User Manual of Multi-Layers Model Matlab package

March 13, 2008

Use of Multi Layer Model and its parameter settings.

The Multi-Layers Model (MLM) software is composed by several files and functions; in the main file *ml_method.m* it is possible to specify a generic dataset input file and to set all the model parameters. The default name for the input dataset is *hybdata.mat*, it is a matlab mat file which stores by a record named Z the following fields:

- Z.hyb: hybridization values;
- Z.contigs: $n \times 2$ matrix that contains in each row the bounds (in base pairs) of each one of the n chunks of contiguous signal fragments stored in Z.hyb.

The input dataset file must be placed in the subdirectory *./dataset*. In Table 1, a detailed list of all the MLM parameters is reported. An input dataset example is already present in the *./dataset* directory of the software package.

Table 1: Parameters of MLM

parameter	**description**
contigs_list	Contigs (Contiguous chunks of signal) list of input signal to classify.
plot_on_off	Enable {1} or disable plot {0}.
nucleosome_length	Nucleosome length.
probe_resolution	Resolution of the microarray in base pairs.
probe_overlap	Overlap between probes in base pairs.
n_cut	Number of thresholds.
permanence_percentage	Percentage of the total number of thresholds used to judge a pattern interesting.
smoothing_on_off	Enable or disable smoothing.
smoothing_window	Window used for smoothing (see Table 2 for some examples).
bias_correction_area	For test purpose set to 0.
bias_correction_interval	For test purpose set to 0.
k	The parameter in the dissimilarity function.
offset_around_max	Number of probes around a maxima point.
k_std	Used to set the range for the linker class.

Table 2: Smoothing windows

window	**description**
$[1, 2, 1]/4$	Three points smoothing.
$[-3, 12, 17, 12, -3]/35$	Five points smoothing.
$[-2, 3, 6, 7, 6, 3, -2]/21$	Seven points smoothing.
$[-21, 14, 39, 54, 59, 54, 39, 14, -21]/231$	Nine points smoothing.
$[-36, 9, 44, 69, 84, 89, 84, 69, 44, 9, -36]/429$	11 points smoothing.
$[-11, 0, 9, 16, 21, 24, 25, 24, 21, 16, 9, 0, -11]/143$	13 points smoothing.
$[-78, -13, 42, 87, 122, 147, 162, 167, 162, 147, 122, 87, 42, -13, -78]/1105$	15 points smoothing.
hamming(n_point)	Hamming window of n points.

The output of the MLM are :
(a) A structure named *classification* that contains, for each contiguous fragment of signal (contig), the class (*well positioned, decentralized, fuzzy*) and the coordinates of each of its nucleosomes.
(b) A binary string named *nucleosome_regions_map* in which, for each probe, 1 means *nucleosome*, 0 *linker*.
The output is automatically saved in a .mat file named *./output/output_of_<input_file_name>_<date>.mat* where *<input_file_name>* is the input file name and *<date>* is the current date.
Setting the *plot_on_off* parameter to 1 will produce 3 plots similar to the ones shown on Fig. 1, 2, 3 .

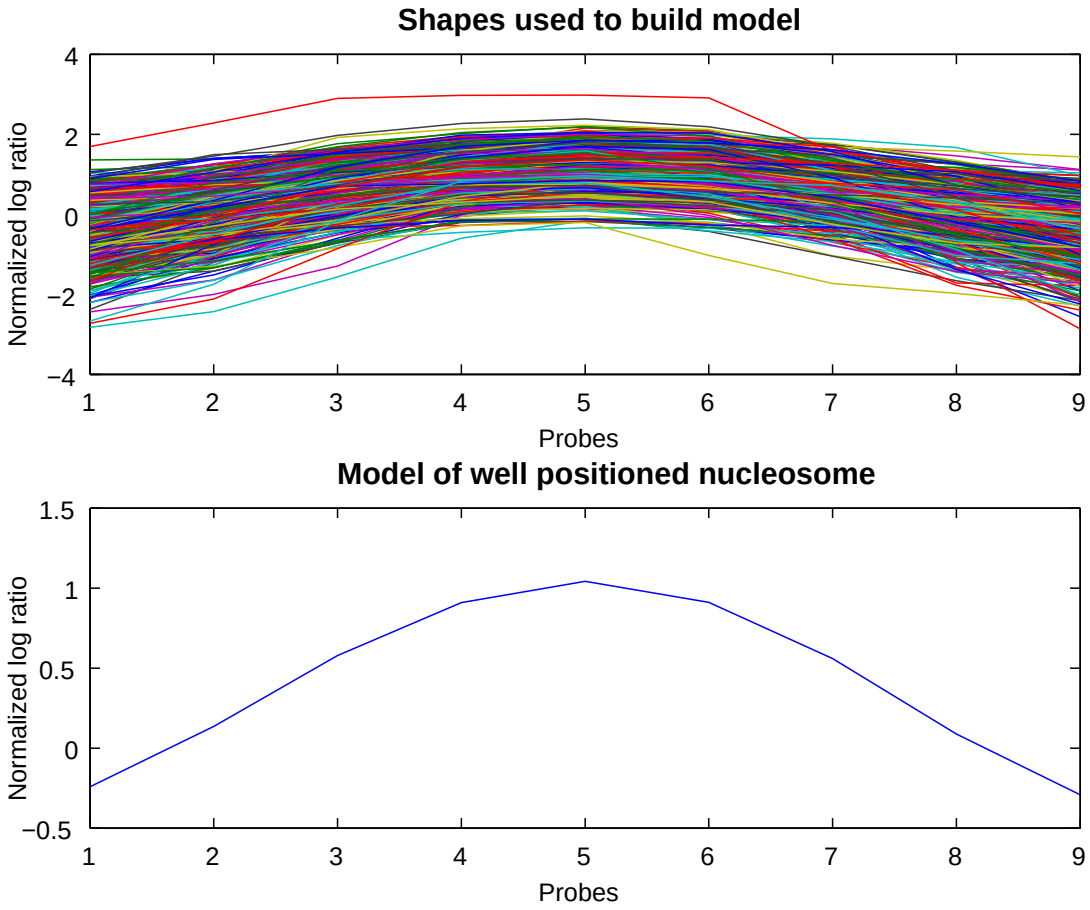

Figure 1: Building of the well positioned nucleosome model

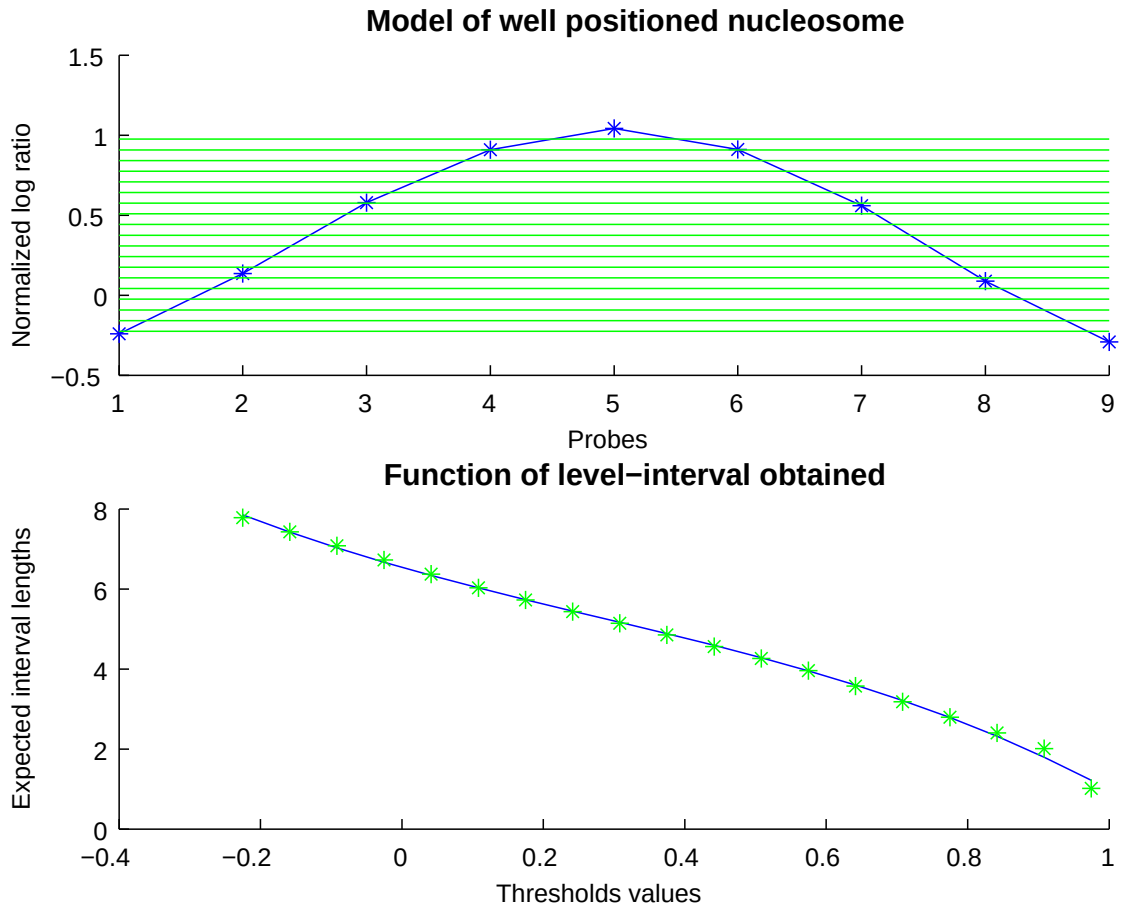

Figure 2: Building of the function that maps threshold levels with expected interval lengths.

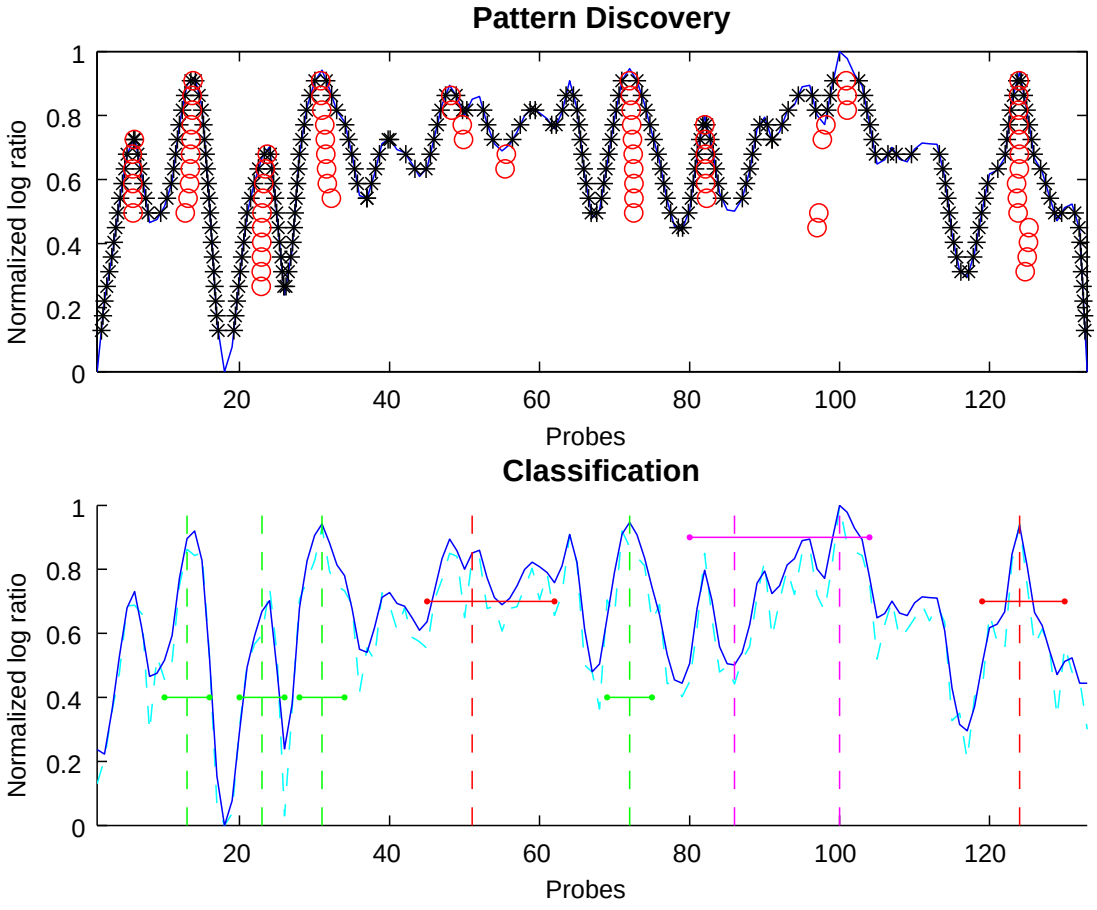

Figure 3: Plot of MLM patterns discovery and classification. In particular, the top plot shows the discovered pattern, while the bottom one reports the nucleosome classification (green intervals means *well positioned*, red intervals means *decentralized*, purple intervals means *fused*).

Microarray Synthetic Signal Generator.

The Synthetic Signal Generator is able to generate signals respecting a particular tiling microarray approach that has been recently used to face with nucleosome identification (for details see [1]). The matlab code implementing such task is *synthetic_microarray.m*.
The prototype of the function is:

function [signal,ideal_signal,signal_map]=synthetic_microarray(number_of_nucleosomes,nucleosome_length,...
expected_time_between_nucleosomes,resolution,overlap,n_rep,decentralized_percentage,...
decentralized_range,purification,noise_spot_variance,SNR,relative_abundance,plot_on_off)

The input parameters of this function are listed in Table 3. It outputs the synthetic signal (*signal*), a binary string (*ideal_signal*) that represents the truth data classification (1 means *nucleosome* while 0 *linker*) and a cell array (*signal_map*) that contains base pairs values for each nucleosome starting and ending, together with its probe coordinates and a label to distinguish between *well positioned* and *delocalized* nucleosomes.

Table 3: Smoothing windows

parameter	**description**
plot_on_off	Plot (1) or not (0) figures.
number_of_nucleosomes	Number of nucleosomes we want to add to the synthetic signal.
nucleosome_length	Length of a nucleosome.
expected_time_between_nucleosomes	Mean of the poisson distribution used to model the expected distances between adjacent nucleosomes.
resolution	Resolution of microarray in base pairs.
overlap	Overlap between probe in base pairs.
n_rep	Number of spotted copies (replicates) of nucleosomal and genomic DNA on each probe of the microarray.
decentralized_percentage	Percentage of the delocalized nucleosomes over the total number of nucleosomes.
decentralized_range	Represents the range which limits the delocalization of a nucleosome in each copy of nr. It is defined in base pairs.
purification	Percentage of DNA purification, which is the probability that each single DNA fragment of the nr copies appears in the microarray hybridization.
noise_spot_variance	Variance of the green signal in each probe, even in absence of nucleosomes due to the cross hybridization.
SNR	Linear signal to noise ratio of the synthetic signal to generate. Note that the noise is assumed to be gaussian.
relative_abundance	Relative abundance between nucleosomal and genomic DNA.

Testing MLM and Micorarray Synthetic Signal Generator

You can test the MLM and Microarray Synthetic Signal Generator packages using the example script *MLM_Example_On_Syntethic_Signal.m*. In particular, it generates a synthetic signal, analyzes it by MLM and outputs precision and recognition confusion matrices.

Bibliography

[1] Yuan G-C., Liu Y. J., Dion M. F., Slack M. D., Wu L. F., Altschuler S. J., Rando O. J. (2005), Genome-Scale Identification of Nucleosome Positions in S. cerevisiae, *Science*, **309**, 626–630.