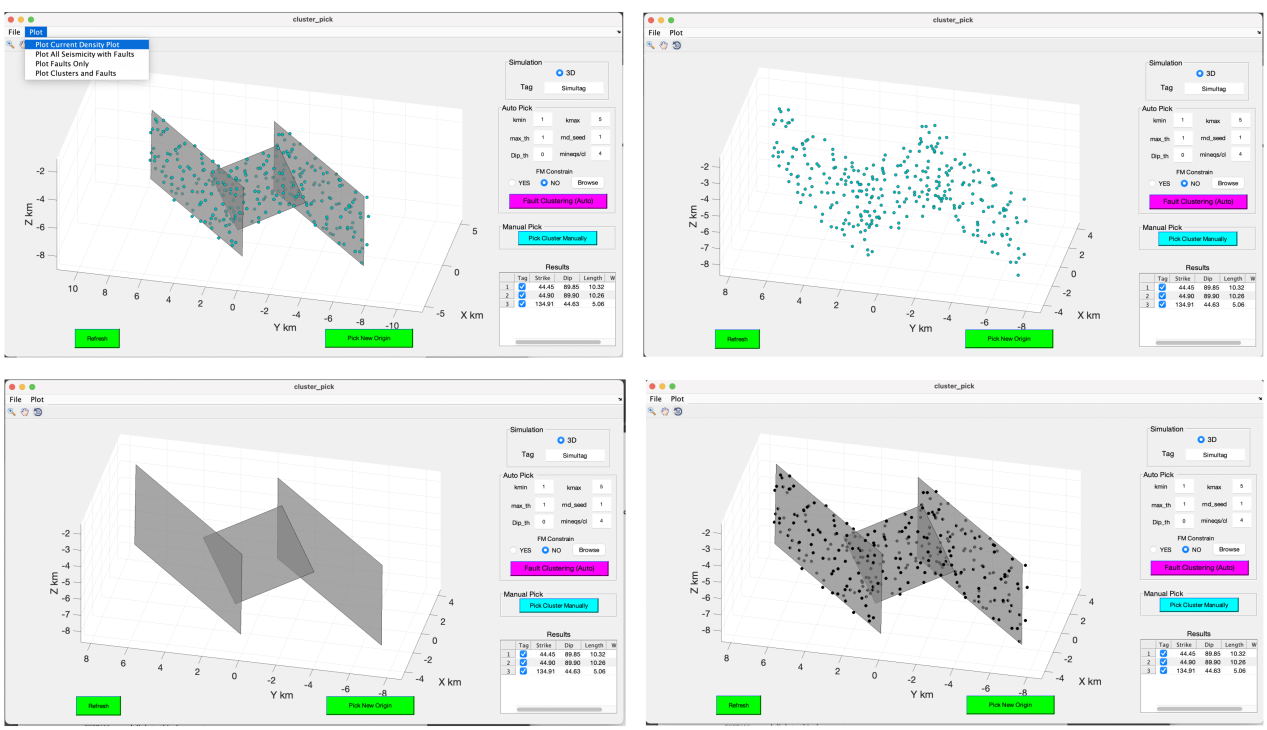


Figure 8: Different plotting options.

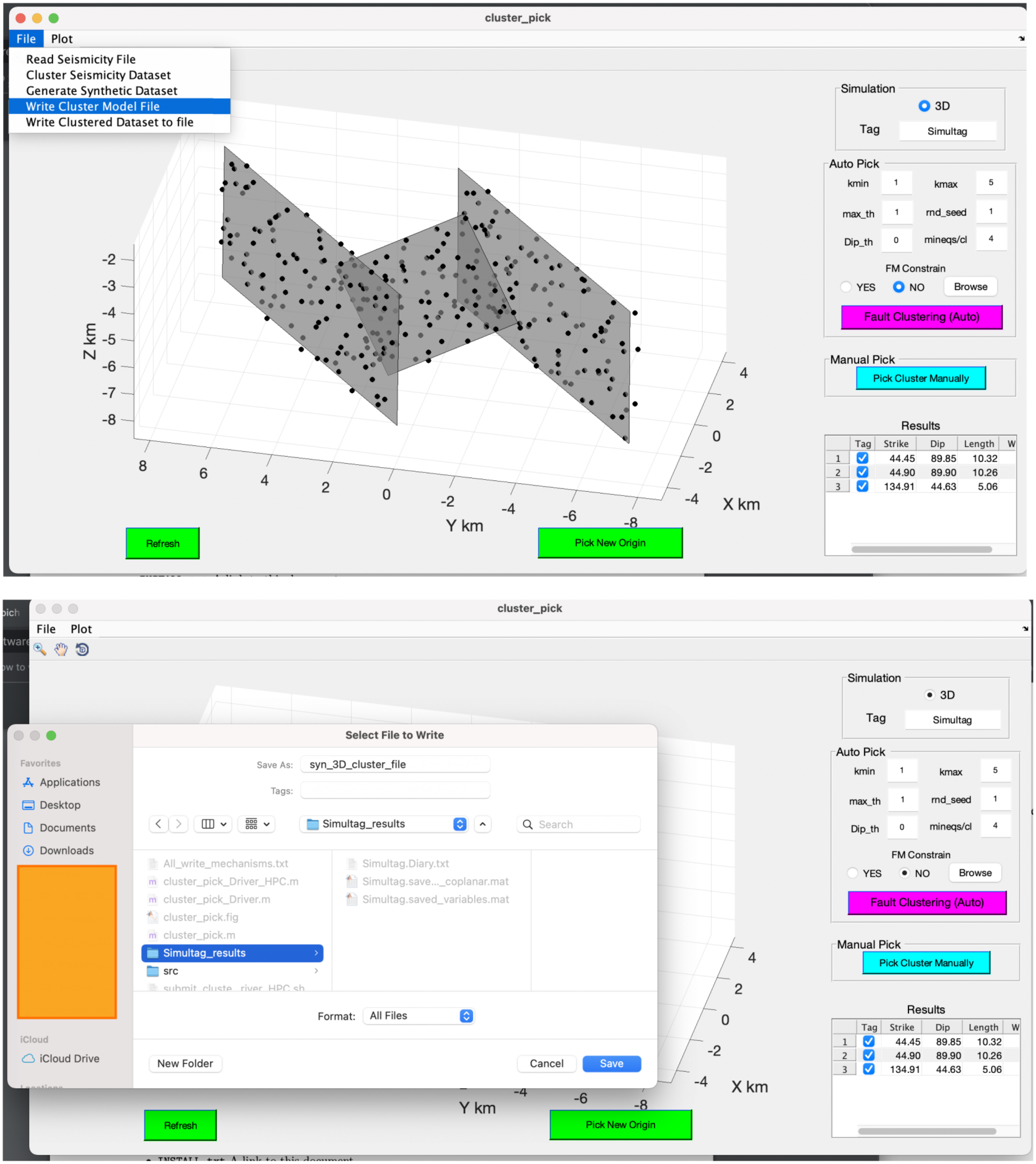


Figure 9: Saving fault geometry result

2.4 Generating Synthetic Seismicity Dataset

Cluster Pick can be used to generate 2D and 3D synthetic datasets using multiple fault parameters similar to the seismic zone of interest. This feature can be opened using the **Generate Synthetic Dataset** under the **File** menu. A separate GUI will appear consisting of a graph panel where the generated synthetic dataset is plotted (Fig. 10). Editable fault parameter fields to input desired parameters, and the faults can be generated and plotted by clicking the **Generate Datasets** button. More faults can be added using the **Add Cluster** button. Default fault parameters can be generated using the **Default Inputs** button.

Figure 11 shows examples of generating 2D datasets using the parameters in the parameter window. The two examples have different cloud thicknesses for the fourth line cluster. Figure 12 shows two examples of generating 3D datasets for two and four faults, respectively. Default parameters can also be used for templates (Fig. 13).

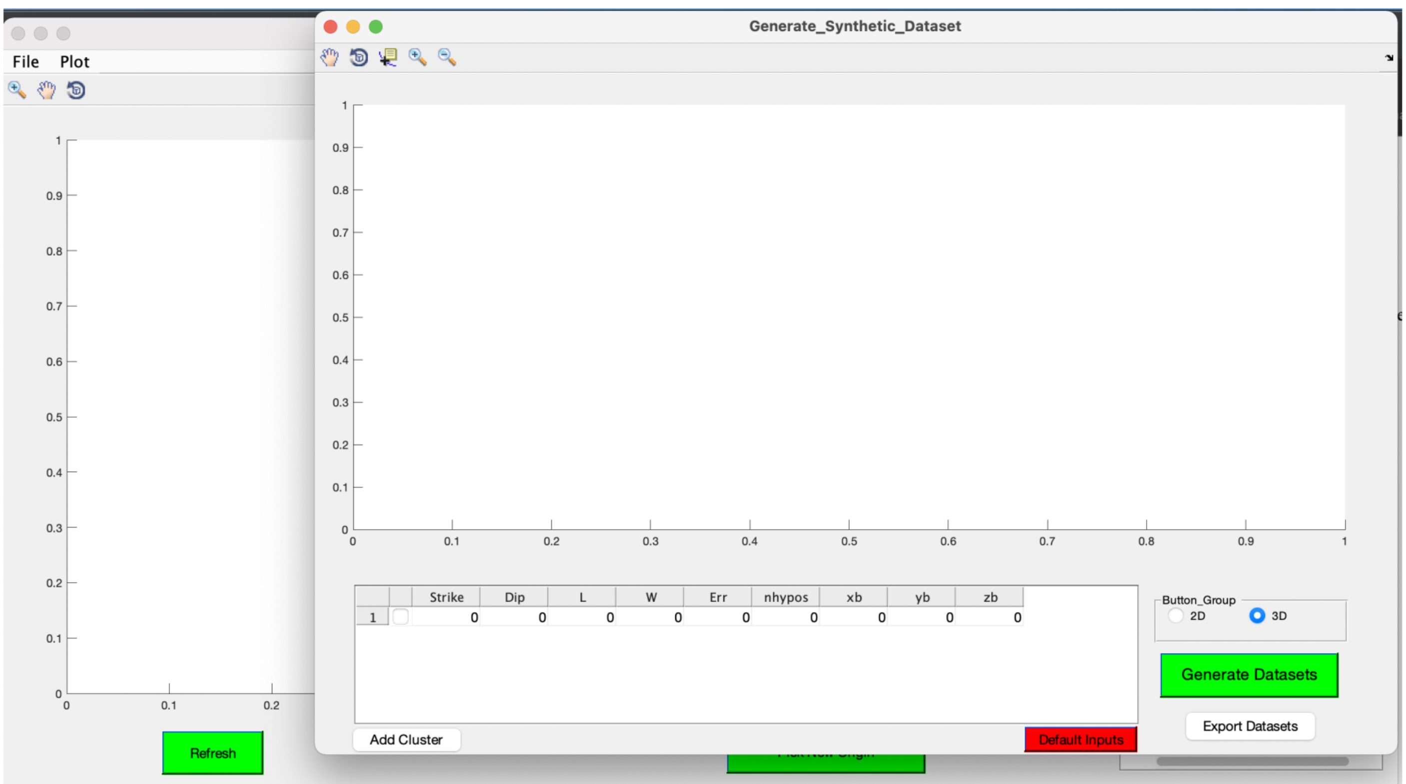


Figure 10: Generating 2D synthetic dataset

2.5 Clustering Seismicity Dataset

Cluster Pick can also be used to perform a cluster analysis to highlight the faults by separating clustered and diffuse seismicity using the cumulative tetrahedra volume method of Ouillon and Sornette (2011). It can be started by clicking the **Declustering Seismicity Dataset** under the **File** menu. This opens a separate GUI containing four panels to display the original and clustered hypocenters, a combination of clustered and diffuse hypocenters and the Cumulative distribution function (CDF) for the original and random hypocenters generated in the process (Fig. 14). Figure 14 also shows how to load the original hypocenters.

For the clustering analyses, a randomized version of the input hypocenters is needed to model the diffuse dataset corresponding to the size and shape of the seismic zone. We then determine the cumulative distributions of the volumes of the observed and randomized catalogs, i.e., N(V) and N0(V0), and separate the diffuse earthquakes from the observed earthquakes by removing all hypocenters in the observed catalog with volumes above a certain volume threshold (Vthresh), noting that the calculated tetrahedra volume for clustered earthquakes is generally less than the tetrahedra volume for diffuse earthquakes. Following Ouillon and Sornette (2011), we assume that the 5% quantile (i.e., probability of 0.05) of the volume distribution of the randomized catalog corresponds to Vthresh and can be used to model diffuse earthquakes. We remove all hypocenters in the input catalog with volumes greater than Vthresh to determine a clustered subset of the observed earthquakes. The resulting clustered earthquakes will be used as input to **Cluster Pick** to generate a fault geometry model.

To generate the randomized catalog, the user will pick a polygon containing the input hypocenters to highlight the extent of the dataset. This is done using **Pick Random Dataset Polygon** on the **File** menu, and a cross hair can be used to pick the polygon on the input datasets display (Fig. 15). After picking the extent of the hypocenters, the algorithm will separate the clustered and diffuse hypocenters and plot the results and the CDFs. Figure 16 shows a better view of the clustered and diffuse hypocenters using the standard rotate button. The default 5% can be changed using the slider bar at the bottom or by typing the desired percentage in the editable field, and the plots of clustered and diffuse hypocenters will be updated on the fly. Figure 17 shows examples of using 10% and an extreme 84.5% values where the majority of the hypocenters are considered clustered. Note that the CDF plots also change. The clustered dataset can then be saved to a file by clicking the **Write Clustered Data** on the **File** menu (Fig. 18).

2.6 Running Cluster Pick on the Terminal or HPC for multiple random seed generator

We provide a version of **Cluster Pick** which enables the user to determine the fault geometry that fits the input datasets using multiple random seed generators in one run on the terminal. To run the terminal version of **Cluster Pick**, open the **cluster_pick_Driver.m** file and edit the input parameters as needed. The results for each random seed generator will be put in different sub-folders with the simulation tag and the seed value as the folder name. The content of the **cluster_pick_Driver.m** file is shown below. The code loops over different random seed generators from 1 to 100 (See line 15). Other seed generator values can be specified in the for loop (e.g., for i in [12, 34, 123, 43] in line 15).

cluster_pick_Driver.m:

```
1  clear all; close all; clc; tic
2  
3  % Add the path to the source codes
4  addpath('src')
5  
6  arg_kmin = 1; arg_kmax=100; arg_err_av = 1;
7  arg_hypo_infile = 'testdata.txt';
8  arg_clus_mineqs = 4; arg_N_loop = 1;
9  arg_dip_threshold = 10; arg_FM_file = '';%'All_write_mechanisms_99.txt'; %'';%
10 arg_PLOT_FLAG0 = 1;   % =0, no plots at all
11 arg_PLOT_FLAG1 = 0;   % =0, no intermediate loop plots of data and planes
12 arg_comb_coplanar = 1; % =0, Do not check or combine coplananr faults
13 arg_plot_avg_FM = 0; % =1, Plot intermediate average FM focalsphere
14 
15 for i= 1:100
16 
17     arg_simul_tag = ['Syn_Faults.i.' num2str(i)];
18 
19     OADC_3D_now_12_4_20(i,arg_kmin,arg_kmax,arg_err_av,...
20             arg_hypo_infile,arg_clus_mineqs,arg_N_loop,arg_simul_tag,...
21             arg_dip_threshold,arg_FM_file,arg_PLOT_FLAG0,arg_PLOT_FLAG1,...
22             arg_comb_coplanar, arg_plot_avg_FM);
23 
24     close all;
25 end
26 toc
```

Cluster Pick can also be used on the HPC using the **cluster_pick_Driver_HPC.m** file. The corresponding submission script (**submit_cluster_pick_Driver_HPC.sh**) contains an example of how **cluster_pick_Driver.m** can be run using 1 to 500 random seed generator values (line 4), submitted in a batch of 50s in order to be considerate to other HPC users.

cluster_pick_Driver_HPC.m:

```
function cluster_pick_Driver_HPC(i)
tic
% Add the path to the source codes
addpath('src')

```

```
arg_kmin = 1; arg_kmax=100; arg_err_av = 2.5;
arg_hypo_infile = 'testdata.txt';
arg_clus_mineqs = 4; arg_N_loop = 5;
arg_dip_threshold = 10; arg_FM_file = 'All_write_mechanisms.txt'; %'';%
arg_PLOT_FLAG0 = 1;   % =0, no plots at all
arg_PLOT_FLAG1 = 0;   % =0, no intermediate loop plots of data and planes
arg_comb_coplanar = 1; % =0, Do not check or combine coplananr faults
arg_plot_avg_FM = 0; % =1, Plot intermediate average FM focalsphere

arg_simul_tag = ['Simul3_Faults.i.' num2str(i)];

OADC_3D_now_12_4_20(i,arg_kmin,arg_kmax,arg_err_av,...
            arg_hypo_infile,arg_clus_mineqs,arg_N_loop,arg_simul_tag,...
            arg_dip_threshold,arg_FM_file,arg_PLOT_FLAG0,arg_PLOT_FLAG1,...
            arg_comb_coplanar, arg_plot_avg_FM);

close all;
toc
```

submit_cluster_pick_Driver_HPC.sh:

```
#!/bin/bash
#SBATCH --partition computeq
#SBATCH --cpus-per-task 1
#SBATCH --array 1-500%50
#SBATCH --time=25-00:00:00
#SBATCH --account=account_name
#SBATCH --job-name=88_20_1

module load matlab/R2018a
matlab -nodisplay -r 'i=${SLURM_ARRAY_TASK_ID};cluster_pick_Driver_HPC(i)'
```

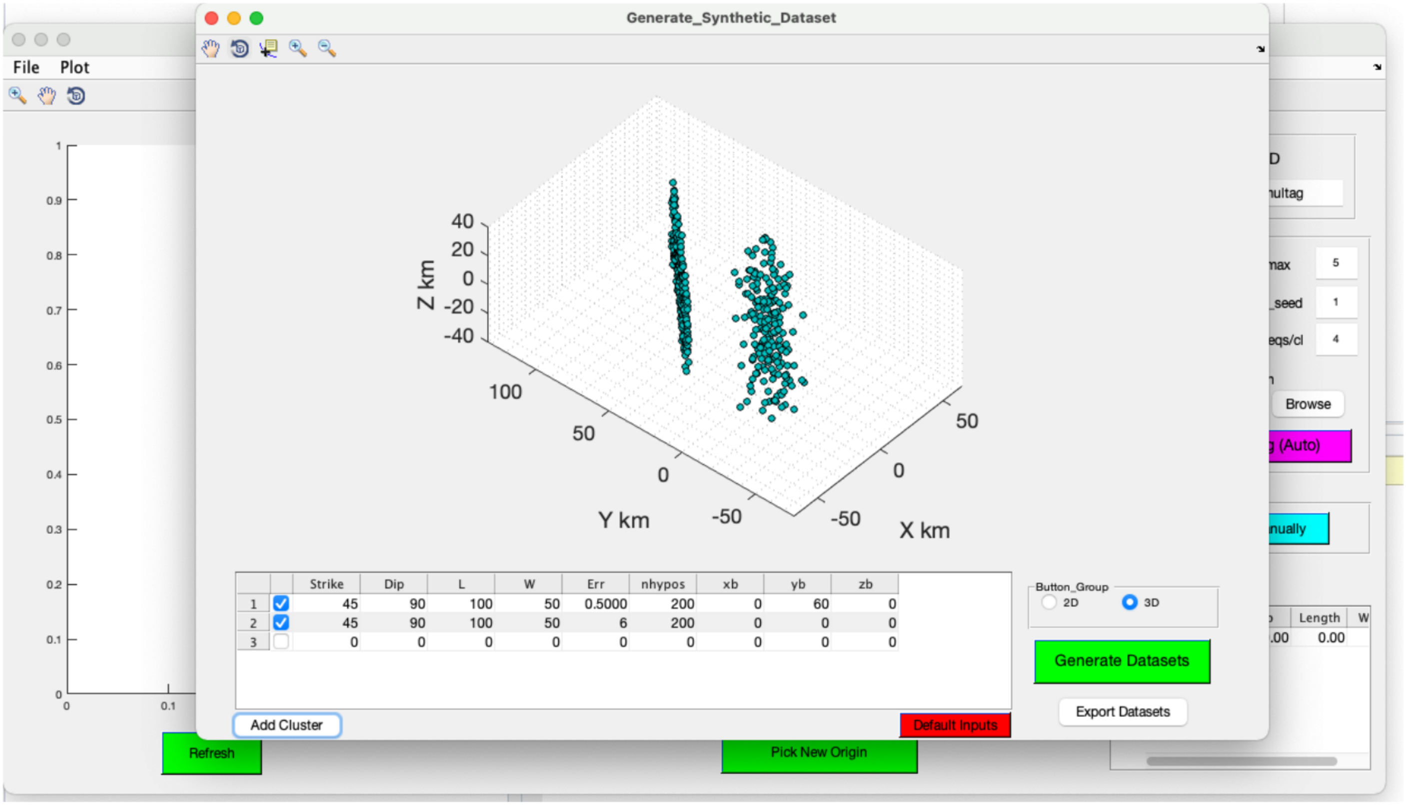
3 Conclusions

We demonstrated how **Cluster Pick** could be used to determine the realistic fault geometry of a seismic zone given the hypocenter locations. It can be run on the terminal or remotely on the High-Performance Computers (HPC) to determine a suite of fault geometries using several random-number generator seeds that fit the input hypocenters. We also demonstrated how **Cluster Pick** could be used to generate 2D and 3D synthetic hypocenters with complicated geometries given the strike and dip, fault centers, and length and widths of the desired number of fault planes. Lastly, we described how **Cluster Pick** could also be used to perform declustering analysis on the original hypocenters using the cumulative tetrahedra volume method of Ouillon and Sornette (2011) to highlight the faults by separating clustered and diffuse seismicity.

4 References

- Fadugba, O. I. (2021). Waveform and Geodynamic Modeling of Seismicity Associated with the Charlevoix Seismic Zone (Doctoral dissertation, The University of Memphis).
- Ouillon G. C., C. Ducorbier and D. Sornette (2008). Automatic reconstruction of fault networks from seismicity catalogs: Three dimensional optimal anisotropic dynamic clustering. Journal of Geophysical Research, Vol 113, No. B01306, https://doi.org/10.1029/2007B005032.
- Ouillon G. and D. Sornette (2011). Segmentation of fault networks determined from spatial clustering of earthquakes. Journal of Geophysical Research, Vol. 116, No. B02306, https://doi.org/10.1029/2010JB007752.

Figure 11: Generating 2D synthetic dataset



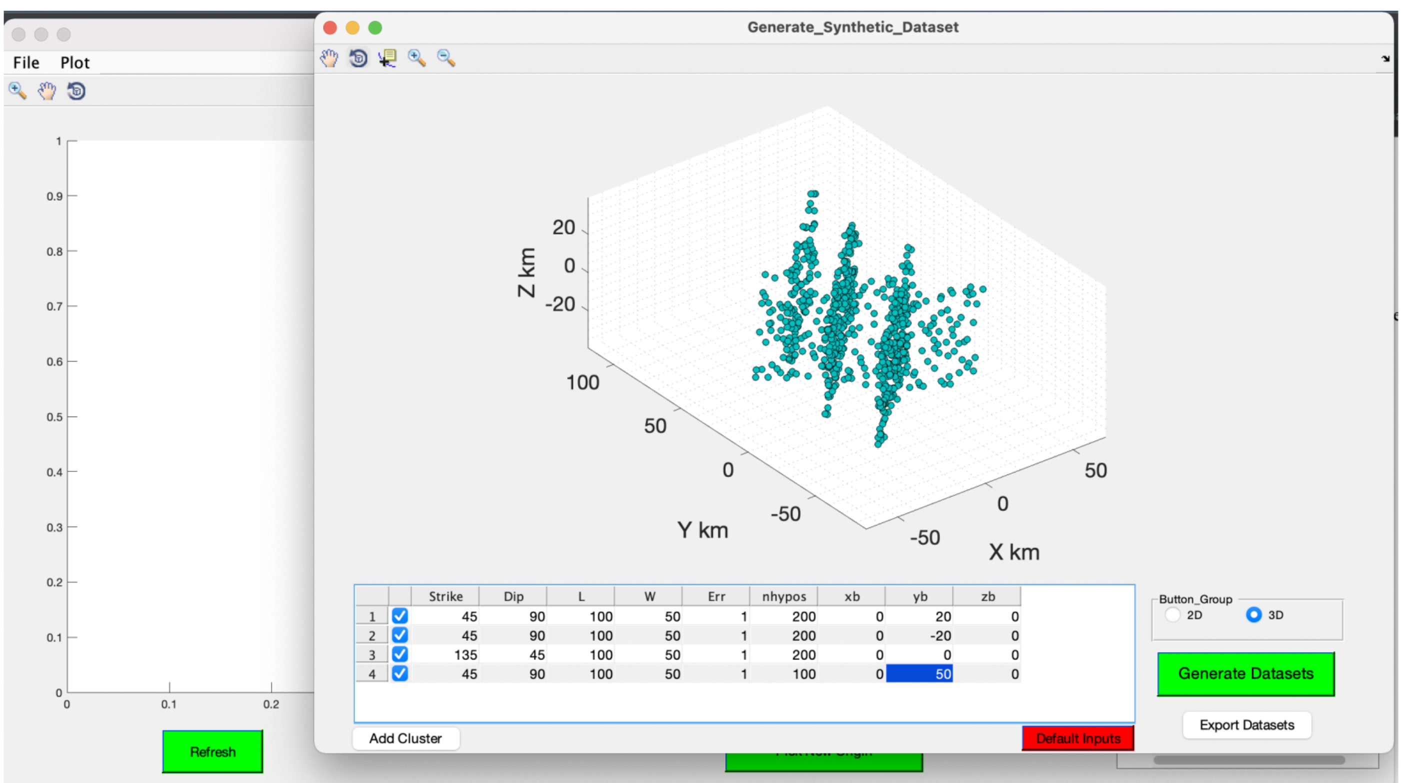


Figure 12: Generating 3D synthetic dataset

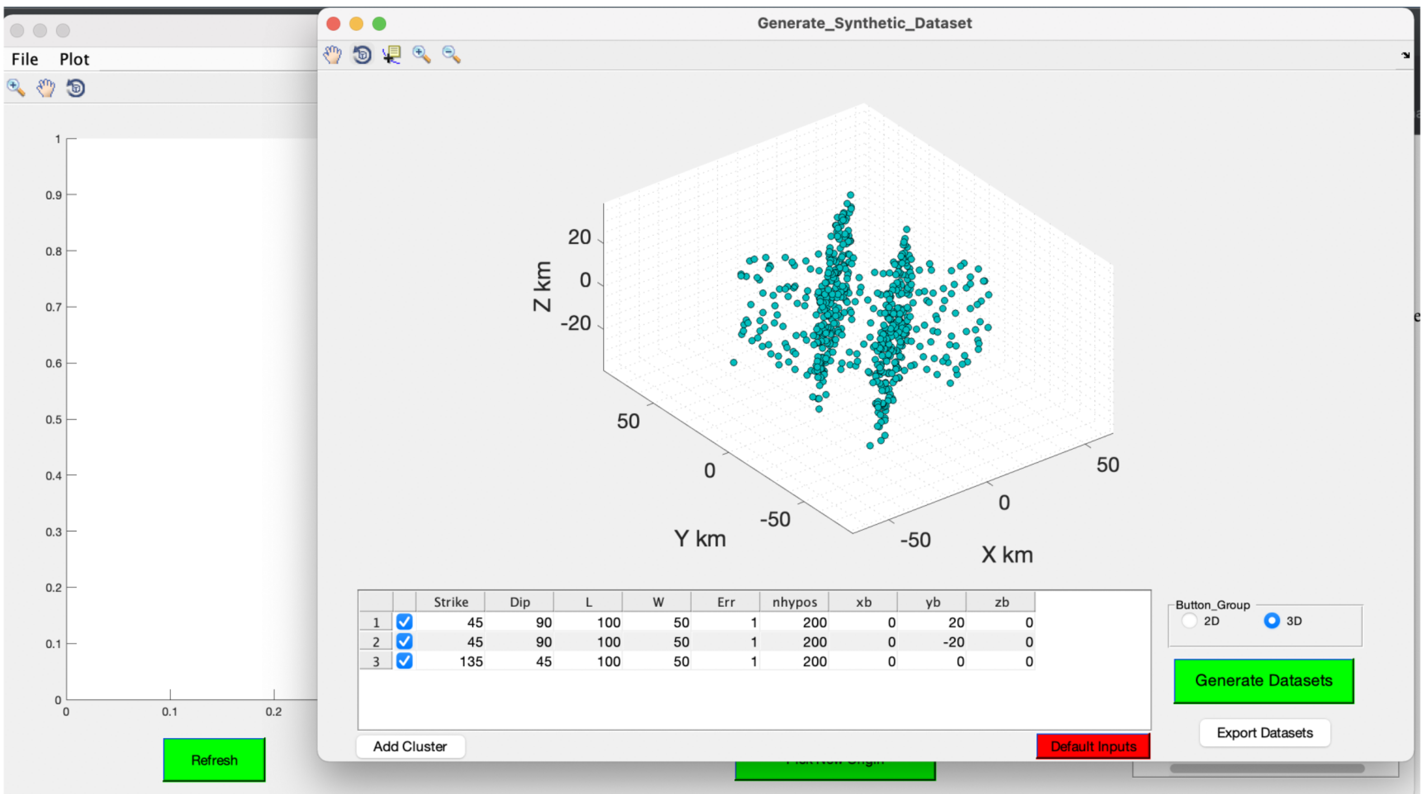


Figure 13: Generating 3D synthetic dataset with default parameters

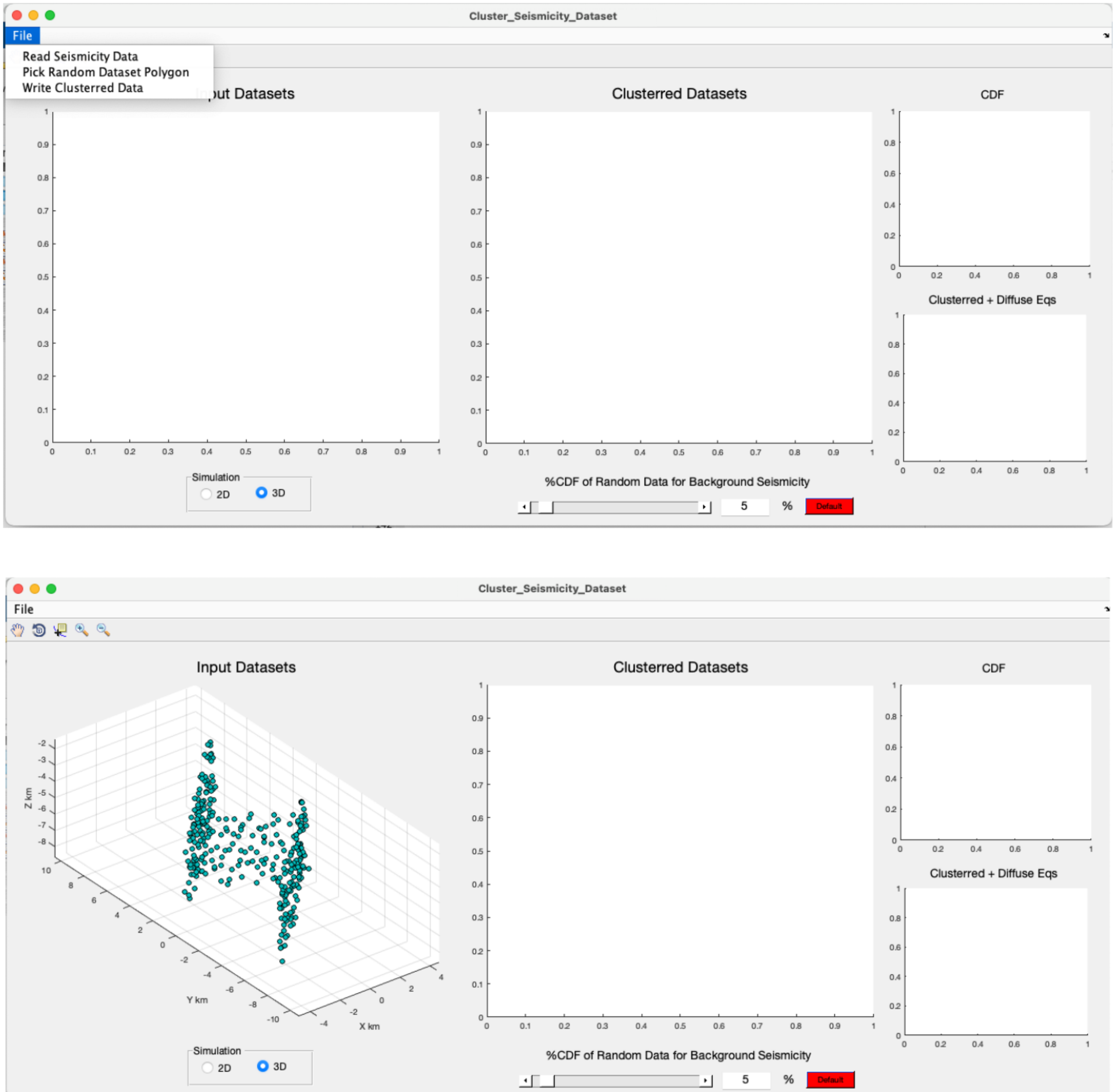


Figure 14: Clustering seismicity dataset

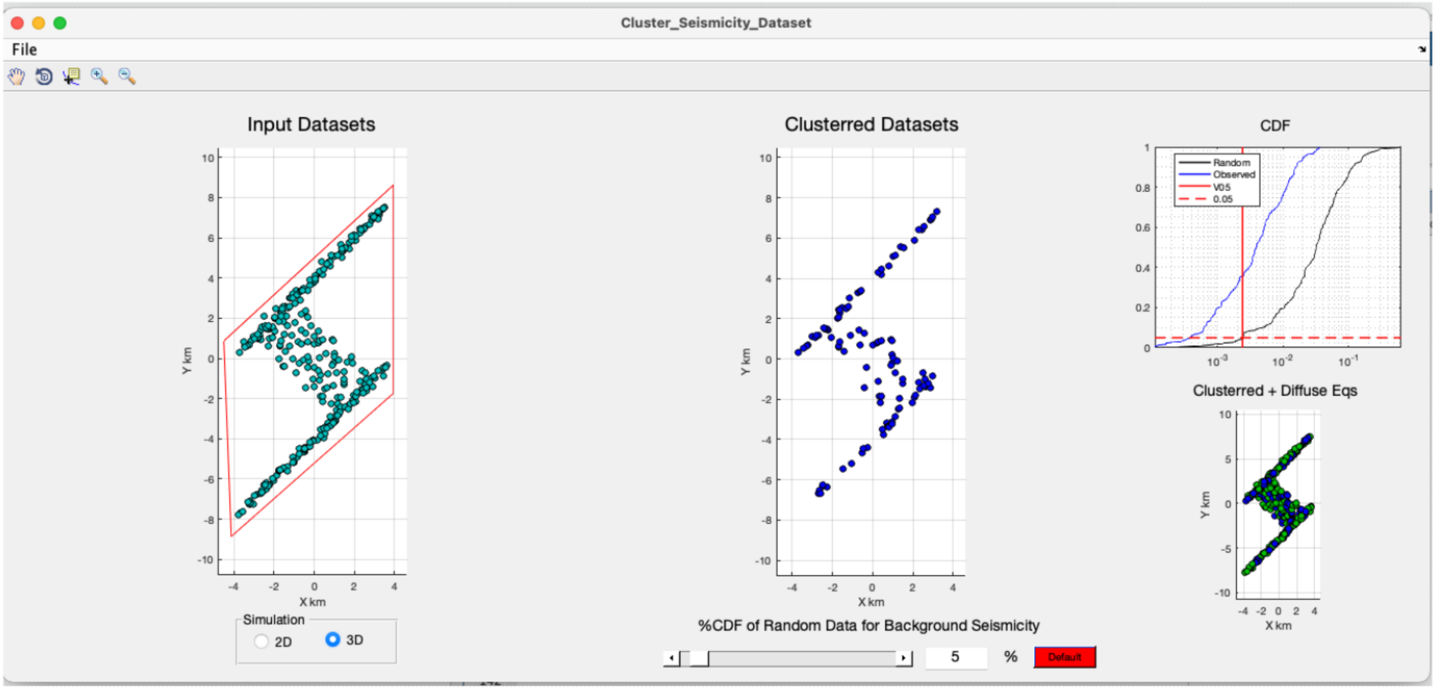


Figure 15: Drawing random dataset polygon

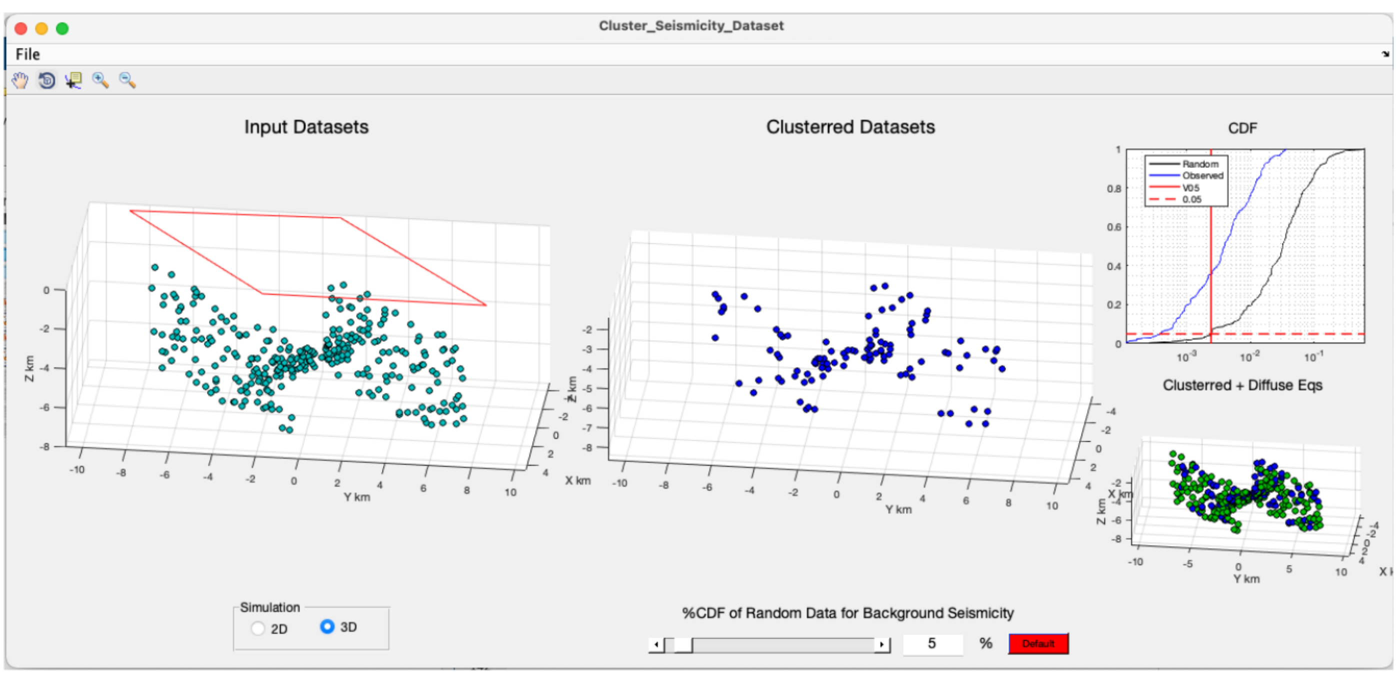
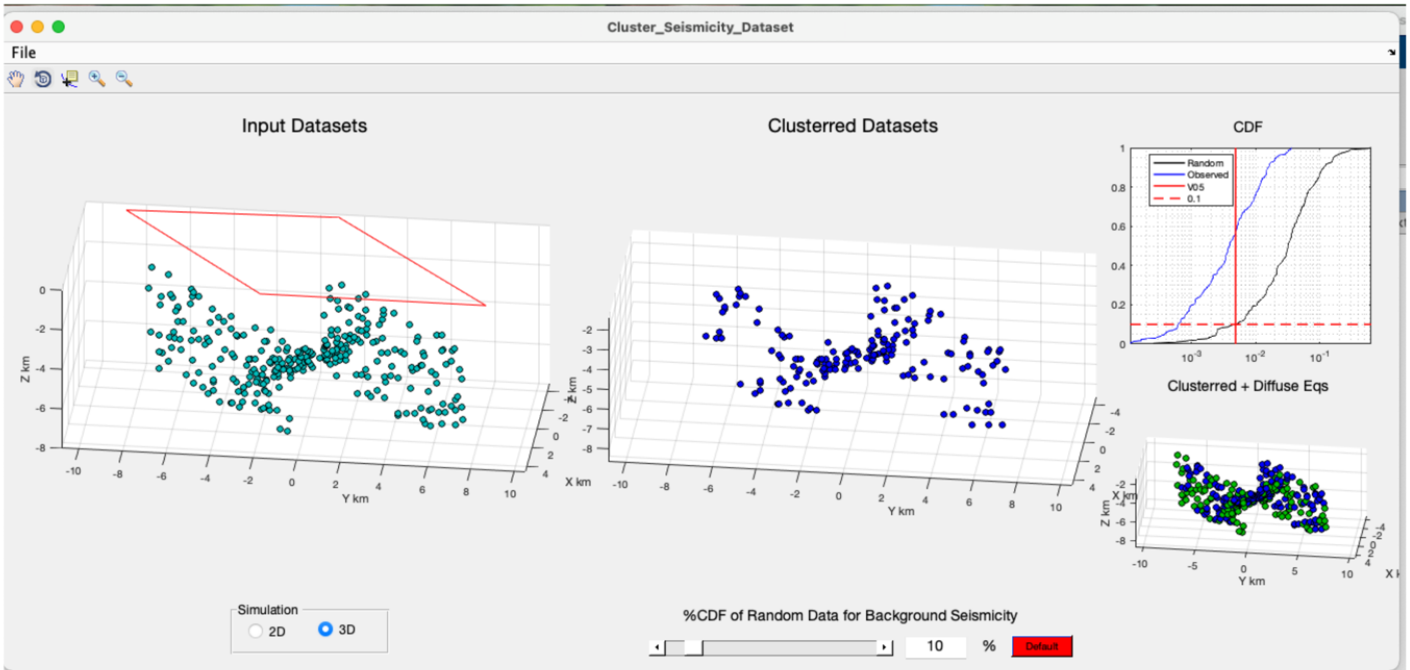


Figure 16: Clustering Seismicity using the default 5% threshold.



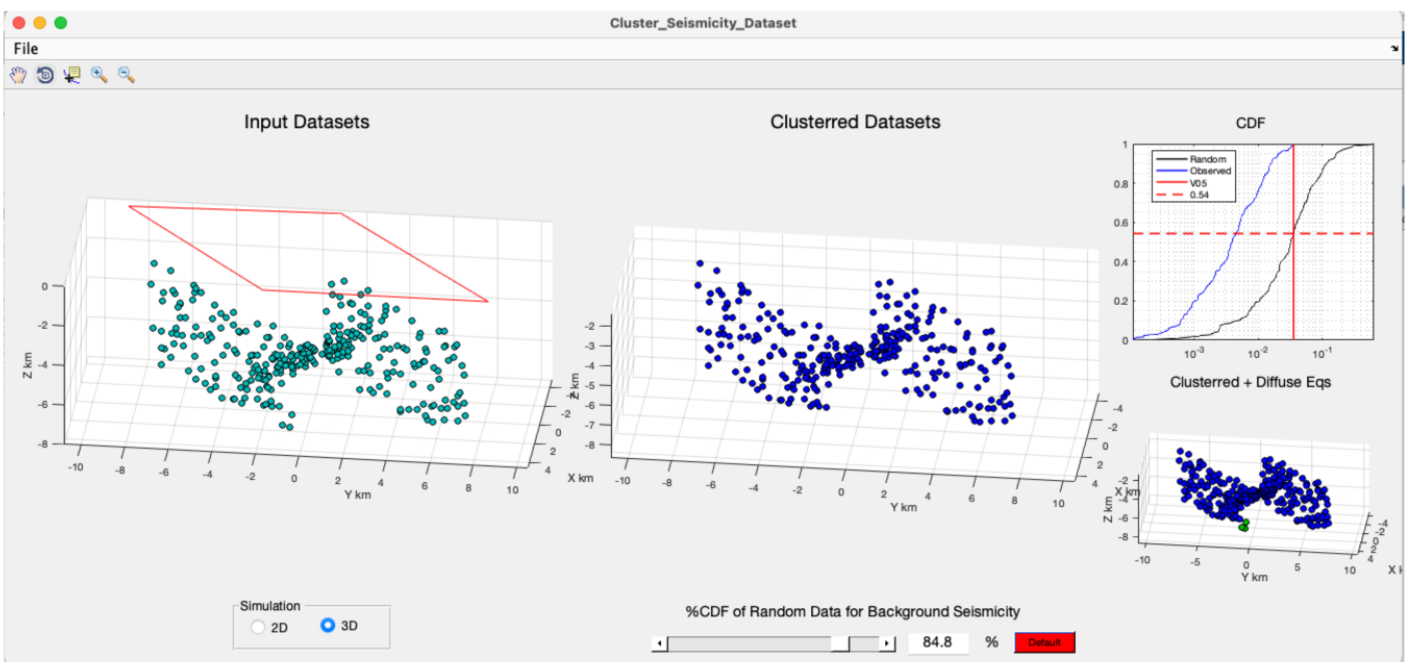


Figure 17: Clustering Seismicity Dataset using different percentages.

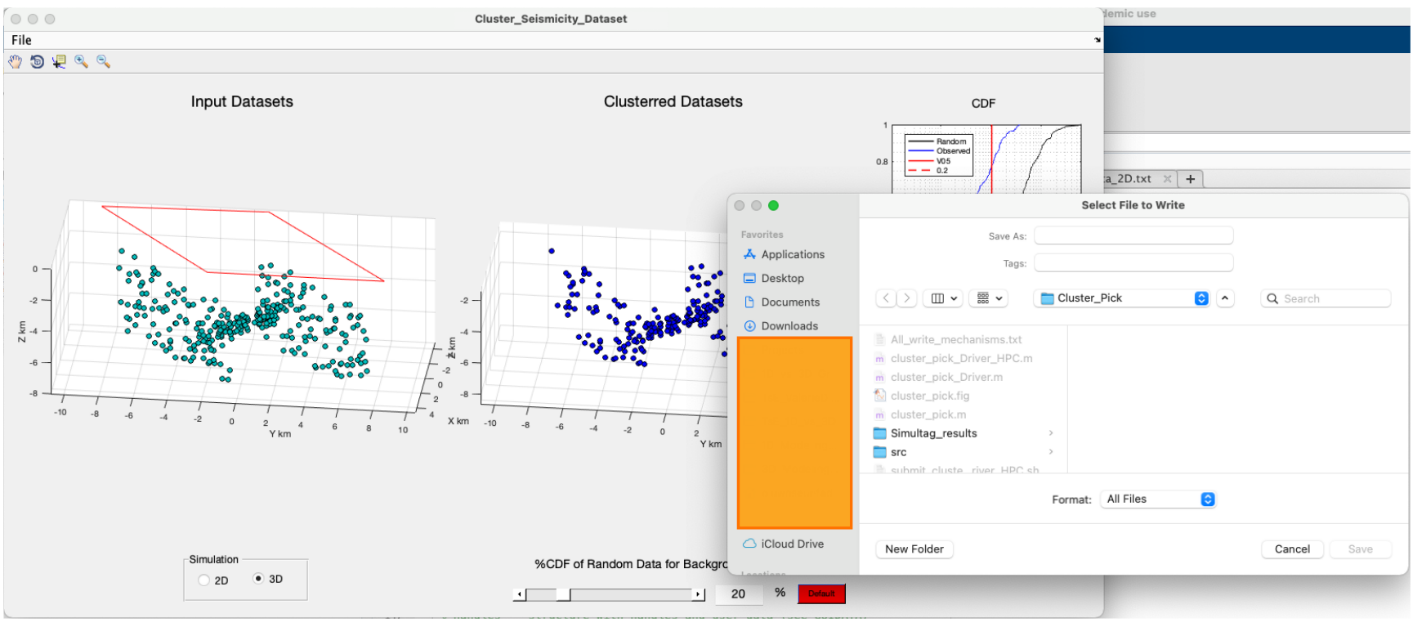


Figure 18: Saving clustered hypocenters.