MEPSA 1.2 user manual

Iñigo Marcos-Alcalde

February 2018

Contents

Introduction

MEPSA (Minimum Energy Path Surface Analysis) is a small GUI-based tool for the analysis of energy landscapes (called maps in MEPSA). Some of its most relevant features are: identification of all the barriers and minima (called nodes in MEPSA) of the landscape at once, description of maxima edge profiles, detection of the lowest energy path connecting two minima and generation of transition state theory diagrams along these paths.

It offers a built-in plotting system and most of the data obtained can be easily saved into column formatted plain text files for further processing or plotting using other software.

To follow this manual a test map is provided in the MEPSA downloads page:

http://bioweb.cbm.uam.es/software/MEPSA/

Availability and contact

MEPSA can be downloaded from:

License **GPLv3 (See the LICENSE file distributed with MEPSA)**

Contact: pagomez@cbm.csic.es and imarcos@cbm.csic.es

Citation:

Iñigo Marcos-Alcalde, Javier Setoain, Jesús I. Mendieta-Moreno, Jesús Mendieta, and Paulino Gómez-Puertas. (2015) MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* 31: 3853-3855

Requirements and Installation

To run MEPSA 1.2 you need python 3.4.x installed in your system as well as the following packages:

- tkinter
- tkinter.constants
- tkinter.filedialog
- numpy
- tkinter.messagebox
- sys
- matplotlib
- tkinter.simpledialog
- math
- tkinter.font

An easy cross-platform (Linux, Windows and Mac OS) solution to get all the required packages is the Anaconda platform by Continuum Analytics from https://www.continuum.io/. Just download the python 3.4.x version specified for your operative system and install it. Then open a terminal window (in Linux or Mac OS) or an "Anaconda Prompt" (in Windows) and run the following lines:

- conda install numpy

- conda install matplotlib

To run the program just call the python interpreter in Anaconda folder using mepsa_1.2_3.4.x.py as argument. For Windows users, which may be less used to terminal handling, there are some options to try when calling mepsa_1.2_3.4.x.py:

- Use "Open with" over mepsa_1.2_3.4.x.py and select pythonw.exe (present in the Anaconda installation folder in case it is not directly visible).

- Run "python mepsa_1.2_3.4.x.py" without the quotes from "Anaconda Prompt" (which should be accesible from start menu) after navigating to the mepsa_1.2_3.4.x.py location.

- Create a .bat file and write "your_Anaconda_instalation_path\pythonw.exe mepsa_1.2_3.4.x.py" without the quotes. Save and run it.

For linux users, if Anaconda is not to be used, in a standard Ubuntu installation you should only have to install the following packages:

- sudo apt-get install python3-python3-tk

- sudo apt-get install python3-numpy

- sudo apt-get install python3-matplotlib

Auto-plot

This option enables/disables plots to be automatically generated after the use of most of the MEPSA functionalities. For example, if auto-plot is enabled a plot of the map will be automatically generated after loading a new map and after every modification. The automatic plots resulting from the use of the "connectivity analyses" window functionalities vary in their content depending on the tool used.

Loading/Unloading a map file

In order to load a map click the "Load map file" button, find the map file, and click open.

To unload a map file just use the "Unload map file" button.

5.1 Supported map format

The supported format for the map files consists on three column formatted plain text files using either space or tab symbols as separators. The three columns will be interpreted as X, Y and Z coordinates respectively. If you have format problems check that there are no headers or any other kind of special line.

Currently MEPSA is designed to work only with uniformly distributed rectangular maps, but expect this to change in future versions.

Some of the mechanics present in MEPSA are thought for smooth maps with a small number of minima. More complex maps can still be used but it might be harder to obtain useful information from some of the analyses (e.g. global connectivity analysis).

Plot map

In order to plot the map that MEPSA is currently using, click the "Plot map" button (Fig. 1).

Take into account that, if any modifications are done to the map through the map editor, MEPSA will be working with the latest modification done (see Chapter 7: Map editor).
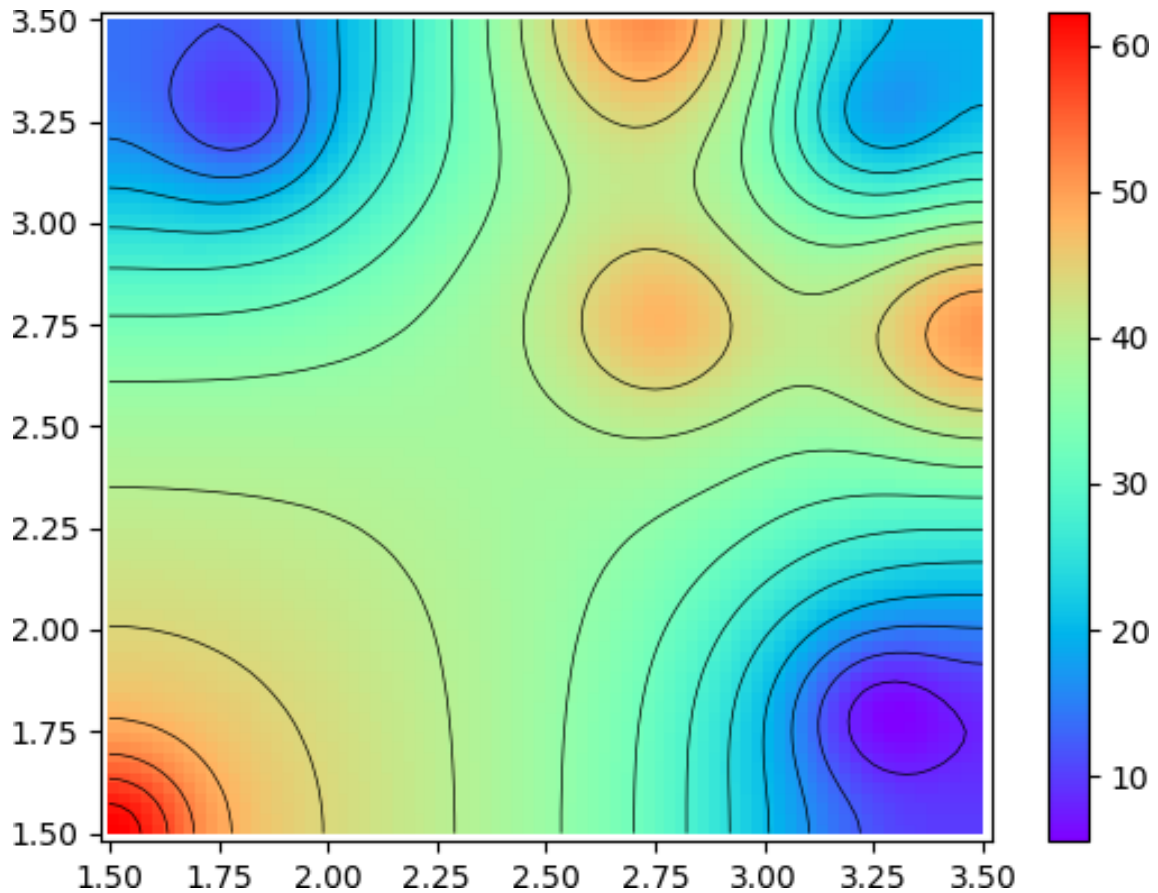

Figure 1: Example of the output of the "Plot map" button.

Map editor

To access the map editor use the "Map editor" button. The map editor window is divided in three sections:

- MODIFY/INVERT MAP
- SMOOTH MAP
- UNDO/SAVE

When "MODIFY MAP", "INVERT MAP" or "SMOOTH MAP" buttons are used, MEPSA stores the modified map as a new copy in memory. This way there will be as many stacked copies of the map as modifications have been done. MEPSA will always work with the latest one. "UNDO" button deletes the last version of the map and therefore makes MEPSA to work with the immediately previous one.

7.1 Modify map

There are two working modes for the "MODIFY MAP" button, "Rectangle" (Fig. 2) and "Circle" (Fig. 3). "MODIFY MAP" button adds a stamp defined by the argument boxes. "Value to add" can be positive or negative. Using "UNDO" button after any modification reverts the map to the previous state.
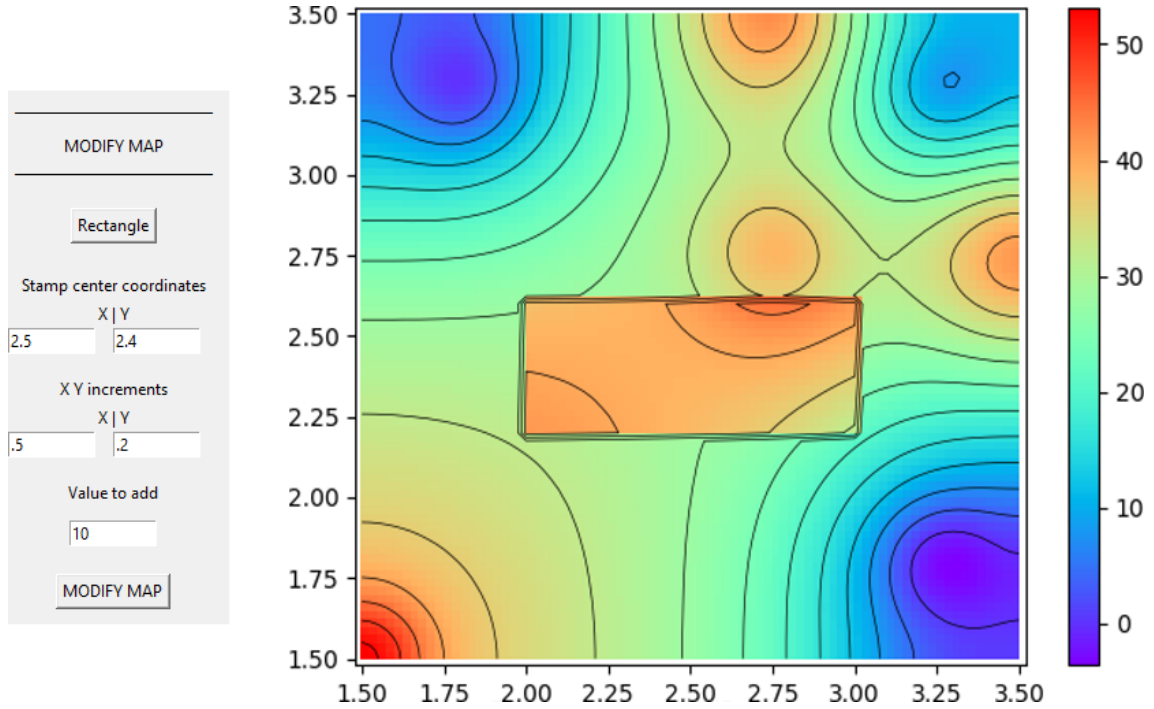

Figure 2: Example of "Rectangle" modification with the indicated values.
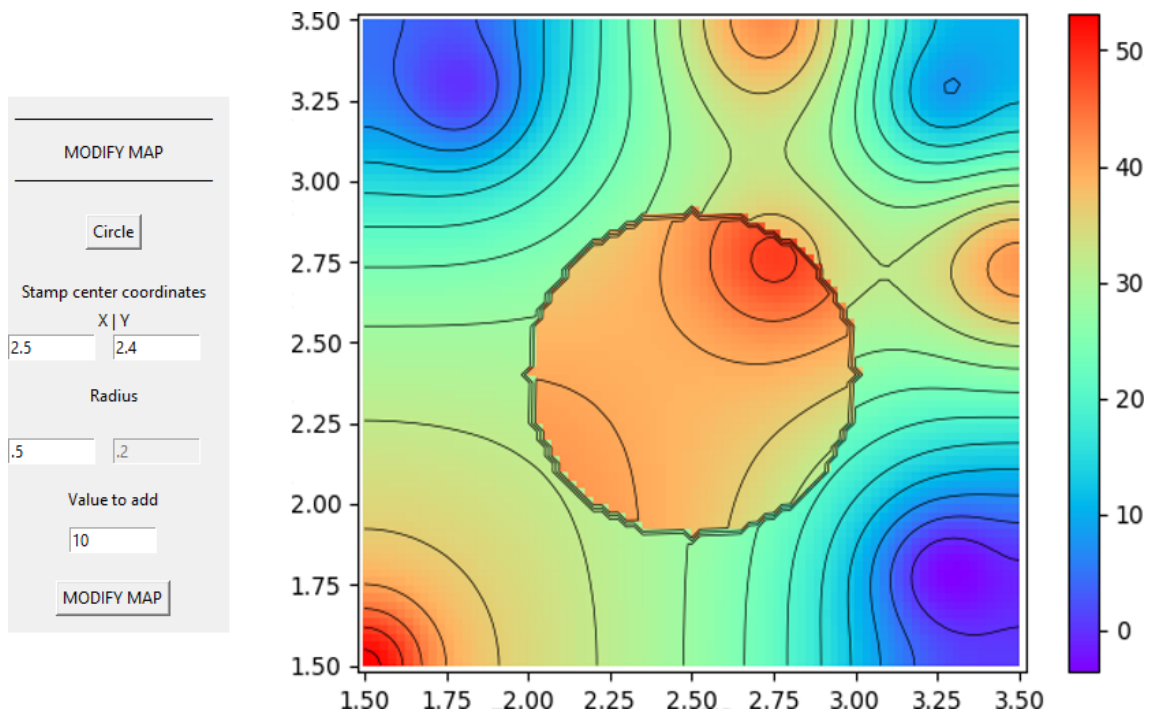

Figure 3: Example of "Circle" modification with the indicated values.

7.2 Invert map

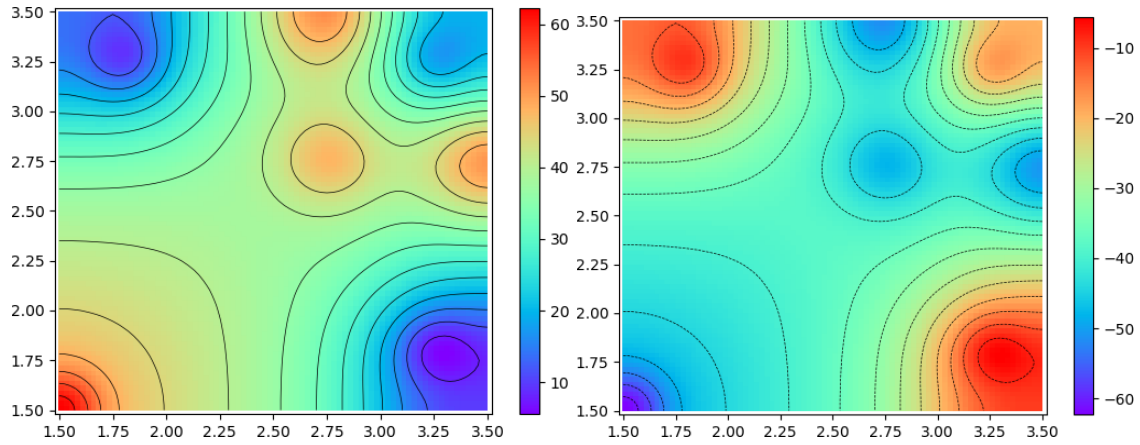Invert map multiplies the energy values of the map by -1 (Fig. 4).

Figure 4: Example of a map inversion with "INVERT MAP" button.

7.3 Smooth map

The smoothing technique present in MEPSA is rather simple non distance-weighted running average. It is just present for fast on-the-run tests removing undesired local minima. It should not be used for final works. Using the "UNDO" button after the smoothing reverts the map to the previous state.

There are two smoothing modes:

- "2 axis smooth": for each iteration, the new value of each point [x,y] is defined as the average value of the points [x,y],[x-1,y],[x+1,y],[x,y-1] and [x,y+1].

- "4 axis smooth": for each iteration, the new value of each point [x,y] is defined as the average value of the points [x,y],[x-1,y],[x+1,y],[x,y-1],[x,y+1],[x-1,y-1],[x+1,y+1],[x+1,y-1] and [x-1,y+1].
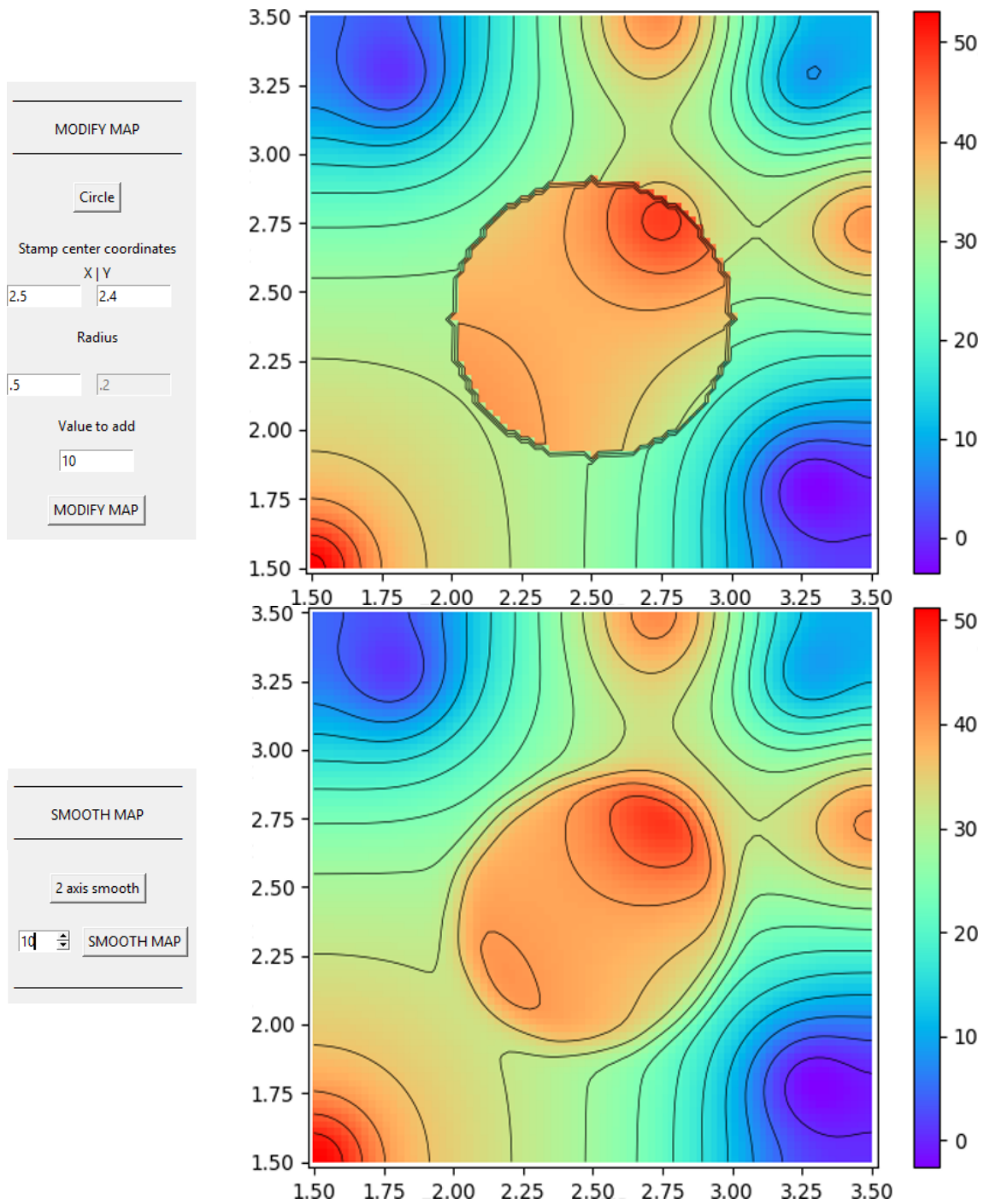
Figure 5: Example of the use of "SMOOTH MAP" on the same "Circle" modification shown in figure 3 with 2 axis smoothing mode and 10 iterations.

7.4   Save map

The current version of the map can be saved into a three column formatted plain text file using the "SAVE" button.

Connectivity analyses

"Connectivity analyses" button opens the "connectivity analyses" window, from which three main tasks can be done

- Global connectivity analysis

- Path generation

- Well sampling analysis

Before any of these tools becomes available, MEPSA needs to define nodes on the map with the "Find nodes" function. The rest of the buttons become accessible as their individual requirements are fulfilled. All the data generated with the path generator can be saved into column formatted plain text files and all the plots can be saved as images in several image formats.

8.1   Find nodes

"FIND NODES" button makes MEPSA analyze the map surface and find the available nodes (the map minima) (Fig. 6). For each point of the map MEPSA analyses the four accessible points (three in the borders and two in the corners) and check if their value is higher, equal or lower than the point that is being evaluated. There are two working modes available: "MIN ONLY" and "MIN AND FLAT". The difference between them is the kind of accessible points required to consider a candidate point a node:

- MIN ONLY: a least one accessible point with higher energy and no accessible points with lower energy.
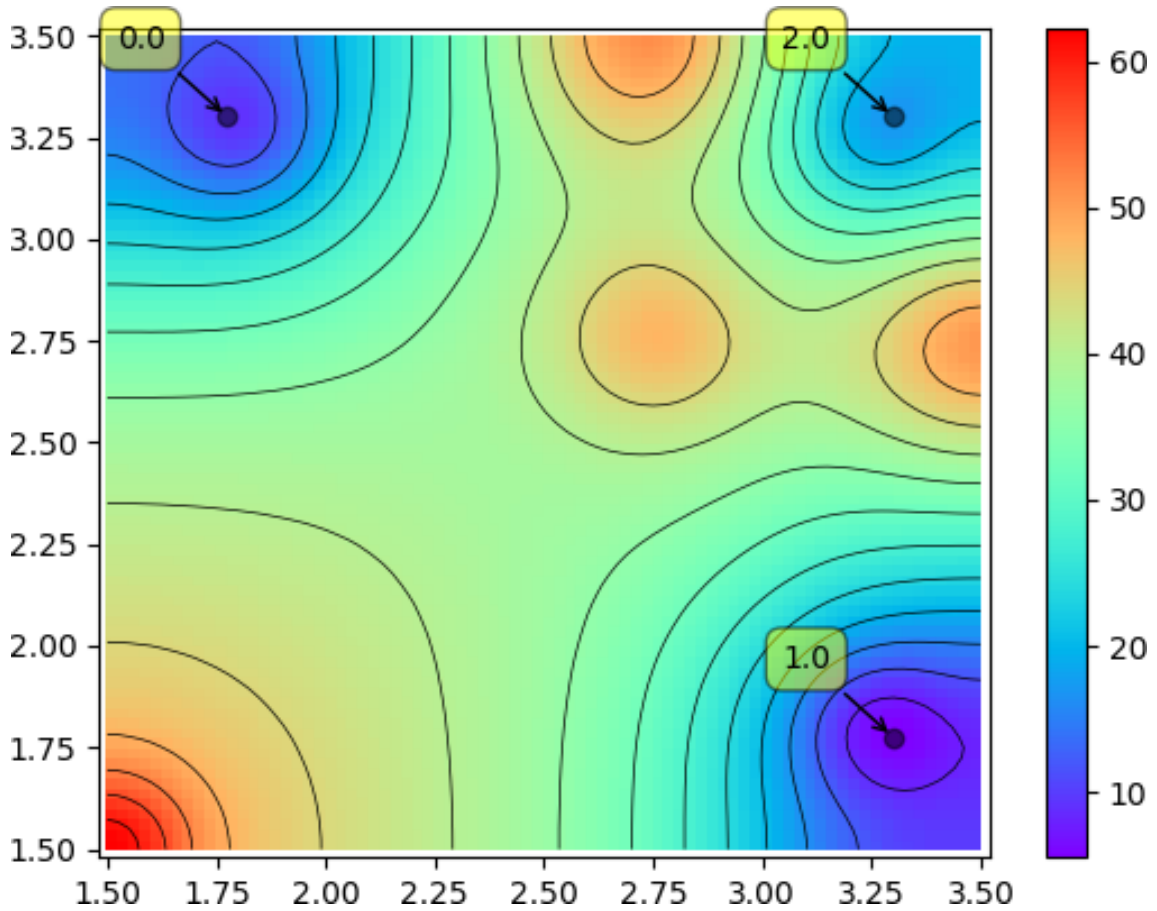
- MIN AND FLAT: no accessible points with lower energy.

Figure 6: Example of a "NODES PLOT". Numbers in round squares are the identifiers that MEPSA has given to the nodes. These identifiers can be used to select origin and target points in the path generation.

8.2 Analyze connectivity

"ANALYSE CONNECT." button makes MEPSA the region of the map dominated by each node and the barriers connecting these regions. Two sampling modes are available, "FULL" (Fig. 7) and "MINIMAL" (Fig. 8). "FULL" samples the whole map, revealing all minima and barriers at once while "MINIMAL" only samples the map to the point in which every node is connected to, at least, another one. The yellow annotations refer to minima points and the purple ones to saddle points. The numbers show the energy values of those points. The color overlay in the map divides the map in the different regions dominated by each minimum.
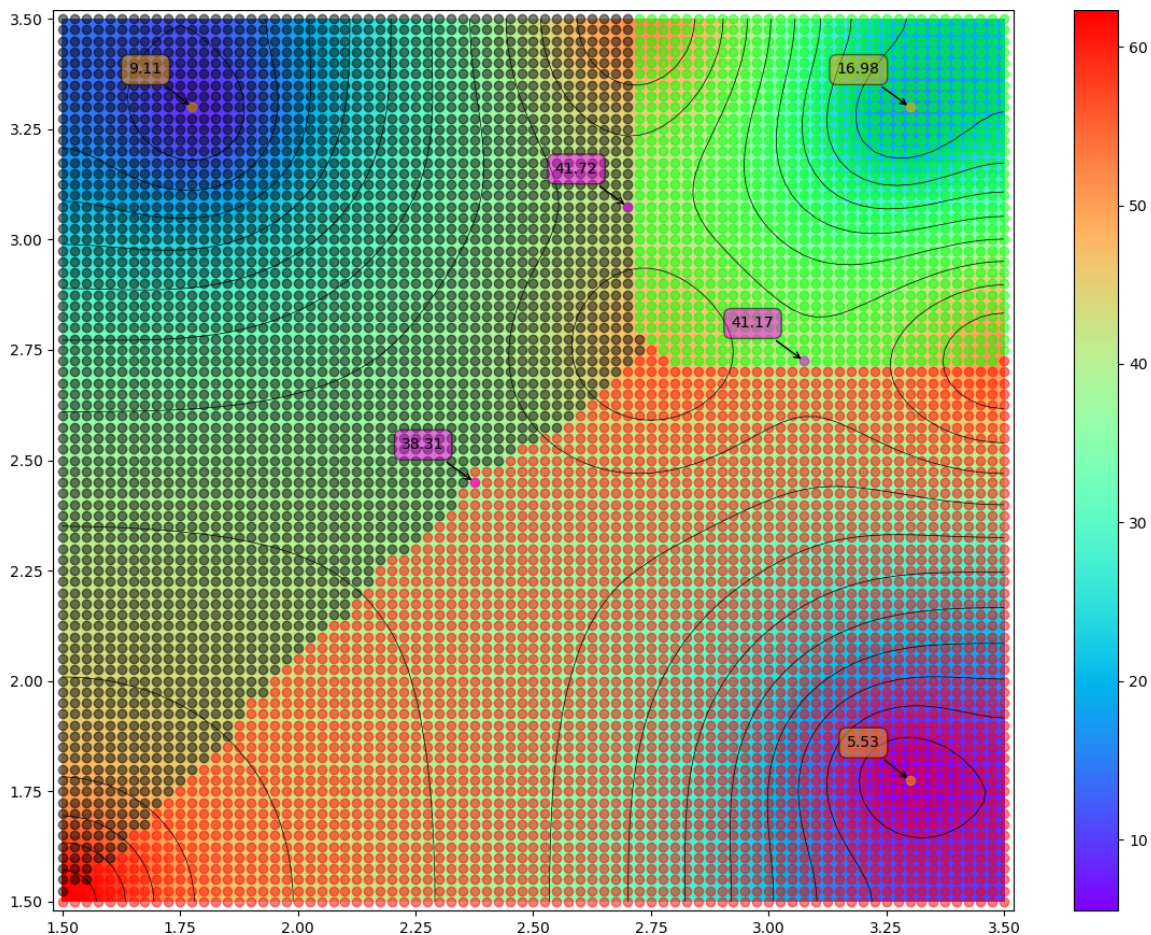

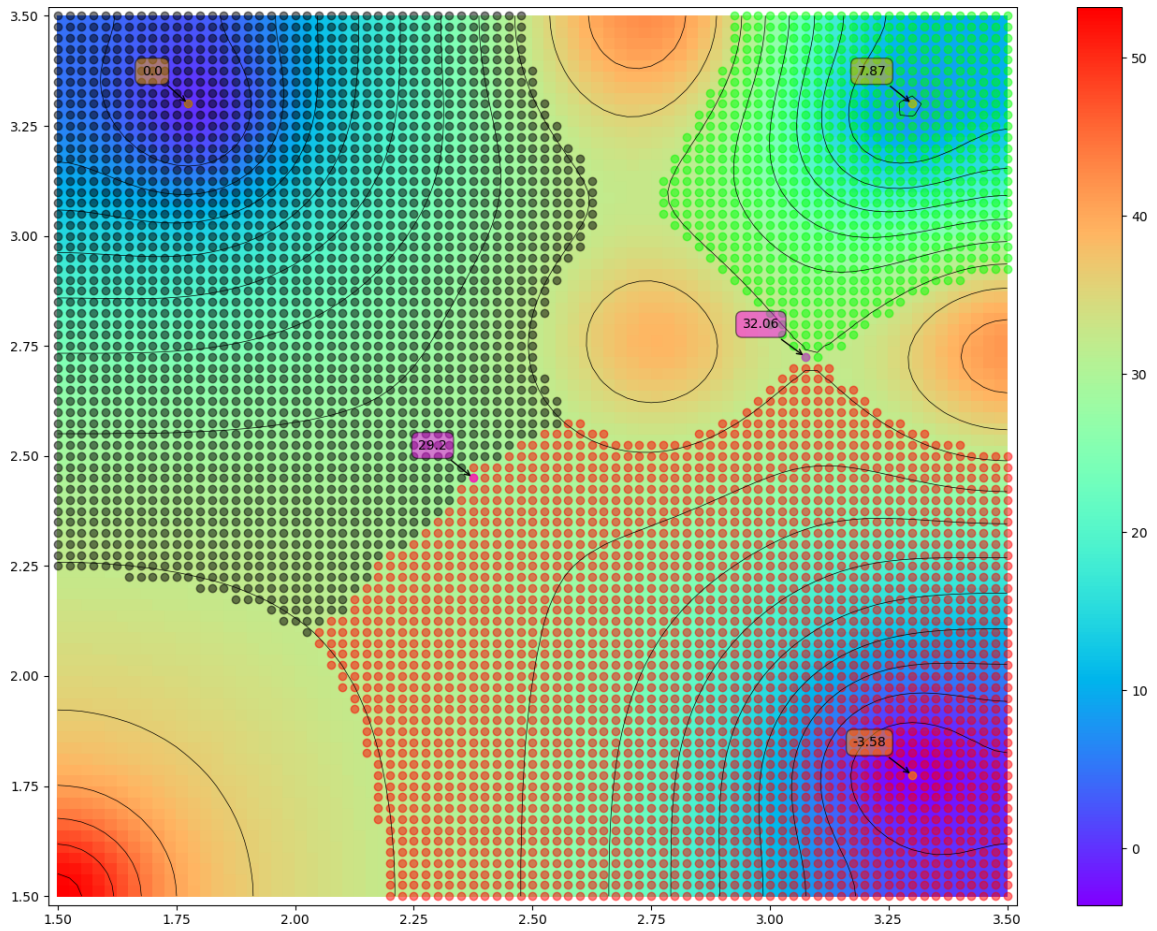Figure 7: Example of a connectivity plot in "FULL" mode.

Figure 8: Example of a connectivity plot in "MINIMAL" mode.

8.3   Path generator

In order to calculate the lowest energy path connecting two points MEPSA needs to know the origin point, the target point and the path generation method to be used.

8.3.1   Select origin and target points

MEPSA offers three ways to select the origin and target points:

- Directly selecting the point by the node identifier.

- Searching for the lowest energy node in a rectangular area of the map defined with the arguments in the parameters boxes with the "USE SEARCH" option.

- Selecting arbitrary points (that do not necessarily have to be nodes) with the "ARBITRARY" option. In this case only the "search center coordinates" are used. With arbitrary selection MEPSA selects the point of the map closer to the specified coordinates.

If "ARBITRARY" selection method is used, "node by node" sampling method is disabled.

After configuring the node selection the "SET O&T" button passes that selection to MEPSA and defines the origin and target points. "O&T PLOT" shows the selected origin and target points in order to let user check if the selected points are the expected (Fig. 9).
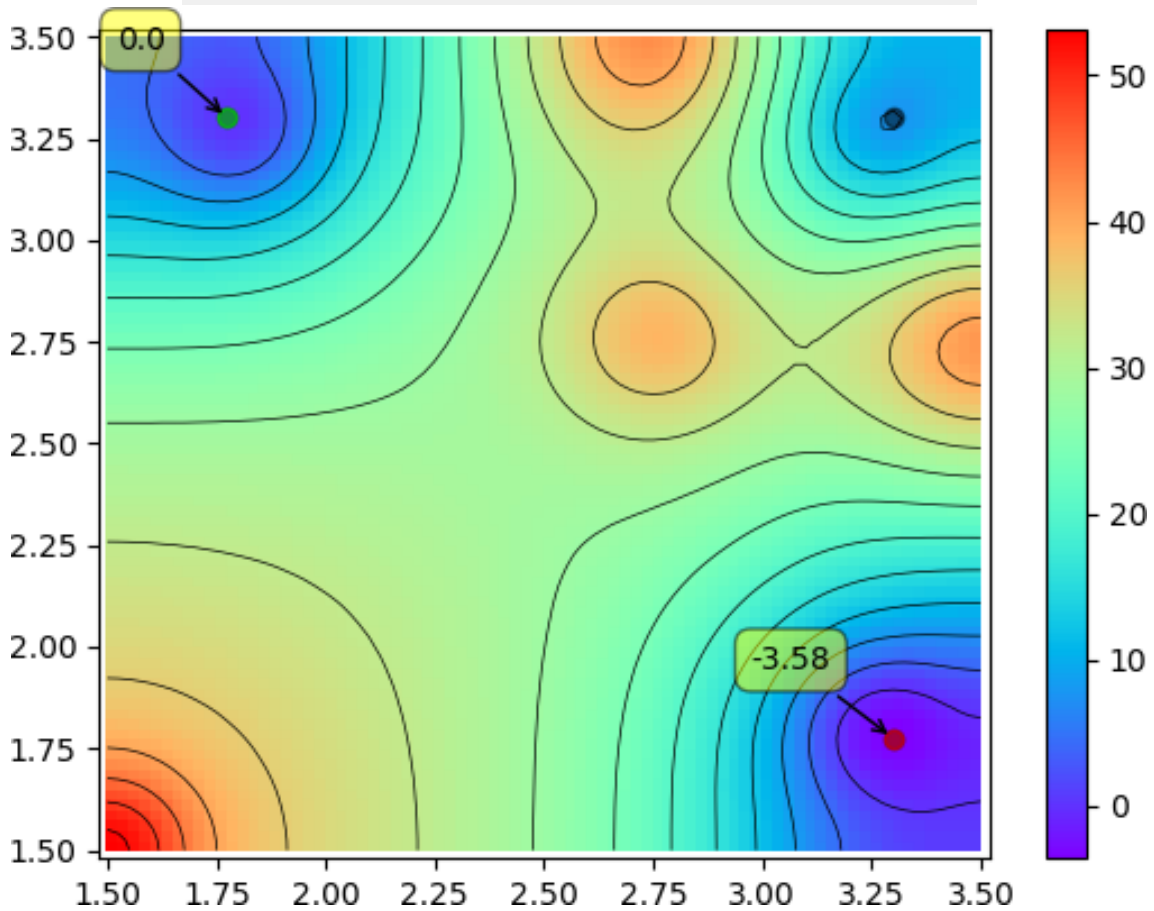
Figure 9: Example of origin (green) and target (red) points selection (O&T PLOT). Labels show the energy value of both points. An example how to select these nodes by node id (0) and by searching the minimum in a specified region (USE SEARCH) is shown in the top connectivity analyses window.

8.3.2 Path generation methods

In order to find the lowest energy path between two points (origin and target), MEPSA propagates from the origin point always choosing the lowest energy points available until the target point is reached. The resulting path is the same for two particular points, no matter which one is the origin and which one is the target.

There are two sampling modes available (Fig. 10):

- GLOBAL: it samples the map starting from the origin point, iteratively occupying the lowest energy points available, until the target point is reached. MEPSA keeps track of the iteration in which each particular point has been occupied and uses this information to perform a trace back from the target point to the origin through the points with the lowest iteration number. This mode is faster but does not guarantee that all the intermediate nodes are visited by the predicted path (Fig. 11).

- NODE BY NODE: it performs a "GLOBAL" sampling first, using data generated to determine which node domains are visited during the minimum energy path calculation and in which order. Then it perform a series of short "GLOBAL" runs connecting these nodes, giving as output the combination of all these paths. This strategy forces the shortest lowest energy path to visit every node whose domain is visited. If origin or target were selected with the "ARBITRARY" selection method, "NODE BY NODE" sampling mode cannot be used. This mode is slower but guarantees that all the intermediate nodes are visited by the predicted path (Fig. 12).

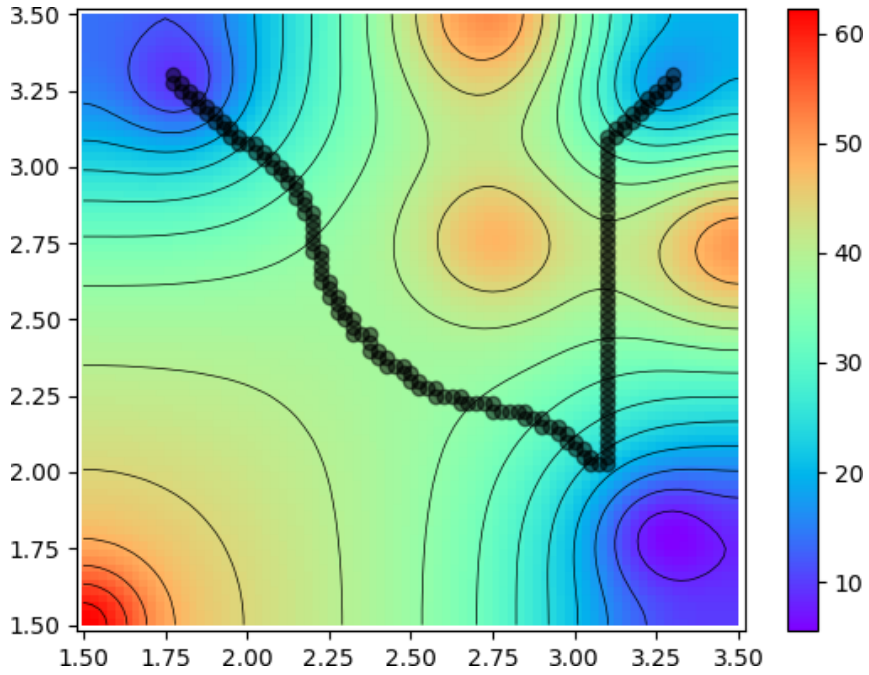Figure 10: Path generation selection tab.

Figure 11: Example of a path calculated in "GLOBAL" mode not properly visiting an intermediate node.
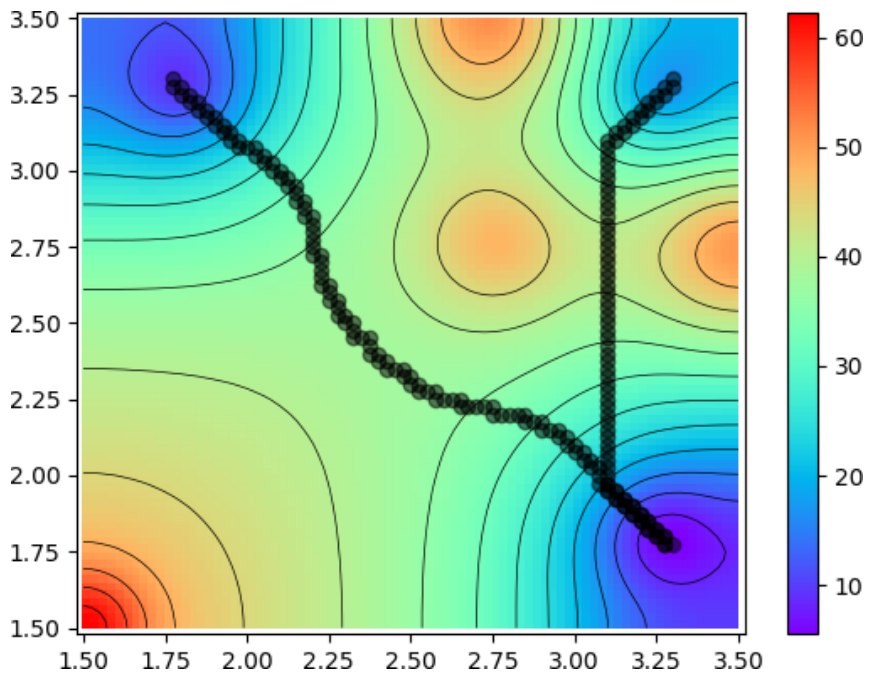

Figure 12: Example of a path calculated in "NODE BY NODE" mode correctly visiting the intermediate node.

8.3.3   Plot types

After a path has been successfully generated, MEPSA offers a range of built-in plot options.

Path plot

"PATH PLOT" button plots the generated path with black dots over the map (Fig. 13).

Figure 13: Path plot.
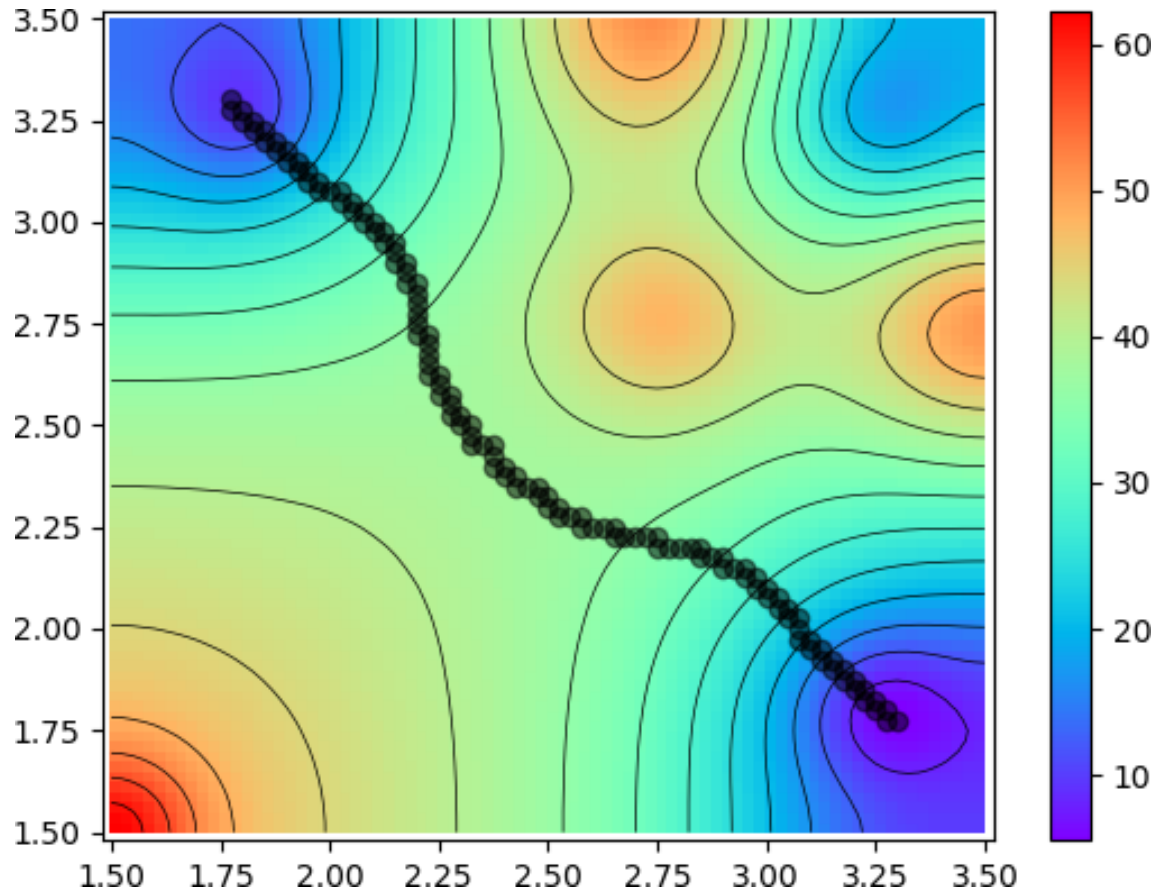
Full plot

"FULL PLOT" button generates a similar plot to the "PATH PLOT" one but containing more information (Fig. 14).
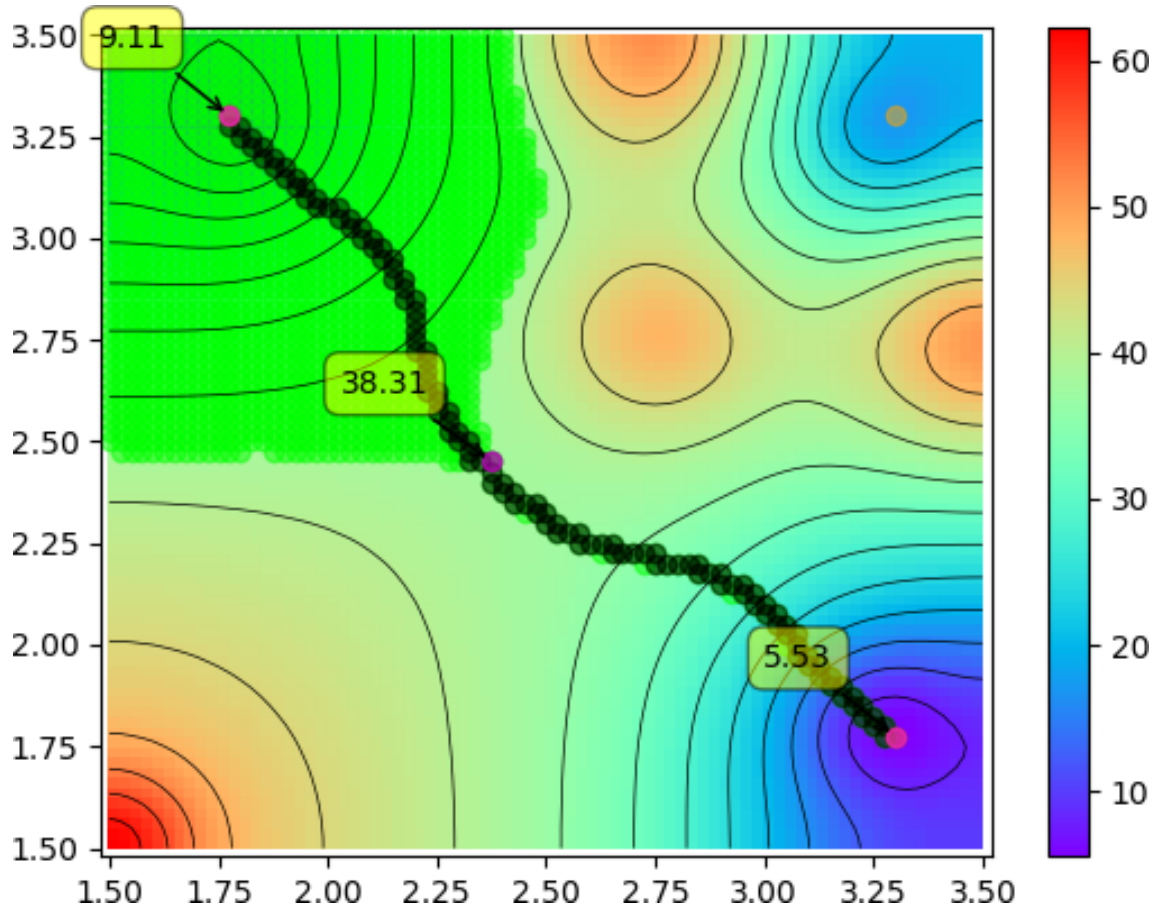

Figure 14: Full plot. The black dots (may look like dark green due to the mix of light green and black) mark the generated path, green dots the area visited during the propagation run and yellow labeled purple dots mark the points of interest along the path.

Energy plot

"ENERGY PLOT" button plots the energy profile along the path (Fig. 15).
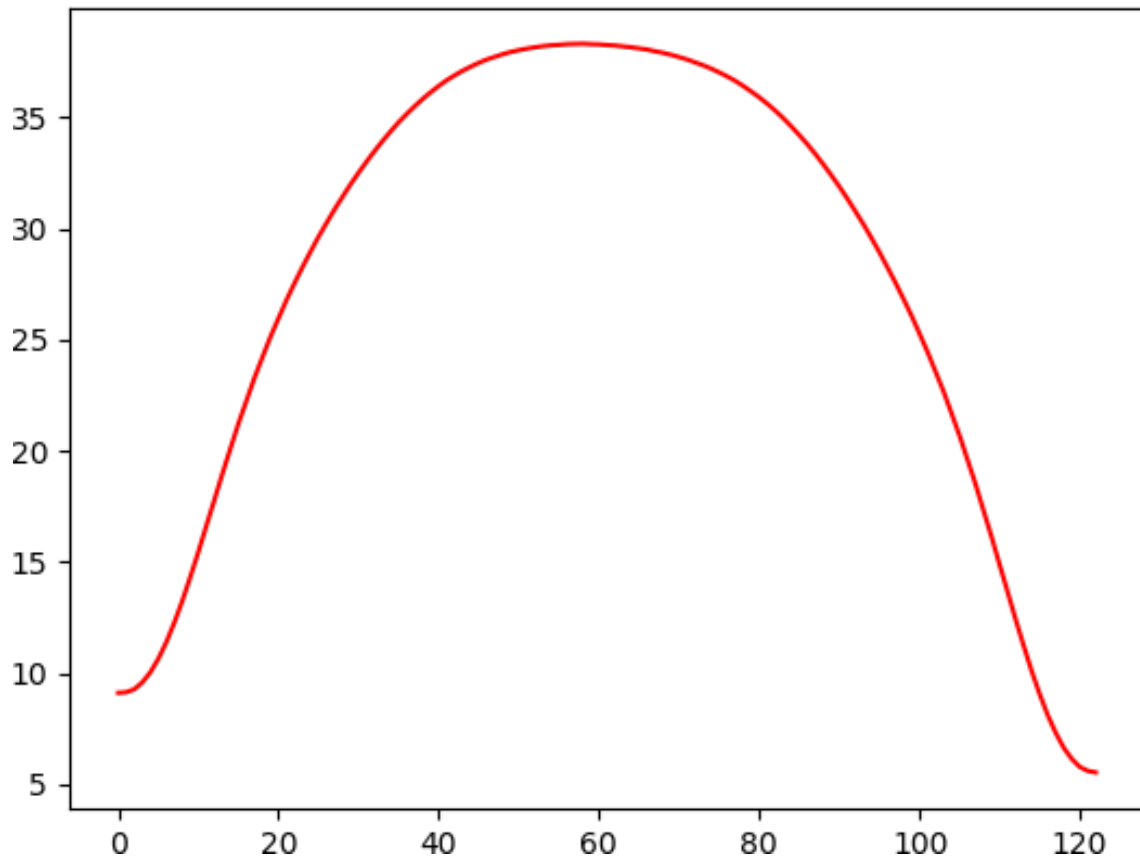
Figure 15: Energy plot.

8.3.4 Path smooth

MEPSA offers a simple tool to smooth the generated path. It is just a cosmetic adjustment in case the user wants a smoother path figure (Fig. 16). Origin and target are fixed and the rest of the points get their positions averaged with the next a previous point on each iteration. Unlike map smoothing, path smoothing is always performed over the original path. For example, if a 5 iterations smoothing is done after performing a 10 iterations one, the obtained path is the result of 5 smoothing iterations, not 15.

Figure 16: Example of path smoothing.

8.3.5   Stacking multiple paths

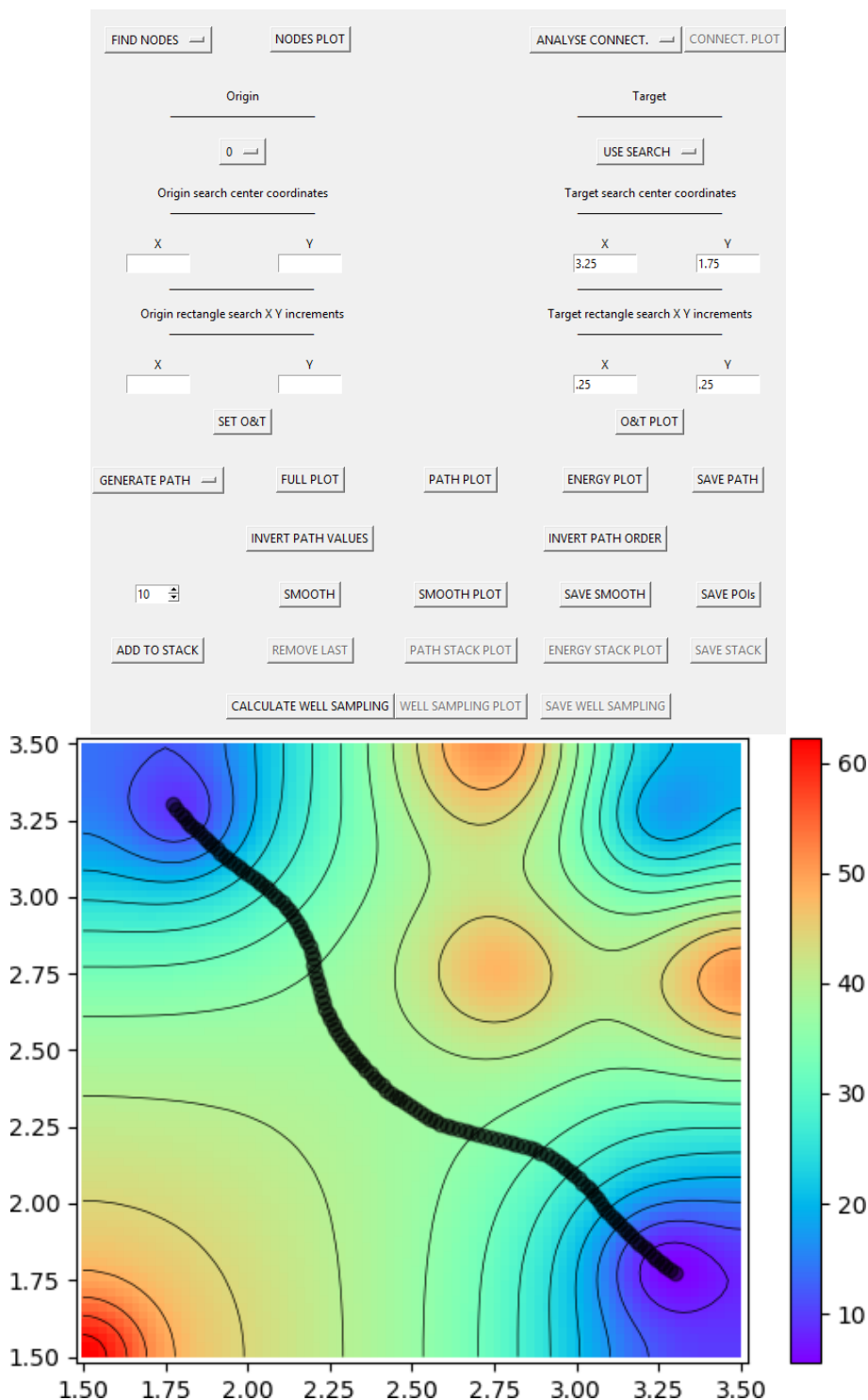Whenever a new path is calculated or a map is modified MEPSA deletes all the information about the last path calculated.

Sometimes it can be interesting to store the previously calculated paths (see "Chapter 9.2: Use map editor to check alternate paths").

This is why MEPSA path generator has a stack in which to store the generated paths.

"ADD TO STACK" button stores the current path in the stack and the "REMOVE LAST" button removes the last path stored in the stack.

"PATH STACK PLOT" creates a similar plot to the "PATH PLOT" one but plotting all the paths stored in the stack at the same time (Fig. 17).

"ENERGY STACK PLOT" creates a similar plot to the "ENERGY PLOT" one but plotting all the energies of the paths stored in the stack at the same time (Fig. 18).
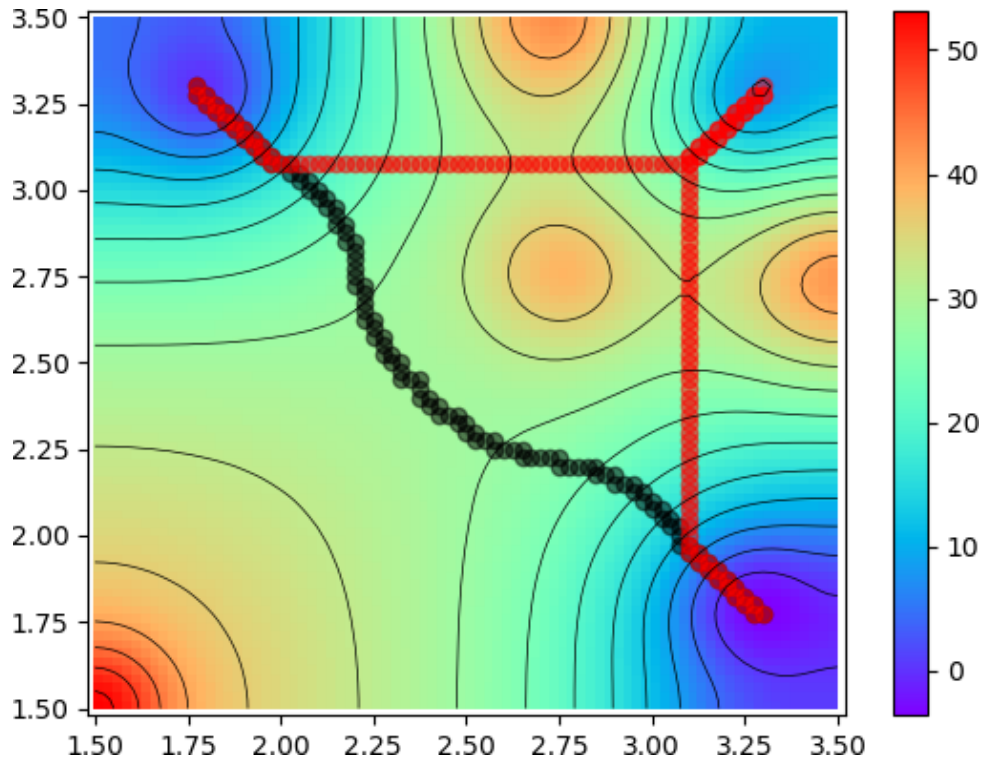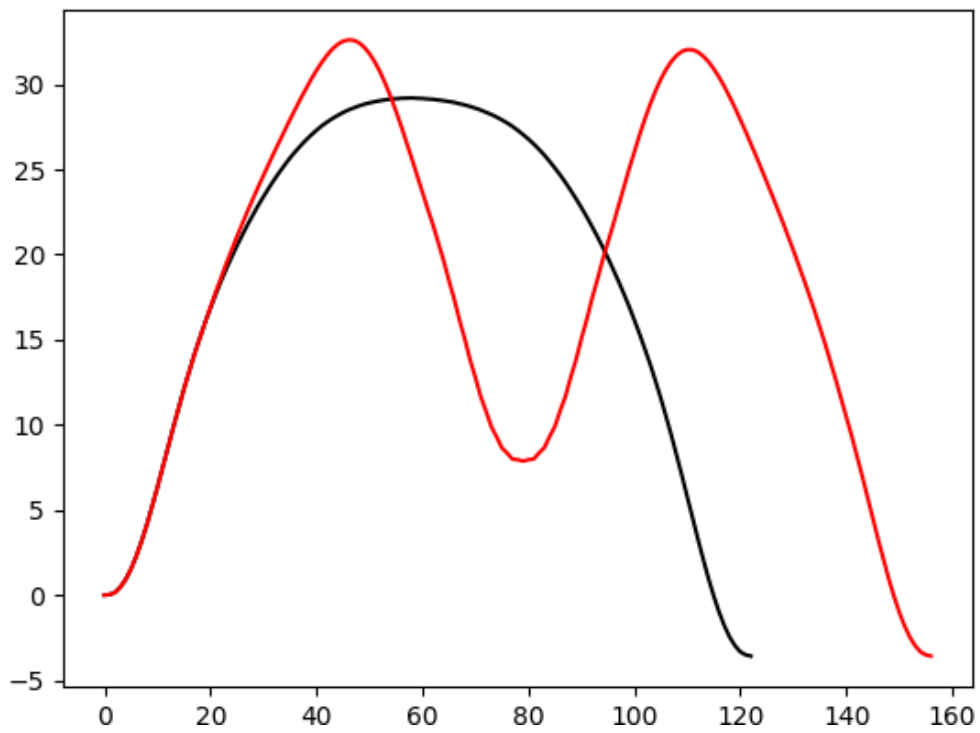
Figure 17: Example of path stack plot.

Figure 18: Example of energy stack plot.

8.3.6   Invert path values and order

The order of the calculated path points can be inverted with the "INVERT PATH ORDER" button.

The energy values of the calculated path points can be multiplied by -1 with the "INVERT PATH VALUES" button.

8.3.7   Saving data

The data generated by the path generator can be saved in different ways.

All the plots can be directly saved into image files (with several formats available) using the floppy disk icon at the plot windows.

There are also several ways to store data in column formatted plain text files:

- "SAVE PATH" button saves the X, Y and Z values of the path points. Therefore, saved data includes the path trajectory and the energy profile.

- "SAVE POIs" button saves the X, Y and Z values of the interest points (minima and saddle points).

- "SAVE SMOOTH" button saves the X and Y coordinates of the smooth path.

- "SAVE STACK" button saves the X, Y and Z values of the path points stored in the stack. In the .txt file the different files are separated by PATH_ followed by the path number (PATH_0, PATH_1, etc.).

8.4 Well sampling analysis

After succesfully selecting origin and target, "well sampling analysis" functionality becomes available. It calculates the amount of points that has to be visited to get to the target node's domain from the origin node and *vice versa*.
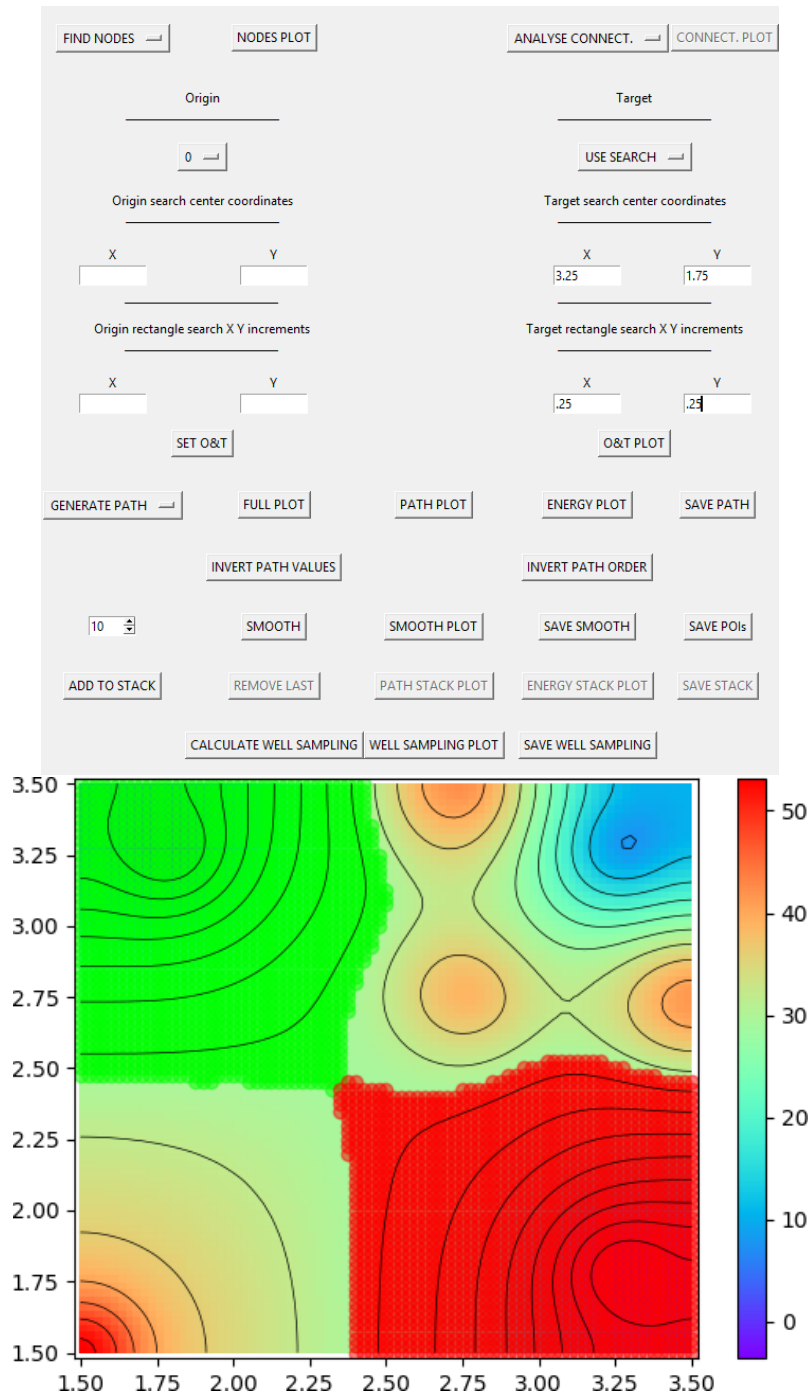

Figure 19: Example of well sampling analysis.

Examples

9.1 Maxima edge profile instead of minimum energy path

Sometimes it might be useful to obtain the profile of the highest energy points connecting two maxima.

To do this the user has to:

- Load the map (Fig. 20A).

- Use the invert function in the map editor (Fig. 20B).

- Run the path generator selecting the desired nodes (Fig. 20C).

- Use the invert path values function to get the correct energy values (Fig. 20E).

- Add the path to the stack.

- Use the undo function in the map editor.

- Use the path stack plot and the energy stack plot buttons to represent the maxima edge profile (Fig. 20D and 20E).

As expected, the path obtained with this approach goes along the boundaries of the regions defined by the global connectivity analysis performed over the original map.
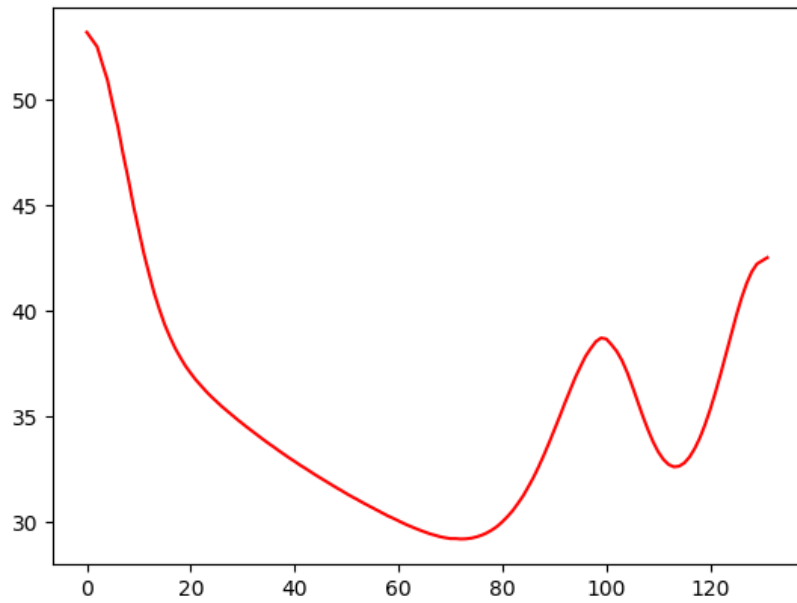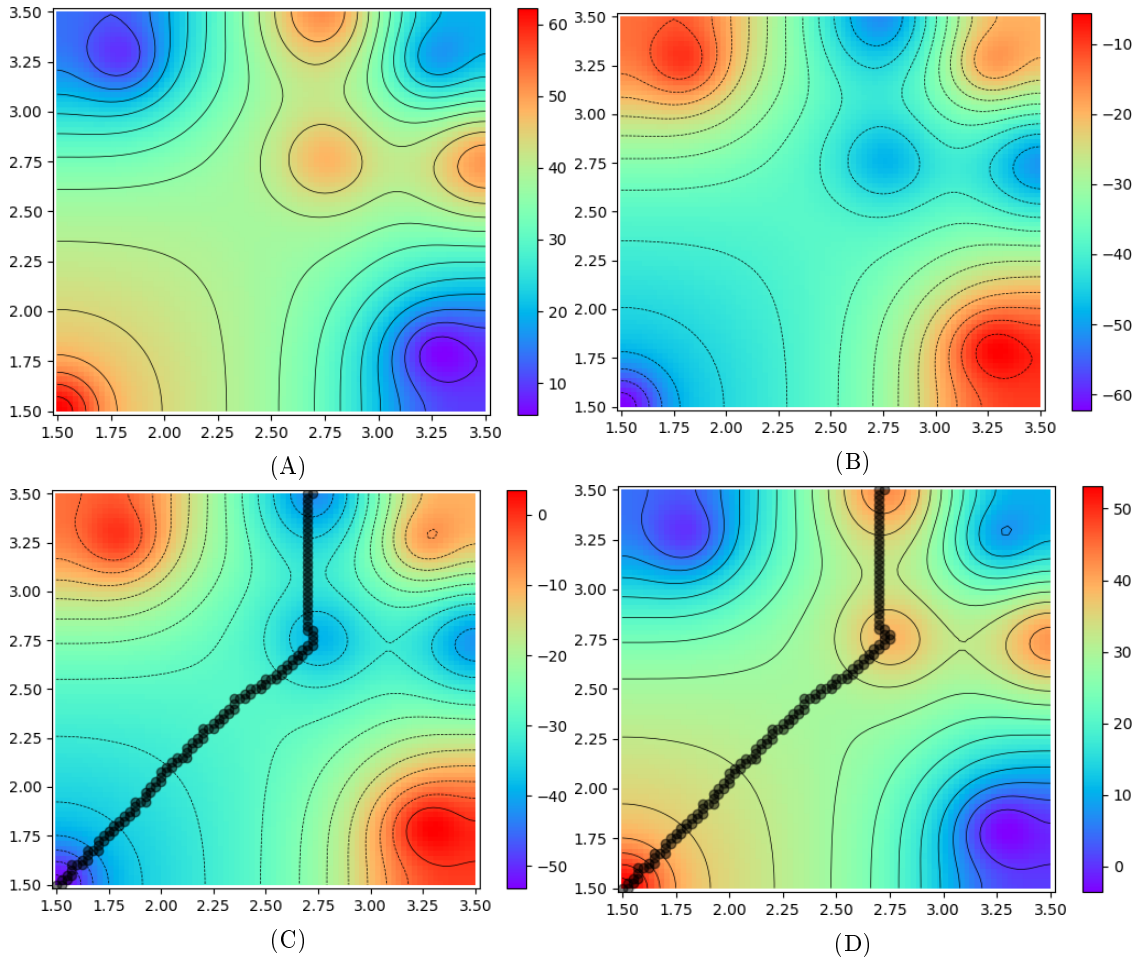
Figure 20: Steps of the maxima edge profile generation. A) Loaded map. B) Inverted map. C) Path calculated over the inverted map. D) Calculated path plotted after undoing the map inversion. E) Energy plot of the path after performing a path energy values inversion.

9.2   Use map editor to check alternate paths

Using the map editor, favorable paths can be blocked forcing MEPSA to sample different ones. As paths can be stored in the path stack, several alternate paths can be compared at once (Fig. 21).

To do this the user may follow these steps:

- Load the map and calculate the minimum energy path between the points of interest (Fig. 21A), storing the resulting path in the path stack.

- Use the the map editor to add energy to a desired region in order to force the path generator to sample other trajectories (Fig. 21B).

- Calculate the minimum energy path between the points of interest (Fig. 21C 21D), storing the resulting path in the path stack.

- In order to represent both paths over the original map use the undo button in the map editor until you get to the unmodified map and then use the path stack plot (Fig. 21E).

- Both paths can be simultaneously represented over the map by using the path stack plot (Fig. 21E).

- To compare the energy along both paths use the energy stack plot (Fig. 21F).

Figure 21: Steps of the alternate path generation. A) Regular path that is stored in the path stack. B) Map edition blocking the most favorable path. C) Path calculated over the modified map in "GLOBAL MODE". If you want the top right minium to be visited by the path, use "NODE BY NODE" mode instead (D). Once the desired path is obtained, it has to be stored in the path stack too. E) Path stack plot over the original map. F) Energy stack plot.

9.3 Converting a path into molecular dynamics restraints

"SAVE PATH" button generates a three column formatted plain text file and, therefore, its data can be easily extracted and reformatted with scripts.

A simple Perl script to convert these output files into Amber restraint definitions can be downloaded from:

http://bioweb.cbm.uam.es/software/MEPSA/

This script creates files with the restraint definitions needed to run a two parameters biased molecular dynamics simulation along the estimated path.

In order to convert a path .txt file into restraints files compatible with other molecular packages, the user may use this script as scaffold and adapt the output to the desired format.

9.4 Use of MEPSA over highly complex landscapes

Some of the MEPSA functionalities are designed with smooth, topologically simple surfaces in mind. For example "ANALYSE CONNECT." and "FULL PLOT" may not perform well with highly complex landscapes and their output will probably contain an excessive amount of information to extract any useful conclusion from it. Using smoothing techniques over those maps may alleviate these issue. In any case, the path generator and all the tools of the map editor should still perform correctly.

As an example of this, the following images show the detection of the lowest elevation path (as this map contains elevation instead of energy values) between Geneva and Turin with MEPSA (Fig. 22 and 23) as well as the corresponding well sampling (Fig. 24). The topographic data was extracted from the SRTM (Shuttle Radar Topography Mission) 30 ARC-SECOND ELEVATION v2.1 data set through the ORNL DAAC OGC Spatial Data Access tool: http://webmap.ornl.gov/wcsdown/dataset.jsp?ds_id=10008

The origin and target points were defined in "ARBITRARY" mode, being Geneva (x:6.1422961 y:46.1983922) the origin and Turin (x:7.6868565 y:45.0703120) the target.
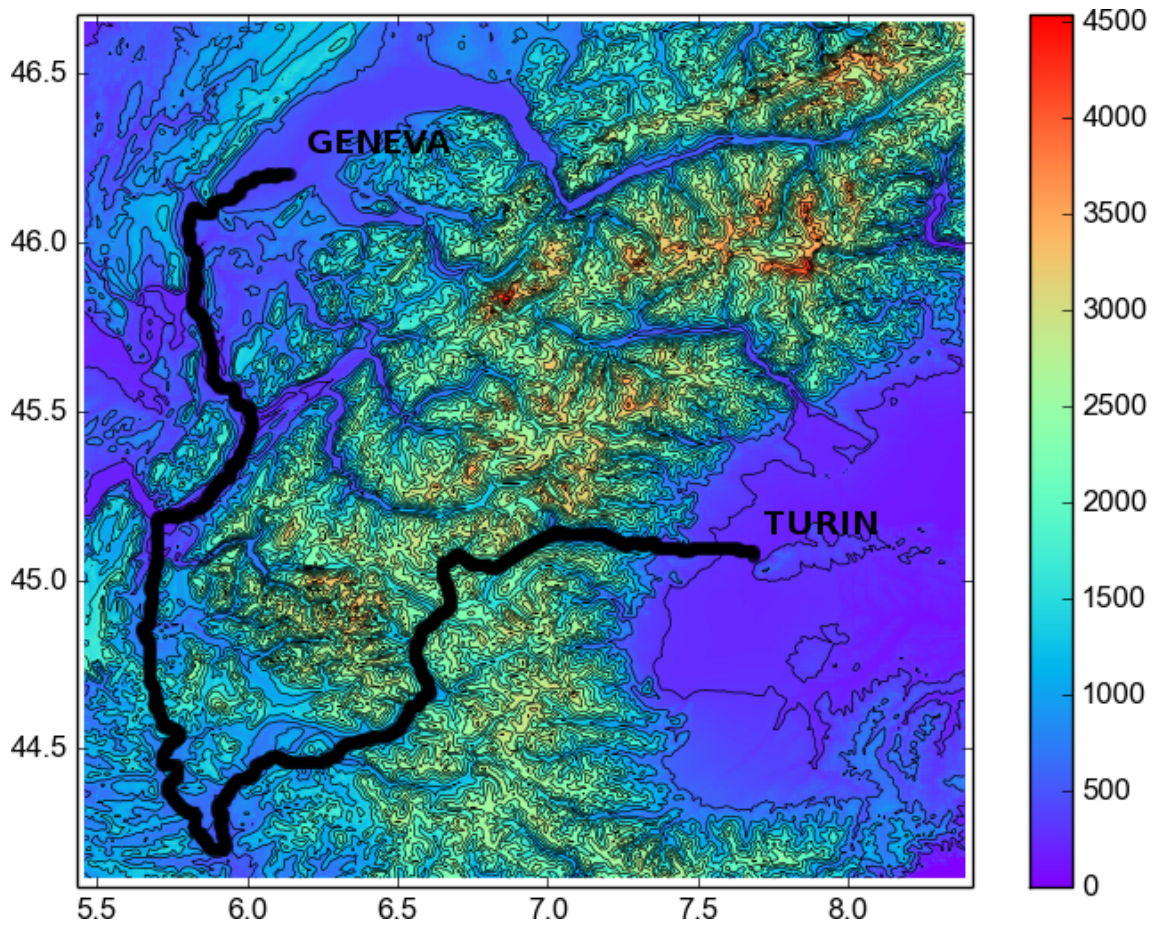
Figure 22: Minimum elevation path between Geneva and Turin. The names of the cities were added with an image editor.
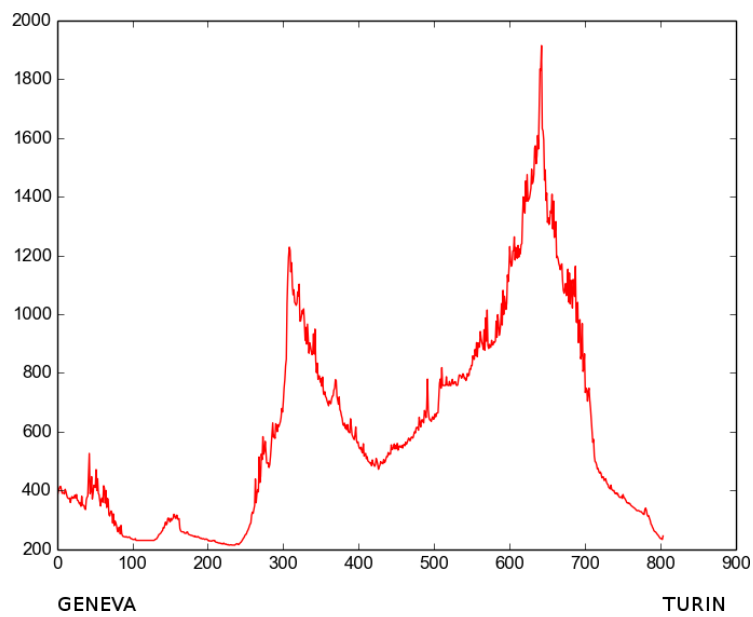

Figure 23: Elevation along the minimum elevation path between Geneva and Turin. The names of the cities were added with an image editor.
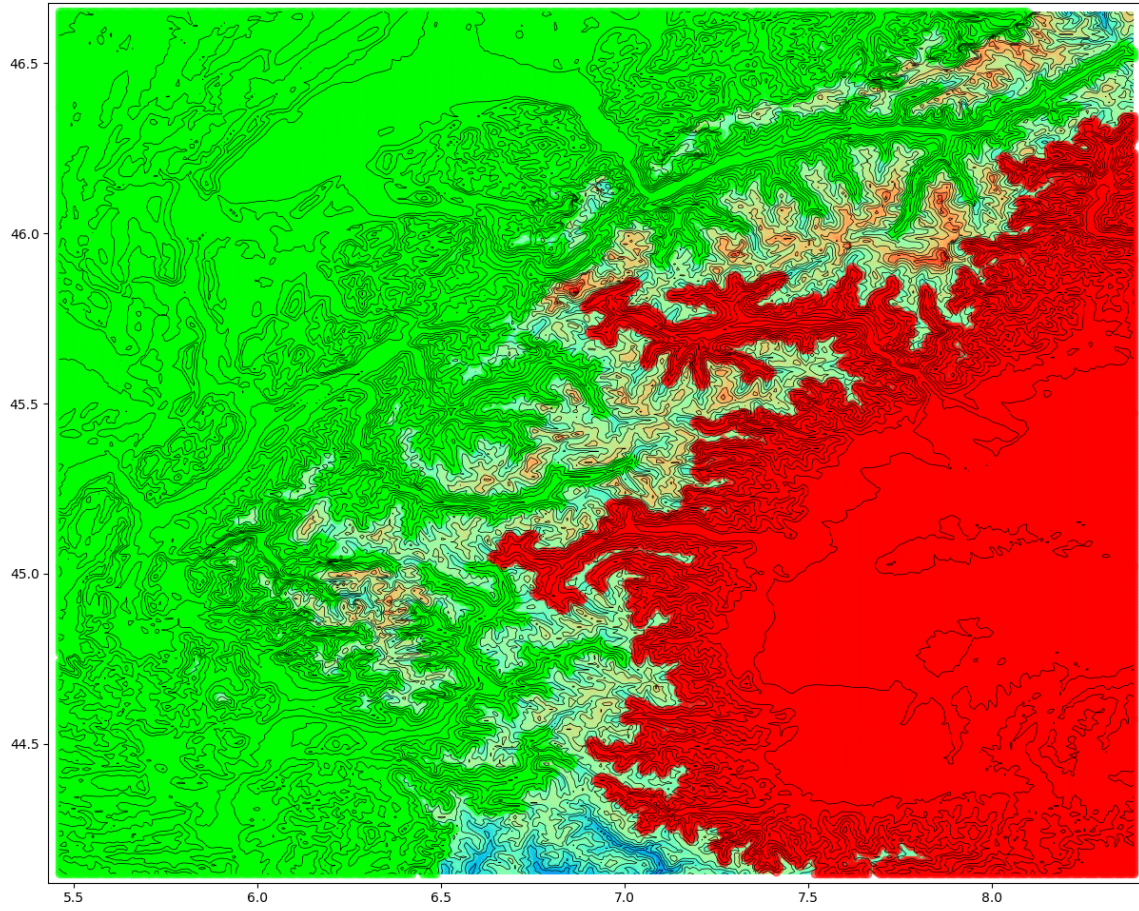
36

Figure 24: Well sampling plot indicating the points to be sampled from Geneva and Turin to detect the minimum elevation path between them.