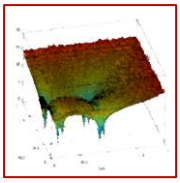

MCLand

User Manual

MCLand: A Python software package to draw Waddington's epigenetic landscape

Ket Hing Chong[1], Xiaomeng Zhang[1], Jie Zheng[2,*]

[1] Biomedical Informatics Lab, School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

[2] School of Information Science and Technology, ShanghaiTech University, Pudong District, Shanghai 201210, China

[*]zhengjie@shanghaitech.edu.cn

URL: https://mcland-ntu.github.io/MCLand/index.html

Last updated: 18 February 2020

1. Introduction

The purpose of this user manual is to guide user on how to use MCLand for plotting and visualizing Waddington's epigenetic landscape for gene regulatory network (GRN). MCLand is a Python software implementing the Monte Carlo method for estimating the probability distribution P(x) and quasi-potential U = – ln P(x) as proposed by Zhang et al. [1]. MCLand enable users to plot and visualize Waddington's epigenetic landscape with ease by using the user-friendly graphical user interface (GUI).

2. Download and Running MCLand

For downloading MCLand please go to the URL: https://mcland-ntu.github.io/MCLand/index.html

2.1 Running MCLand

The Python source code for MCLand Version 1 can be run on Windows, Linux or Mac. Unzip the downloaded MCLand file. For Windows user, the steps for running MCLand are given here. However, for Linux or Mac users the steps should be similar. To run the graphical user interface for MCLand please install Anaconda 3 version 4.3.1 from https://repo.continuum.io/archive/ and then use Anaconda Prompt to install python-libsbml with conda run: conda install -c sbmlteam python-libsbml

Then you have to set up the Python 3.6 interpreter, e.g. add the system variable to path C:\ProgramData\Anaconda3, C:\ProgramData\Anaconda3\Scripts, C:\ProgramData\Anaconda3\Library\bin. After completed the setting, you are ready to use Anaconda 3 Spyder (a Python IDE) to open the MCLand_ver1.py file and run the program to start the Graphical User Interface.

Then the MCLand GUI will be opened as in Figure 1.

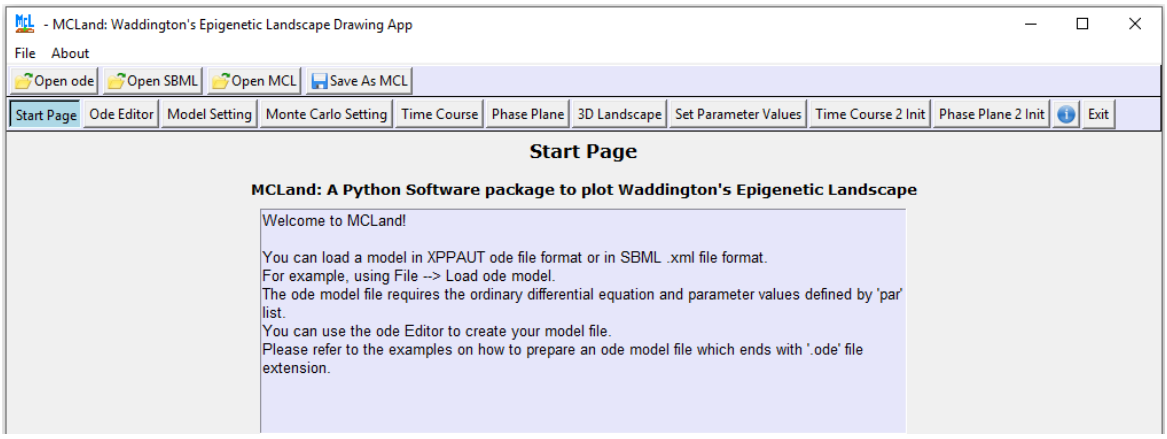

Figure 1. The Start page of MCLand GUI.

3. XPPAUT Model Creation using ode Editor

To start your model creation you need to have a model GRN and its ordinary differential equations. Based on the model equations you can then create the XPPAUT model by using the ode Editor. The ode Editor is used to create and edit the ode model file which end with .ode file extension. The XPPAUT ode file is a standard file used by the software XPPAUT (or XPP). In fact, ode file can be created by any text editor such as notepad. However, we included the ode Editor for making MCLand easy to use.

Some of the standard notations used in XPPAUT are given below:

1. The '#' is used for comment line.
   e.g. # Model equations and parameter values are listed in this ode file
2. The differential equation is defined by 'dx/dt = …' or 'x´ = …'
   e.g. dx/dt = k−c*x
   Or x´ = k−c*x
3. The parameter values are listed starting with param or par:
   e.g. par k=2, c=1
4. The initial condition values are listed with the keyword 'init'. The initial condition is optional. If the initial conditions not defined then by default it will be assumed to be zero for all the variables.

e.g init x=0.2

5. The XPPAUT file end with the keyword ‘done’.

Below is an example of the ode file (save in the file name ‘Wang_PNAS_2011.ode’) for Case study 1:

```
# The developmental process proceeds as moving from
# undifferentiated to the differentiated basin of attractions.
# Wang_PNAS_2011.ode
#
x1'=a1*x1^n/(S^n+x1^n)+b1*S^n/(S^n+x2^n)-k1*x1
x2'=a2*x2^n/(S^n+x2^n)+b2*S^n/(S^n+x1^n)-k2*x2
param b1=1, b2=1, k1=1, k2=1, S=0.5, n=4, a1=1, a2=1
@ total=500, xp=t, yp=x1, dt=0.01, xlo=0, xhi=500, ylo=0, yhi=50,
maxstor=500000
done
```

In MCLand we have created a keyword for defining constants and functions by using the keyword ‘conts’. For example, conts pi=3.14159

The information given above is sufficient for creating the ode model file to be loaded in MCLand. For more details about XPPAUT, users are refer to the book [2] or the XPPAUT website http://www.math.pitt.edu/~bard/xpp/xpp.html.

The GUI interface for the ode Editor is given in Figure 2. User can type the model information in the text area or select the ‘New’, ‘Open’, ‘Save’ or ‘Save As’ as in any text editor.

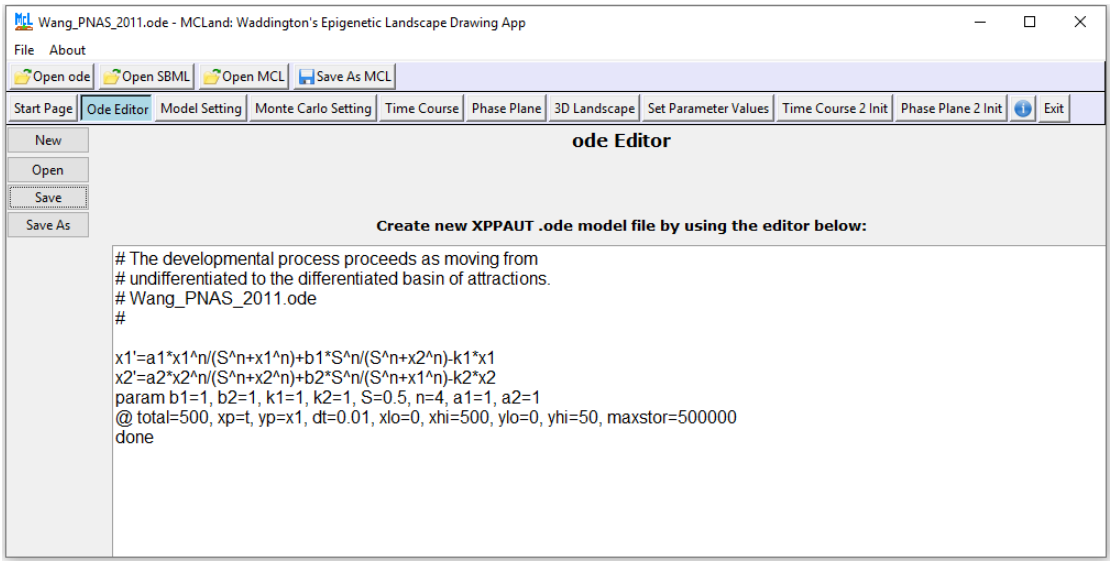

Figure 2. The ode Editor interface.

4.1 Loading a model

After you have created the ode model file, you can load the model file into MCLand by clicking the Open ode (toolbar at the top right of Figure 1) or choose from the menu File -> Load ode Model (See Figure 3). We used Python object oriented design architecture to implement the Python software tool. The model object is defined by the Model class given in Figure 4. Once the ode model file is loaded into MCLand the model information for equations, parameter values and any constants or functions will be extracted into the Model attributes and used for time course simulations, plotting of phase plane analysis and also the Monte Carlo method generation of Waddington's epigenetic landscape.

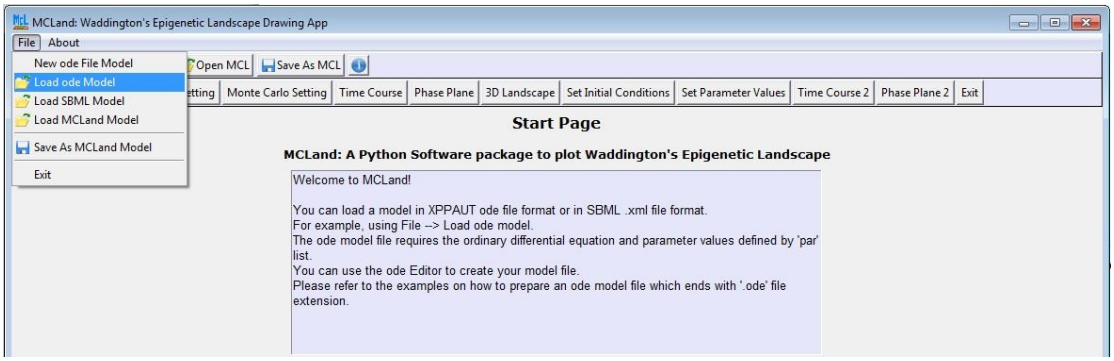

Figure 3. To load ode model using the menu.

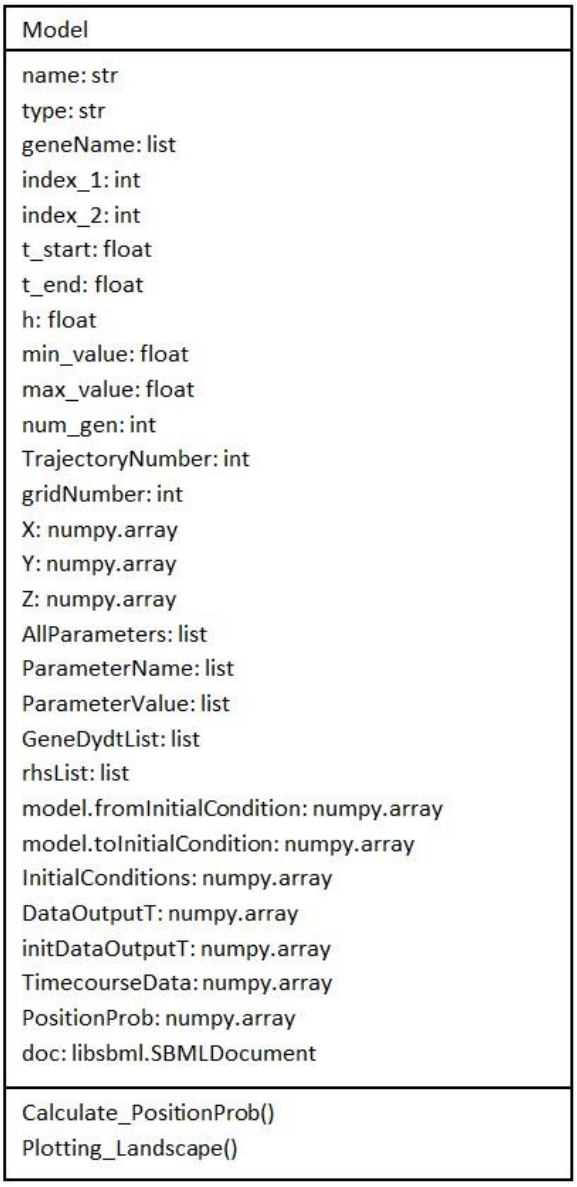

Figure 4. The class defining the Model object is used to store model information and the method to calculate and plot Waddington's landscape.

Once the ode model is loaded into MCLand the model information will be displayed in the Model Setting Page (see Figure 5). The Model Setting page function is to allow user to key in the setting for time-course simulations, phase plane analysis and Waddington's epigenetic landscape. The minimum value is usually set as 0 because protein level is assumed to be a positive real value. The maximum value is determined based on time-course simulations to get a suitable maximum range of value for plotting Waddington's epigenetic landscape. The Setting Page also allows user to select the two biomarker genes for the two dimensions of the Waddington's

epigenetic landscape. In the same way the SBML model (or MCLand model file) can also be loaded into MCLand.

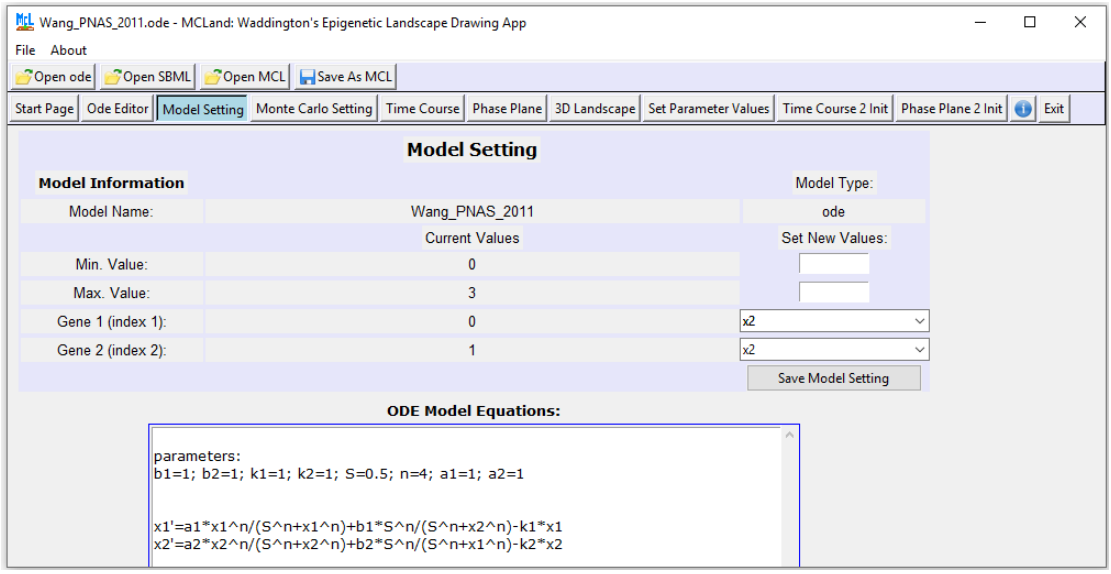

Figure 5. The Model Setting Page.

4.2 Saving a model

Here saving a model is referring to saving an ode file. User can click the ‘Save’ button (see Figure 2) to save the model equation after completing the model information key in. User can also click the Save As to save the ode model into a new model file name.

4.3 Creating a new model

Creating a new model means creating a new ode file. User can click the ‘New’ button to create an empty text area of a new model. Then enter the model information by typing in the text area.

4.4 Modifying a model

Modifying a model is referring to modifying the ode model equation information. User can edit or modify any model information and then save the final ode file by clicking the ‘Save’ button. However, when we user change the parameter value in MCLand the changes is only in MCLand pickle file (by saving landscape in MCLand file as given in Section 5.4) and not in the ode file. That means if you want to change your model parameter value you have to modify the ode file; this design is similar to the usage in XPPAUT.

5. How to plot and visualize Waddington's epigenetic landscape

5.1 Time-course simulations

To decide the maximum value for a variable and how long it takes to reach a stable steady state one needs to examine the time-course simulations. The maximum range of the value for a variable is important because it will be used as a setting for calculating the window of potential values and plotting Waddington's epigenetics landscape. The duration of the time-course simulation is also needed for obtaining a suitable time end for setting of the Monte Carlo method parameter. In other words, we to simulate long enough for the trajectory to reach stable steady state or attractor (e.g. time-course data approaching a horizontal graph as in Figure 6 where the protein levels have reached an attractor). In Figure 6 the example shows that 't_end=30' is suitable as the simulation time is long enough for the protein levels to reach the stable steady state or attractor. By examining the time-course simulations we can also observe if the dynamical system under study contains attractors or not. For example, when time-course simulations were oscillating continuously even when large time duration is used, thus indicating there is no stable steady state or attractor in the system; instead, the system may have a limit cycle attractor.

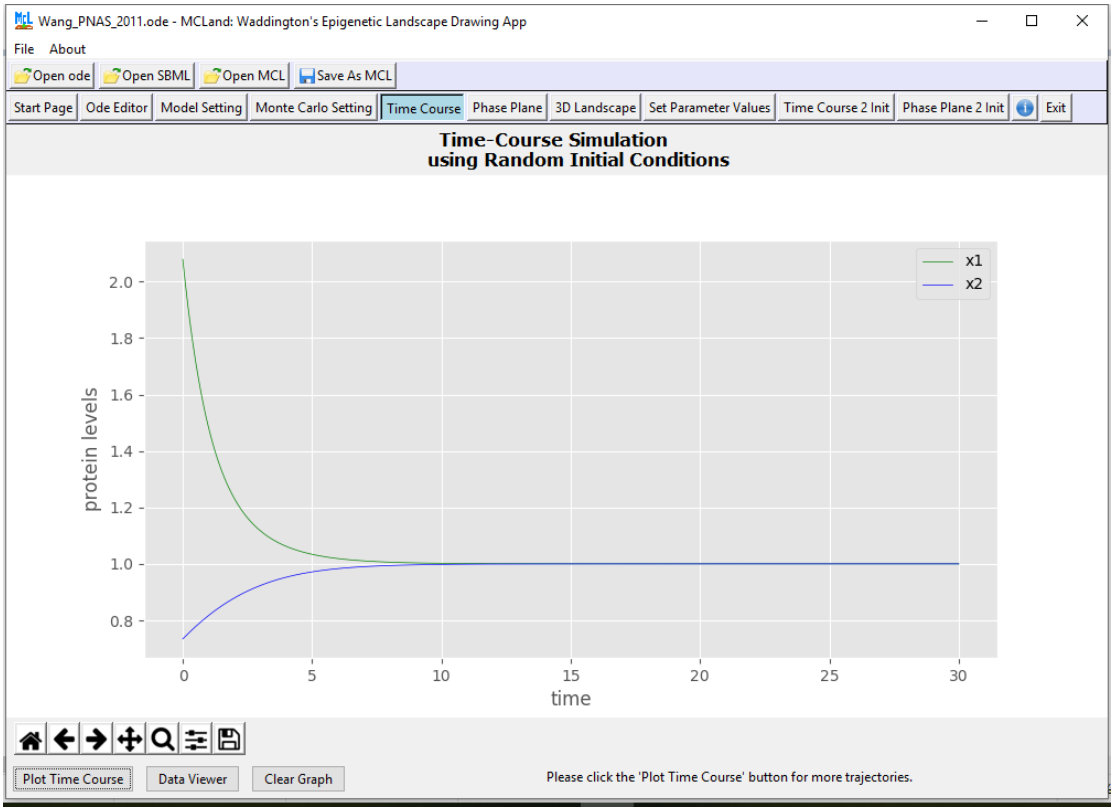

Figure 6. The Time Course Page.

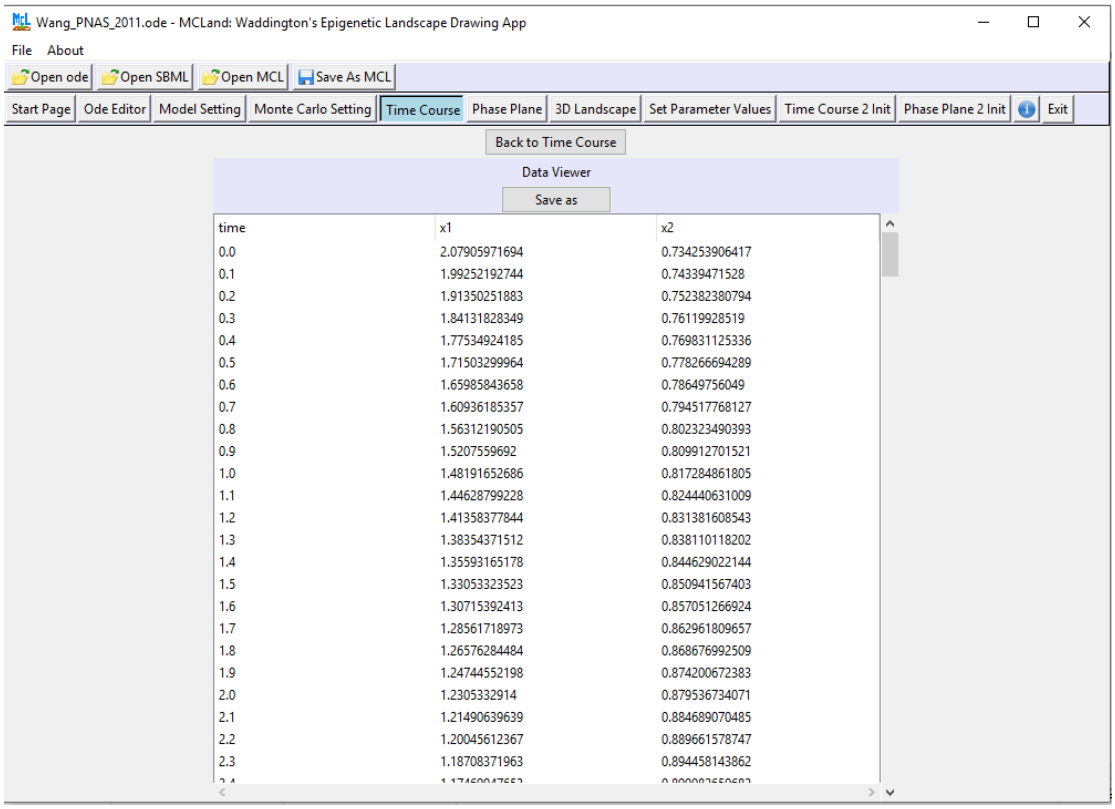

Figure 7. The Data viewer Page.

In the Time Course Page, MCLand will use random initial conditions obtained from within the state space of the dynamical systems of the model set as shown in the Figure 5 (e.g. from the range from minimum value of 0 to maximum value of 3). The data for the time-course simulation can be viewed by clicking the 'Data Viewer' button in Time Course Page (see Figure 7). An example view of the data for time-course simulation is shown in Figure 6. It is a convenient way to see the time-course data. The Data Viewer Page also allows user to save the data into a csv file.
For obtaining user key in initial conditions, user can choose the Set Initial Conditions Page and key in the initial conditions. Follow by selecting the Time Course 2 for simulating the time-course simulation with the user keyed in initial conditions. These additional functions are provided as a common function in simulating time-course graph as other software such as XPPAUT.

5.2 Phase plane analysis

Phase plane analysis is a conventional analysis to visualize the changes of two variables over time in a plane. An example of the phase plane analysis is shown in Figure 8. The phase plane shows a set of trajectories from random initial conditions and the arrows represent the direction of the trajectories with respect to time. The phase plane analysis also shows the flow of the trajectories. For example, in Figure 8 it shows that the trajectories are moving towards three attractors.

For drawing a user input initial conditions user has to first set the initial conditions and then use the Phase Plane 2 Page to draw the phase plane.

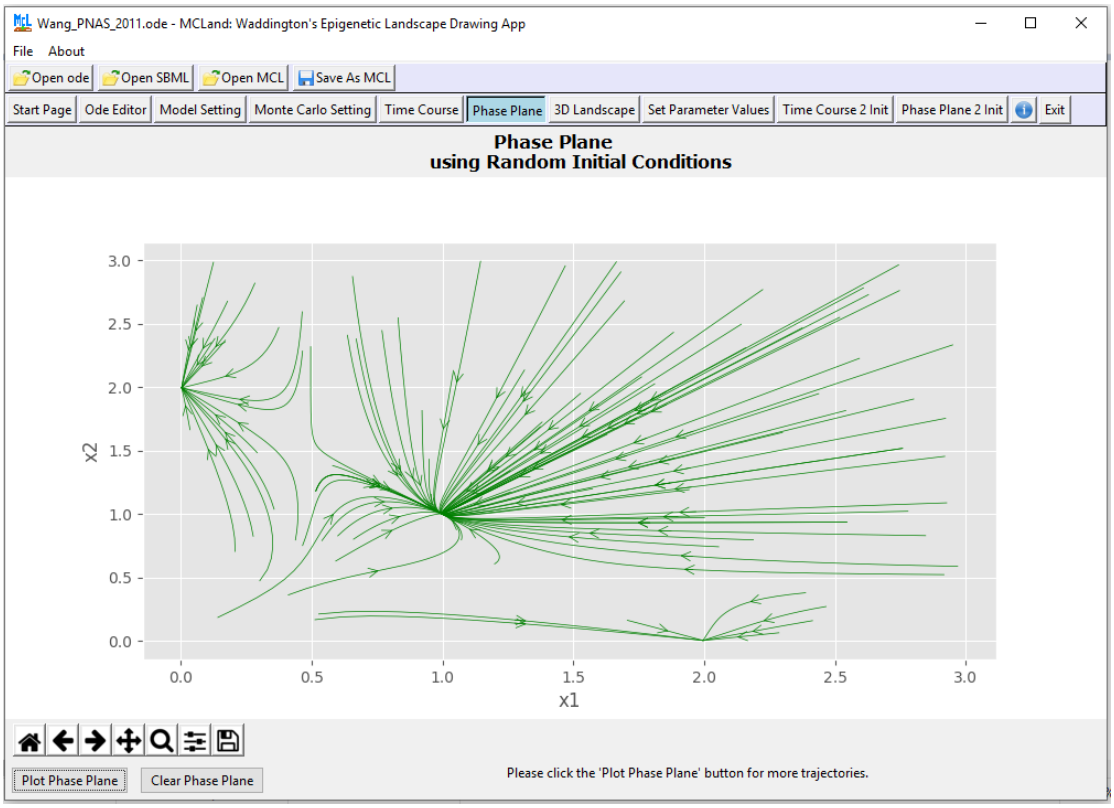

Figure 8. The Phase Plane Page.

5.3 Plotting Waddington's epigenetic landscape

For plotting the Waddington's epigenetic landscape, user needs to click the 3D Landscape Page (Figure 9). By clicking the button 'Calculate Potential U and Plot Landscape', it will execute the Monte Carlo method to simulate and generate the Waddington's epigenetic landscape for us. An example of the results of the Waddington's epigenetic landscape is displayed in Figure 10, which indicates there are three attractors. The landscape in Figure 10 is generated using 10000 number of

trajectory with random initial conditions in the state space of the systems and it will output a prototype of the Waddington's epigenetic landscape in a short period of time (e.g. within 2 minutes). For high quality Waddington's epigenetic landscape user have to navigate to the Monte Carlo Setting Page to change the number of trajectory to a higher number such as 50,000 or 100,000.

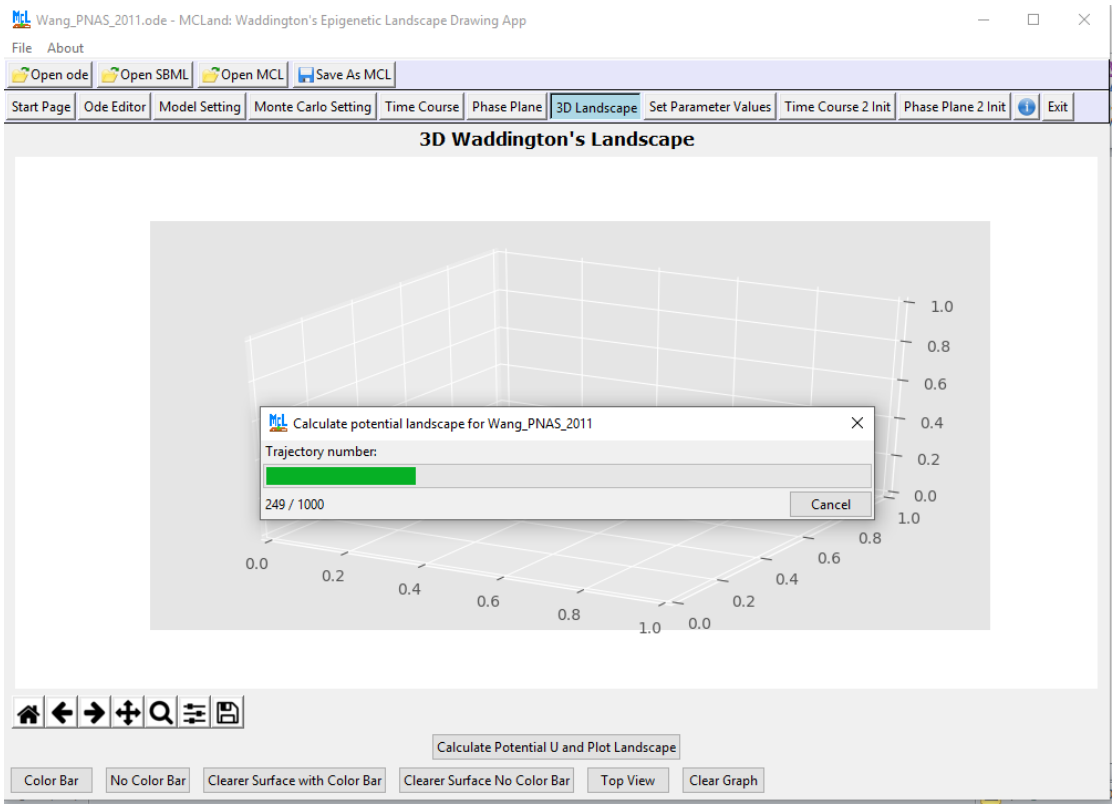

Figure 9. The progress bar showing the calculation of 3D Landscape.

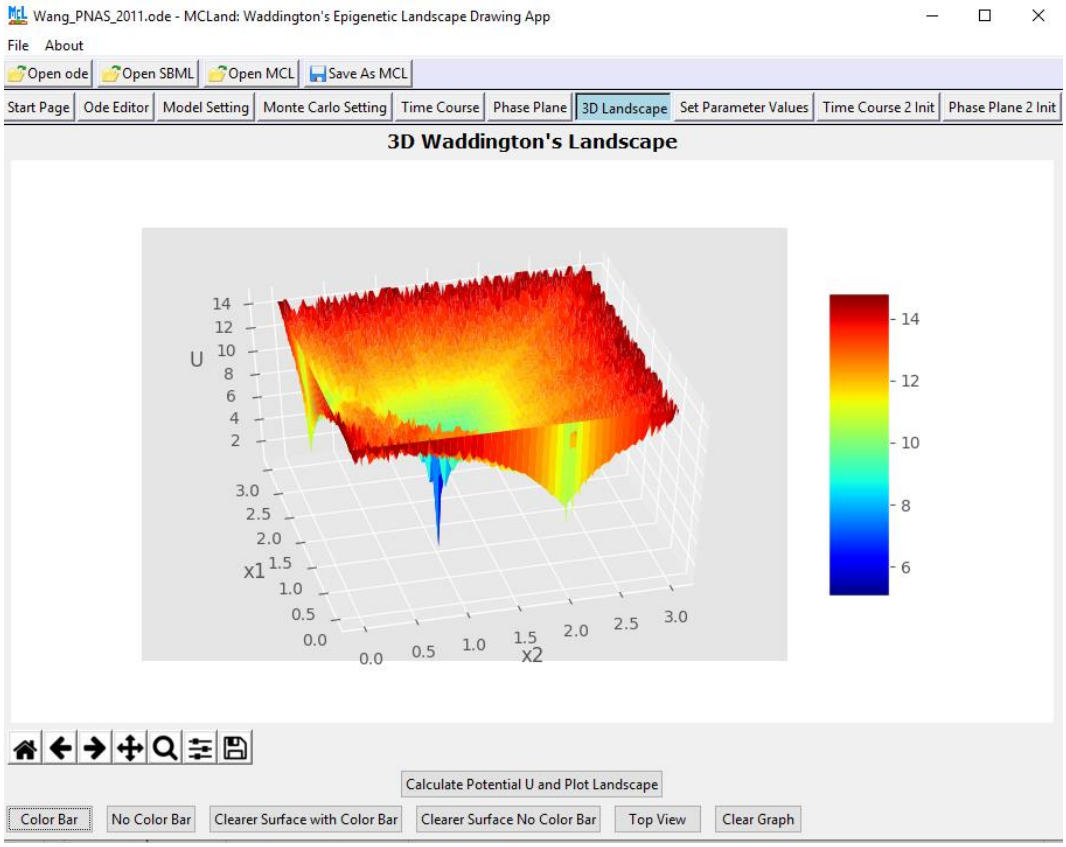

Figure 10. The 3D Landscape Page.

In the case where user wants to plot a Waddington's epigenetic landscape with a different parameter value. Then one has to change the parameter value using the Set Parameter Values Page. After the parameter value has been changed user can draw the Waddington's epigenetic landscape by going to the 3D Landscape Page.

5.4 Saving Landscape into MCLand file

After the Waddington's epigenetic landscape is generated, users can choose to save the figure into different types of file format: Encapsulated Postscript (.eps), Joint Photographic Experts Group (.jpeg), Portable Network Graphics (.png), Portable Document Format (.pdf), Scalable Vector Graphics (.svg) and Tagged Image File Format (.tif). These are the file formats provided by Python Navigation Toolbar and it is sufficient to accommodate user's preference to generate quality Waddington's epigenetic landscape for their systems modeling study. Most importantly, user can save the Waddington's epigenetic landscape into MCLand Model file for future use (see Figure 11). MCLand use Python pickle to save the information for the Waddington's epigenetic landscape coded in the model object. User can load the Waddington's

epigenetic landscape stored in MCLand model file by using the Load MCLand Model from the menu File -> Load MCLand Model.

Figure 11. The function to save into MCLand Model file.

6. Monte Carlo method parameter setting

There are four Monte Carlo method parameters [1] to be adjusted to plot a suitable Waddington's epigenetic landscape (Figure 12). The first method parameter is the numerical simulation to obtain time-course simulation. MCLand is using the **odeint** from the **scipy.integrate** package. The numerical simulation is using the default 0.1 for the time step h. User can set to a prefer time step as needed. For example, a smaller time step of 0.05 can generate more time point for time-course simulations. However, a smaller time step use may require a longer time to simulate and generate the Waddington's epigenetic landscape.

The second method parameter is the Trajectory Number where the default value is 1000 for generating a prototype of the Waddington's epigenetic landscape in a short time e.g. within 10 minute. For high quality Waddington's epigenetic landscape one needs to increase the Trajectory Number to a large number for example 50,000 or 100,000. The third method parameter is the time duration as mentioned earlier. The default time duration is 30. User may need to change the time duration depending on the time frame for achieving stable steady state or attractor. The fourth method parameter is Grid Number for the number of grid boxes in the plane. The default Grid Number is 100. Larger grid number will generate smaller grid boxes.

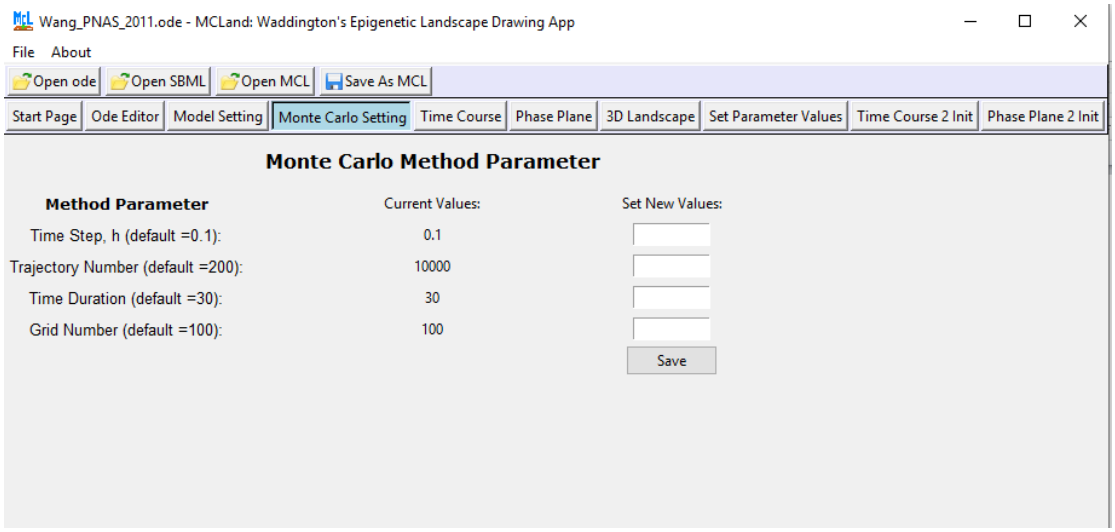

Figure 12. The Monte Carlo Method page.

7. Case Studies

7.1 Example 1: A Model from Wang et al. (2011) [3]

The first example is the biological model for development and differentiation from Wang et al. (2011), which contains two genes x1 (PU.1) and x2 (GATA1) that are mutually inhibiting each other and self-activating (Figure 13). PU.1 and GATA1 are transcription factors. The ode model Wang_PNAS_2011.ode can be found in the 'ode_model' folder. The 3D Waddington's landscape is given in Figure 14 which shows three attractors: multipotent stem cell or myeloid progenitor cell, myeloid cell and erythroid cell. Co-expression of PU.1 and GATA1 correspond to the multipotent stem cell attractor in the middle in Figure 14 [3]. PU.1 promotes the myeloid cell and GATA1 promotes erythroid cell; that means high expression of PU.1 induced multipotent stem cell differentiated into myeloid cell and high expression of GATA1 induced multipotent stem cell differentiated into erythroid cell [3].

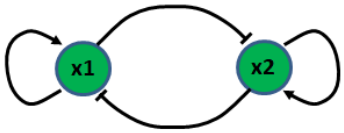

Figure 13. The gene regulatory network proposed by Wang et al. (2011)[3].

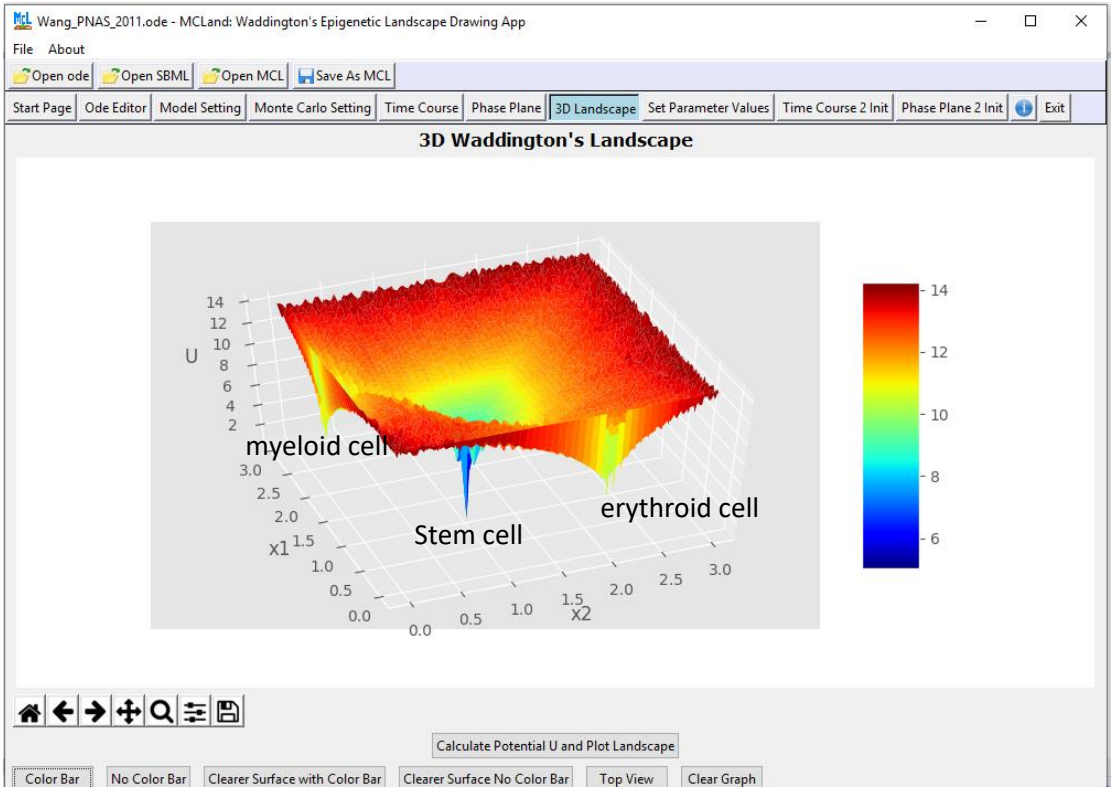

Figure 14. The 3D Landscape generated from 20,000 trajectory number. The attractor in the middle is the multipotent undifferentiated cell which co-expressed PU.1 and GATA1. High expression of x1 or PU.1 represents myeloid cell and high expression of x2 or GATA1 represents erythroid cell [3].

Model equations from Wang et al. (2011) are:

$$\frac{dx_1}{dt} = \frac{a_1 (x_1)^4}{S^4 + {x_1}^4} + \frac{b_1 (S^4)}{S^4 + {x_2}^4} - k_1(x_1)$$

$$\frac{dx_2}{dt} = \frac{a_2 (x_2)^4}{S^4 + {x_2}^4} + \frac{b_2 (S^4)}{S^4 + {x_1}^4} - k_2(x_2)$$

where $a_1 = 1$, $a_2 = 1$, $b_1 = 1$, $b_2 = 1$, $S = 0.5$, $k_1 = 1$, and $k_2 = 1$.

7.2 Example 2: A Model from Li and Wang (2015) [4]

The second example is the biological model for cancer stem cell gene regulatory network from Li and Wang (2015), which contains six genes or proteins (Figure 15). The six proteins are ZEB, Oct4, MDM2, P53, miR-145 and miR-200. The ode model file is Li_and_Wang_CancerRes_2015.ode can be found in the 'ode_model' folder. According to the Waddington's landscape from Li and Wang (2015) there are four attractors: stem cell, cancer stem cell, normal and cancer. However, based on the Monte Carlo method from Zhang et al. (2018) [1] each of the attractor become a pair of attractors which is shown in Figure 16. The Monte Carlo method may have captured the correct number of attractors that are being approximated into four attractors as in Li and Wang (2015) [4]. High P53 or miR-200 or miR-145 expression denotes normal cell state, and the low P53 or miR-200 or miR-145 expression denotes cancer cell state [4]. Meanwhile high ZEB or OCT4 expression denotes stem cell state and on the other hand low ZEB or OCT4 expression denotes non-stem-cell state [4]. Li and Wang (2015) used ZEB and OCT4 as biomarkers to differentiate cancer state and cancer stem cell state.

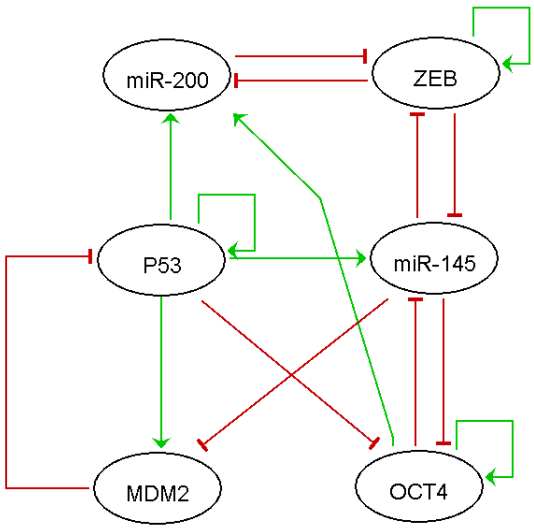

Figure 15. The gene regulatory network of cancer stem cell proposed by Li and Wang (2015) [4].

The model equations for Li and Wang (2015) are given below:

$$\frac{\mathrm{dP53}}{\mathrm{dt}} = \frac{\mathrm{sa}\ (\mathrm{P53})^4}{\mathrm{S^4} + \ \mathrm{P53^4}} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ \mathrm{MDM2^4}} - \mathrm{k(P53)}$$

$$\frac{\mathrm{dMDM2}}{\mathrm{dt}} = \frac{\mathrm{a}\ (\mathrm{P53^4})}{\mathrm{S^4} + \ \mathrm{P53^4}} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{miR145})^4} - \mathrm{k(MDM2)}$$

$$\frac{\mathrm{dOCT4}}{\mathrm{dt}} = \frac{sa_1\ (\mathrm{OCT4})^4}{\mathrm{S^4} + \ (\mathrm{OCT4})^4} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ \mathrm{P53^4}} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{miR145})^4} - \mathrm{k(OCT4)}$$

$$\frac{\mathrm{d(miR145)}}{\mathrm{dt}} = \frac{\mathrm{a}\ (\mathrm{P53^4})}{\mathrm{S^4} + \ (\mathrm{P53})^4} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{OCT4})^4} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ \mathrm{ZEB^4}} - \mathrm{k(miR145)}$$

$$\frac{\mathrm{d(ZEB)}}{\mathrm{dt}} = \frac{sa_2(\mathrm{ZEB^4})}{\mathrm{S^4} + \ \mathrm{ZEB^4}} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{miR145})^4} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{miR200})^4} - \mathrm{k(ZEB)}$$

$$\frac{\mathrm{d(miR200)}}{\mathrm{dt}} = \frac{\mathrm{a}\ (\mathrm{P53^4})}{\mathrm{S^4} + \ (\mathrm{P53})^4} + \frac{\mathrm{a}\ (\mathrm{OCT^4})}{\mathrm{S^4} + \ (\mathrm{OCT4})^4} + \frac{\mathrm{b}\ (\mathrm{S^4})}{\mathrm{S^4} + \ (\mathrm{ZEB})^4} - \mathrm{k(miR200)}$$

where $a = 0.5$, $sa = 0.8$, $sa_1 = 0.8$, $sa_2 = 0.8$, $b = 0.1$, $S = 0.5$ and $k = 1$.

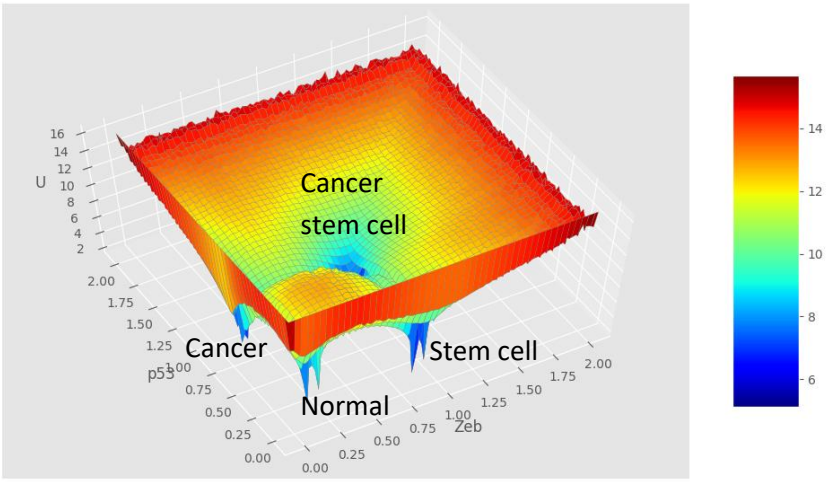

Figure 16. The 3D Landscape generated from 100,000 trajectory number save in Tagged Image File Format (.tif).

8. License

MCLand is distributed with the GPL-2.0 license. MCLand is a free software package for academic and research usage.

9. Contact information bug reporting

For comments and bug reporting you are welcome to write to us:

zhengjie@shanghaitech.edu.cn

Your feedback is valuable for us to improve the current version of MCLand.

References

1. Zhang X, Chong KH, Zheng J (2018) A Monte Carlo method for in silico modeling and visualization of Waddington's epigenetic landscape with intermediate details. bioRxiv: 310771.
2. Ermentrout B (2002) Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students: Siam.
3. Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. Proceedings of the National Academy of Sciences, USA 108: 8257-8262.
4. Li C, Wang J (2015) Quantifying the Landscape for Development and Cancer from a Core Cancer Stem Cell Circuit. Cancer research: canres. 0079.2015.