

SSC-Clustering Tool

The SSC-Clustering Tool provides an implementation and graphical user interface for¹

- Superparamagnetic Clustering
[M. Blatt, S. Wiseman and E. Domany, Physical Review Letters 76, 3251 \(1996\);](#)
- Sequential Superparamagnetic Clustering
[T. Ott, A. Kern, W.-H. Steeb and R. Stoop, Journal of Statistical Mechanics: theory and experiment: P11014, 2005.](#)

Superparamagnetic Clustering is a nonparametric clustering approach that has been applied successfully in a variety of different fields. It is based on the physical properties of an inhomogeneous ferromagnetic model. A data set to be clustered is translated into a ferromagnetic Potts spin model, where clusters appear naturally as regions of aligned spins. Upon an increase of the system temperature T , a natural clustering hierarchy (dendrogram) can be established.

Sequential Superparamagnetic Clustering (SSC) is a substantial extension of the superparamagnetic approach, based on the extraction of clusters and the iterative application of superparamagnetic clustering. The method is able to find the most natural clusters (in a nonparametric sense) even if they are highly different in shape and density.

Further information about the algorithms can be found in the cited articles.

As input to the SSC-Clustering Tool you have to provide a complete similarity matrix, containing the distances between all the data items to be clustered. The matrix must be loaded from a .txt file that contains these single values separated by a space. Detailed instructions on how to use the algorithm can be found below.

The SSC-Clustering Tool is provided as .exe file for Windows and Linux.

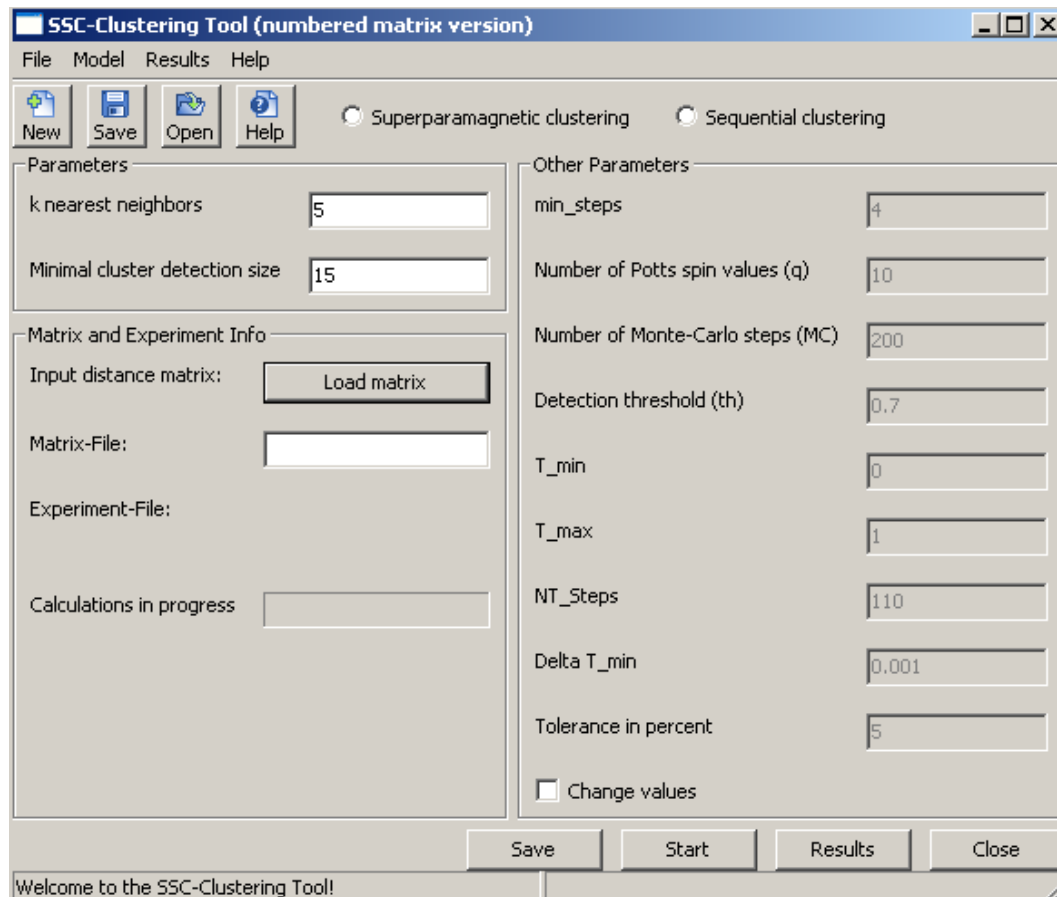
This project has been supported by

- Institute of Applied Simulation, Zurich University of Applied Sciences
- Ethikzentrum, University of Zurich
- Stoop Group, ETH Zurich
- Cogito Foundation, Wollerau Switzerland

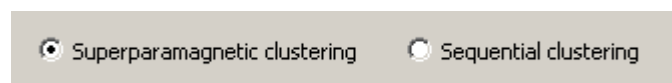
¹ Please cite these articles when publishing results obtained with this tool

Guideline: How to use our clustering tool

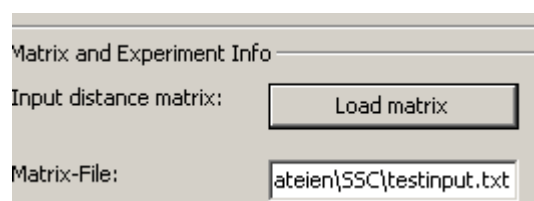
Step 1: Open the GUI



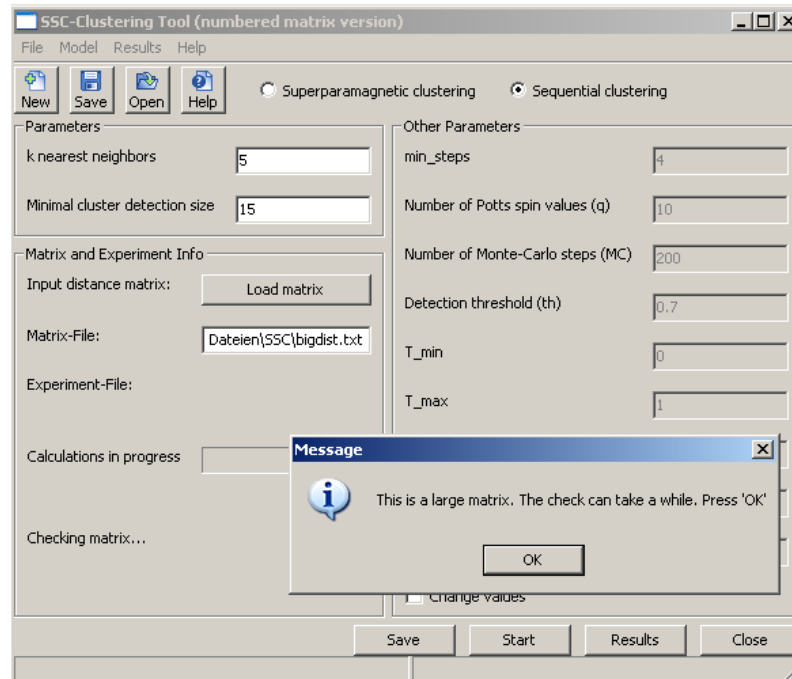
Step 2: Choose a method



Step 3: Load the data matrix



The validity of the matrix will be checked. For large matrices this process can take a while. In this case, the tool will inform you and will request your OK.



The conditions for validity are:

- Dimension (square matrix or square matrix with numbered lines)
- Type of entries (floating point numbers separated by a space)
- The size of the matrix must not exceed 5000 x 5000

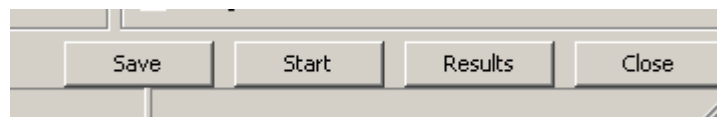
Step 4: Change the parameters if needed. If you do not change the parameters, the default values are used.

List of parameters:

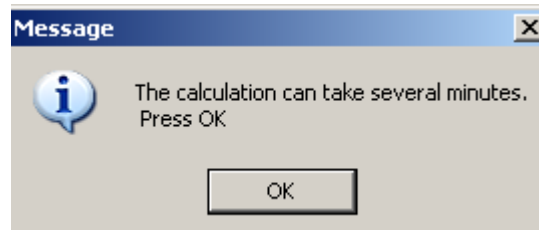
| | |
|--|--|
| k nearest neighbors (integer between 1 and N-1) | defines the connectivity: each point is connected to its k nearest neighbors. Default value: 5 |
| Minimal cluster detection size (integer between 0 and N-1) | defines the granularity for the detection of clusters. Default value: 15 |
| min_steps (integer larger than 0) | defines the minimal number of temperature steps for clusters to be detected. Default value: 4 |
| Number of Potts spin values (q) (integer larger than 0) | defines the number of Potts spin values used. Default value: 10 |
| Number of Monte Carlo steps (MC) (integer larger than 0) | defines the number of Monte Carlo steps used. Default value: 200 |
| Detection threshold (th) (float between 0 and 1) | defines the pair correlation threshold for points to be assigned to the same cluster. Default value: 0.7 |
| T_min (nonnegative float) | defines the lower bound of the temperature interval. Default value: 0 |
| T_max (nonnegative float larger than T_min) | defines the maximal upper bound of the temperature interval. Default value: 0 |
| NT_steps (integer larger than 0) | defines the partition of the temperature interval (maximal number of temperature steps). Default value: 110 |
| Delta T_min (nonnegative float) | defines the minimally possible binning of the temperature axis. Default value: 0.001 |
| Tolerance in percent (float between 0 and 100) | defines the tolerance for the fluctuations of the size of a cluster (in %). Default value: 5 |

A change of the default values can influence the results slightly. You may use these options for either a quicker run of the algorithm or for improving the clustering resolution.

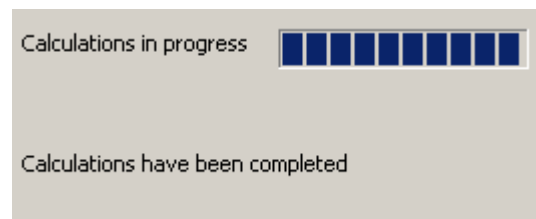
Step 5: Save your experiment and start the calculations



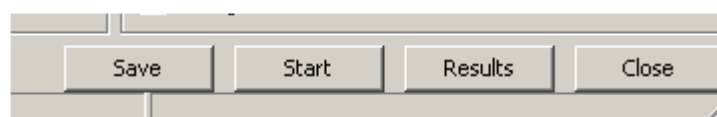
The algorithm is started. For large matrices the calculation can take a while. In this case you will be informed by the tool. Press 'OK' if you want to continue.



A message will inform you when the analysis is completed.

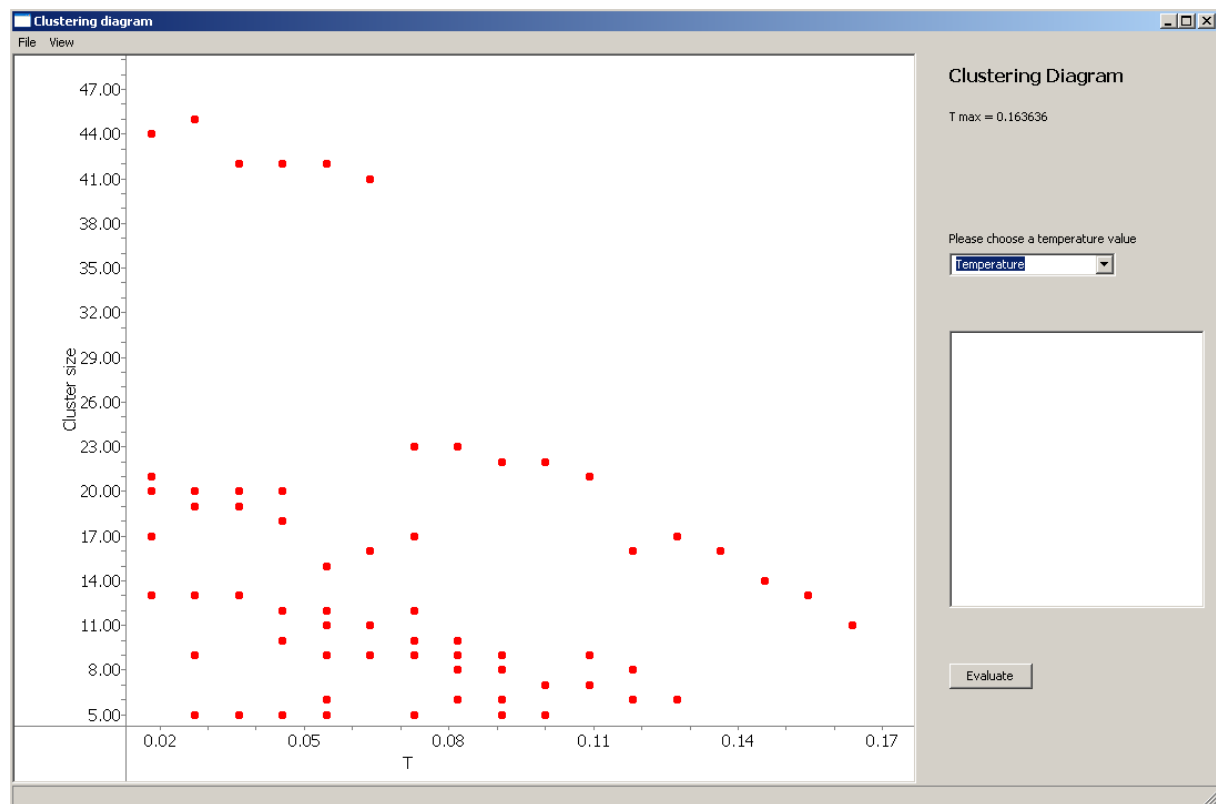


Step 6: You can access the results via the 'Results' button.



Superparamagnetic Clustering

The results will be summarized in a diagram that shows the size of the occurring clusters vs. the temperature T .



When selecting a temperature on the right side, the sizes of the occurring clusters are listed.

Please choose a temperature value

0.036364

At temperature = 0.036364 there are

- 1 cluster(s) with size 42
- 1 cluster(s) with size 20
- 2 cluster(s) with size 19
- 1 cluster(s) with size 13
- 2 cluster(s) with size 5
- 1 cluster(s) with size 2
- 28 cluster(s) with size 1

Evaluate

Press 'Evaluate' to access the detailed composition of the clusters. The results are automatically stored as pdf-file and as .txt-file.

spmc_experiment_T0.036364.pdf - Adobe Reader

Datei Bearbeiten Anzeige Dokument Werkzeuge Fenster Hilfe

1 / 2 113% Suchen

Superparamagnetic Clustering

Parameters:

| | |
|--|---------------------|
| k nearest neighbors = 5 | T_min = 0 |
| Minimal cluster detection size = 15 | T_max = 1 |
| Number of Potts spin values (q) = 10 | NT_Steps = 110 |
| Number of Monte-Carlo steps (MC) = 200 | Delta T_min = 0.001 |
| Detection threshold (th) = 0.7 | min_steps = 4 |

Matrix file: C:\Dokumente und Einstellungen\lotti\Eigene Dateien\SSC\testinput.txt
 Experiment file: C:\Dokumente und Einstellungen\lotti\Eigene Dateien\SSC\spmc_experiment.txt

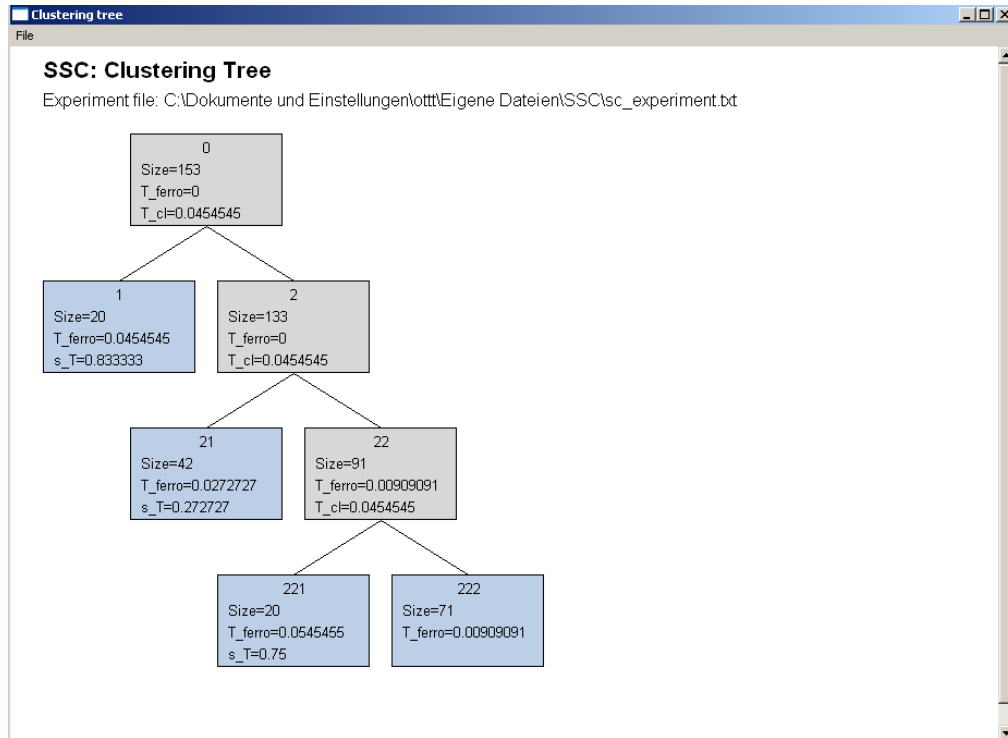
Results:

T = 0.036364

| Cluster number | Cluster size | Members (Points Nr) |
|----------------|--------------|---------------------|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 5 | 3, 8, 9, 10, 13 |
| 4 | 1 | 4 |
| 5 | 1 | 5 |
| 6 | 1 | 6 |
| 7 | 1 | 7 |

Sequential Clustering

After performing a sequential clustering, the resulting clustering tree will pop up.



At the same time the detailed results will be written into a .pdf and a .txt file.

The screenshot shows the Adobe Reader window displaying the PDF file "sc_experiment.pdf". The document content is as follows:

Sequential Clustering

Parameters:

| | |
|--|---------------------|
| k nearest neighbors = 5 | T_min = 0 |
| Minimal cluster detection size = 15 | T_max = 1 |
| Number of Potts spin values (q) = 10 | NT_Steps = 110 |
| Number of Monte-Carlo steps (MC) = 200 | Delta T_min = 0.001 |
| Detection threshold (th) = 0.7 | min_steps = 4 |
| Tolerance = 5% | |

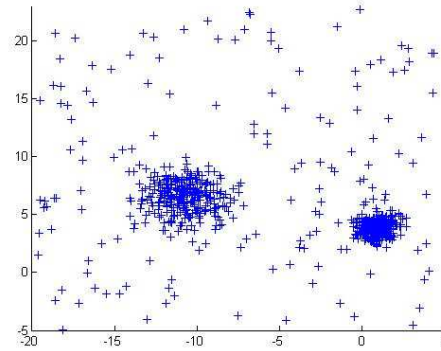
Matrix file: C:\Dokumente und Einstellungen\ottt\Eigene Dateien\SSC\testinput.txt
Experiment file: C:\Dokumente und Einstellungen\ottt\Eigene Dateien\SSC\sc_experiment.txt

Results:

| Cluster | Size | T ferro | T cl | ST | Members |
|---------|------|-----------|-----------|----------|---|
| 0 | 153 | 0 | 0.0454545 | | - |
| 1 | 20 | 0.0454545 | | 0.833333 | 106, 107, 108, 109, 110, 111, 112, 113, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126 |
| 2 | 133 | 0 | 0.0454545 | | - |
| | | | | | 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 127, 128, 129, 130, 131, 132, 133 |

Example

To illustrate the SSC-Clustering Tool we consider a 2-dimensional toy system with two clusters embedded into background noise. The data set with 750 points can be created with the attached MATLAB program.



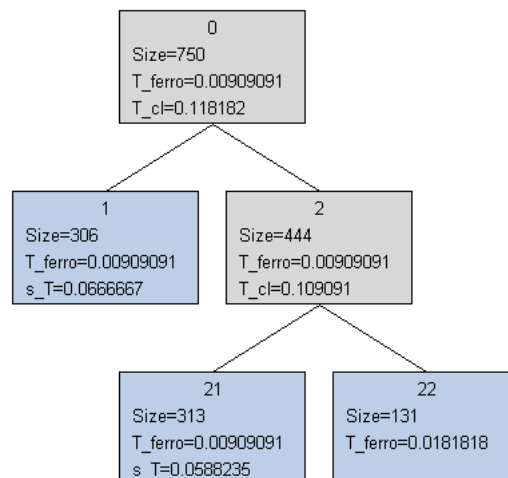
The Euclidian distance of all points is then calculated and the data is stored in the file *2dtoy.txt*. The file should look as follows:

```
2dtoy.txt - Editor
Datei Bearbeiten Format Ansicht ?
0 1.1403 0.91415 1.1783 1.2297 0.85719 0.90954 0.894 0.79506 1.3005 1.2543 2.0532 0.48616 1.8152 1.5796 1.5251 0.40683 0.9371 1.0246 1.5669 1.1337
0.96948 1.8298 0.48822 2.2454 0.71204 1.6524 0.918 1.2998 1.3697 1.4693 1.9697 0.66786 0.20322 1.8941 1.1146 0.9574 0.16083 1.4447 0.84233 1.2894 1
55 0.50716 0.94287 0.48334 1.1504 0.86488 0.99243 0.65272 3.4522 1.4746 0.38658 1.7474 1.8537 0.37933 1.6132 0.74881 0.48618 1.0891 12.848 13.313 1
12.91 12.202 14.289 12.642 9.5936 13.707 12.596 10.347 13.692 12.597 13.617 12.795 12.372 12.505 12.773 13.151 13.522 12.253 14.019 12.165 15.924 1
13.765 13.681 13.45 12.505 12.2 9.9092 15.268 10.488 13.76 13.007 12.706 15.97 12.789 14.877 11.826 13.827 12.464 14.517 11.455 12.77 14.39 4.2249
23.295 22.593 18.662 10.053 20.504 21.556 16.966 25 2.5685 4.4109 5.6607 5.8861 19.018 14.167 5.6431 19.066 20.904 21.393 4.7174 14.981 20.01501.14
9319 2.1347 2.9441 2.4201 1.4482 0.74631 2.8461 1.986 2.0179 0.51103 1.8746 0.74608 1.3324 2.0791 1.8226 1.6839 1.5951 2.6392 1.1611 1.6779 1.4701
474 1.4983 1.708 1.5497 1.0703 1.8417 1.1432 2.364 2.4802 1.4096 0.82069 0.93031 1.7273 1.8218 1.2557 1.6963 3.4625 1.1972 0.76151 1.3786 2.0065 0.
.637 12.115 13.954 14.77 14.781 12.309 13.026 13.175 13.744 13.004 15.079 13.45 10.273 14.283 13.444 11.109 14.411 13.51 14.299 15.686 13.272 13.36
.508 11.824 13.527 14.896 13.783 10.82 13.778 14.499 14.542 14.382 13.315 12.964 10.699 15.998 11.165 14.548 13.788 13.419 16.708 13.394 15.636 12.
81 21.268 24.745 6.7869 17.792 20.389 5.3154 10.111 24.303 23.569 19.099 11.138 21.377 22.652 17.891 26.079 3.6433 5.2375 6.7945 6.1971 20.157 15.0
67 2.3638 1.4755 0.29256 0.54175 1.1167 0.80621 0.79486 0.99556 1.0905 1.0907 0.65296 0.91896 0.74584 0.83467 0.85713 0.94669 1.1809 0.96151 0.9285
.0278 0.5851 1.7541 0.16138 1.085 0.33632 1.3645 0.5852 2.4062 0.94366 1.0107 0.63244 0.38083 0.6676 1.382 1.7586 0.61199 1.0981 1.0466 0.496 0.698
2 12.443 12.579 12.168 11.771 11.24 12.638 9.9449 11.202 12.886 12.25 13.247 11.135 13.419 13.988 12.727 13.526 11.358 14.493 10.528 11.934 12.009
```

We perform a sequential clustering using the default parameters:

SSC: Clustering Tree

Experiment file: C:\Dokumente und Einstellungen\ottt\Eigene Dateien\matlab-dateien\ssc_experiment.txt



Two real clusters (sizes 306 and 313) and a clutter of background points (size 131) have been found. The detailed results are written into the file *sc_experiment.pdf* (and *sc_experiment.txt* respectively). We may use this information to illustrate the clusters using MATLAB again.

