Likelihood landscape plots

1 Introduction

In this document, we will explain how likelihood landscape plots are produced by the `--pl/plot-likelihoods` option of sMap and provide some examples on how they can be interpreted.

The likelihood landscape plots represent which parameter values have been sampled during maximum-likelihood estimation and the corresponding log-likelihood value.

The way the plots are produced depends on how many parameters are being estimated by maximum-likelihood. If only one parameter is being estimated (e.g. in the case of an ER model), a simple 1-D plot is produced (**Figure 1A**), in which the horizontal axis represents the parameter value and the vertical axis represents the corresponding log-likelihood. Intervals of this parameter for which the likelihood has not been computed are highlighted in the plot by question marks.

If two parameters are being estimated, a 2-D plot is produced (**Figure 1B**) with the two axes corresponding to the parameters being estimated. The colour of each "point" (i.e., little square) in the plot corresponds to the log-likelihood for the corresponding parameter values. Parameter values that have not been sampled appear as grey areas.

If more than two parameters are being estimated, first of all the sampled parameter vectors are projected into a 2-D space, with the aim of representing most of the variation in the log-likelihood function (see 1.1). This 2-D space is represented in the plot (**Figure 1C**), in which each parameter vector is associated with a "point" (little square) that is coloured according to the log-likelihood value. If two or more parameter vectors correspond to the same point, the point is coloured according to the highest log-likelihood value between them. Points that do not correspond to any sampled parameter vectors appear as grey areas.

To the right of the plot, there is a summary of how much of the function's variation the two components that are plotted explain; below this is a projection of the orientation of the parameters being estimated in the 2-D space of the plot.

The first page of the PDF produced by sMap contains a plot of all the parameter values that have been sampled. The second page contains only the values for which the likelihood falls within 5 log-units of the maximum-likelihood estimate (MLE), while the third page contains only the values within 1 log-unit of the MLE. In all plots, a red circle highlights the MLE, while a red rectangle encompasses the parameter values that are plotted in the following page.

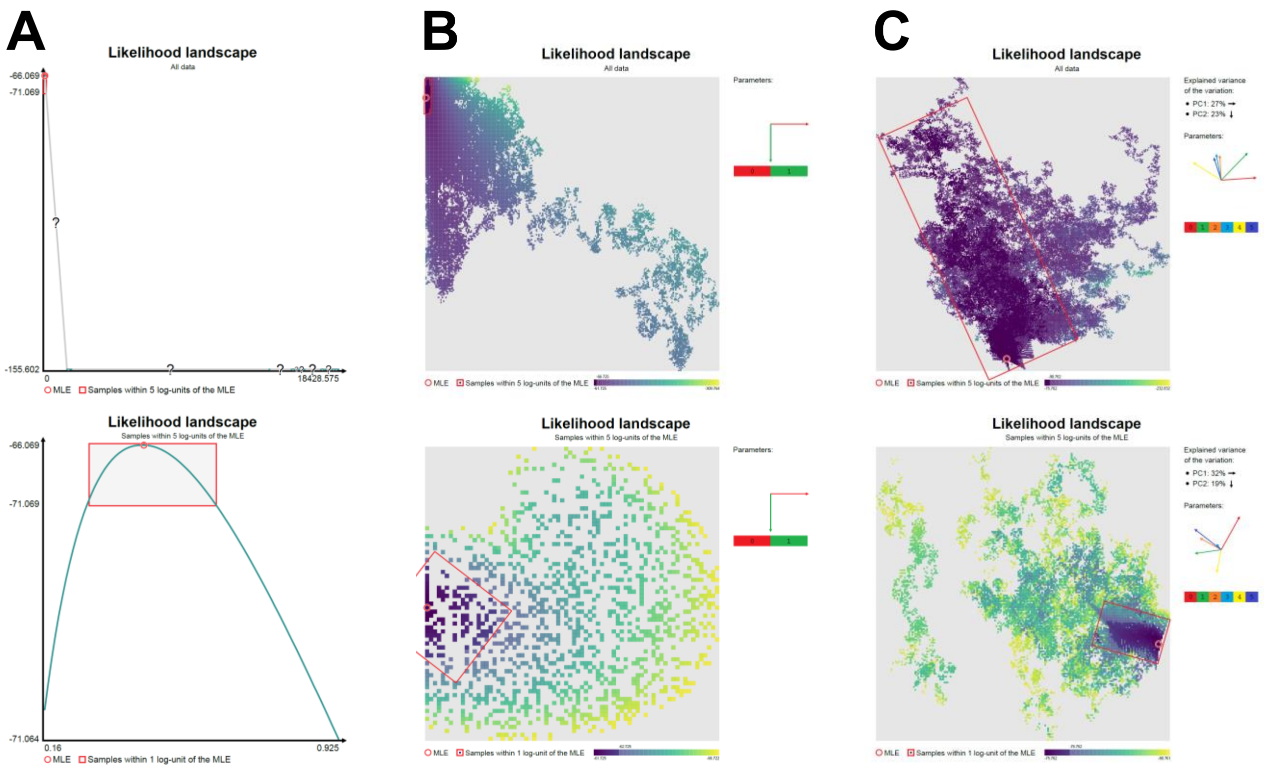


Figure 1. Examples of likelihood landscape plots produced by sMap. **A**: Likelihood function depending on only one parameter. **B**: Likelihood function depending on two parameters. **C**: Likelihood function depending on six parameters. The top row represents plot produced using all sampled parameters, the bottom row represents plots produced only samples within 5 log-units of the maximum-likelihood estimate (MLE).

1.1 Representing a function of many variables in a 2-D plot

To represent a function f of many variables in a 2-D plot, we use an approach that is similar to (and involving) principal component analysis (PCA).

Our objective is, given a set of parameter vectors $\boldsymbol{x_i}$ and the value of f computed at those vectors, to perform a basis change $\boldsymbol{x_i} \to \boldsymbol{x_i'}$ so that the first two components of the $\boldsymbol{x_i'}$ correspond to the directions of “maximum change” in the function, without performing any additional function evaluations (i.e. likelihood computations).

To find the directions of maximum change in the function, we would ideally use the function’s gradient. This is not available, and we do not wish to perform any additional function evaluations, thus we use the following approach:

- Randomly choose 1000 pairs of sampled values, biasing this choice so that pairs of points that are closer are more likely to be chosen than pairs of points that are far away.
- For each pair of points $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, compute the vector[1]:

$$\boldsymbol{v_{ij}} = \left(f(\boldsymbol{x_i}) - f(\boldsymbol{x_j})\right)\frac{\boldsymbol{x_i} - \boldsymbol{x_j}}{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|}$$

- Perform a PCA on the $\boldsymbol{v_{ij}}$. The results of the PCA are a transformation matrix $\boldsymbol{A}$ and a vector of “loadings”.

[1] Note that if f is differentiable, by the mean value theorem there must exist a point where the directional derivative of f along the direction $\boldsymbol{x_1} - \boldsymbol{x_2}$ is equal to $f(\boldsymbol{x_1}) - f(\boldsymbol{x_2})$.

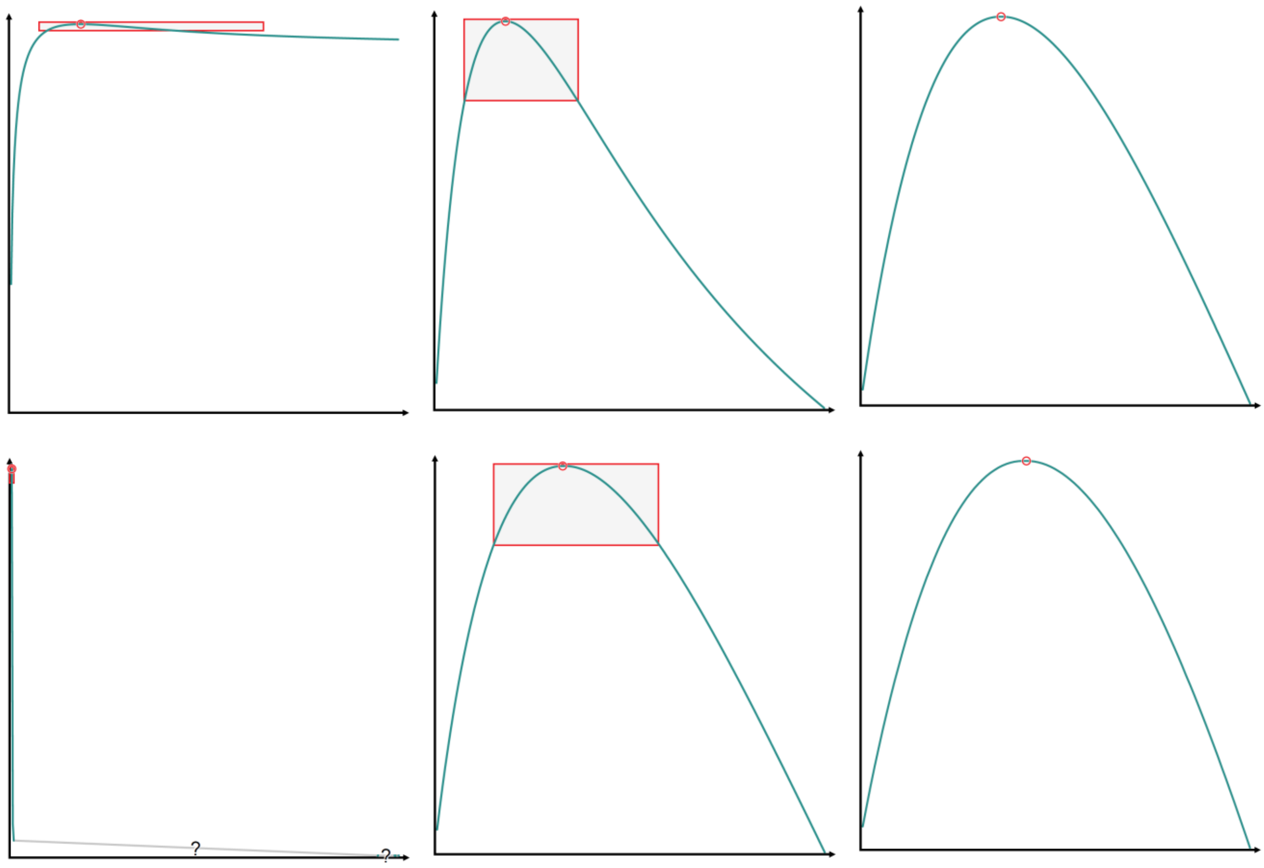
- $\boldsymbol{A}$ is used to perform the change of basis: $\boldsymbol{x}_i' = \boldsymbol{A}\boldsymbol{x}_i$. The loadings represent how much of the variance in the $\boldsymbol{v}_{ij}$ is explained by each component in the new basis. The loadings for the first two components are shown to the right of the plot produced by sMap.
- The first two components of the $\boldsymbol{x}_i'$ are preserved, while the rest are discarded. These two components are used to plot $f(\boldsymbol{x}_i)$ as if it were a function of 2 variables.

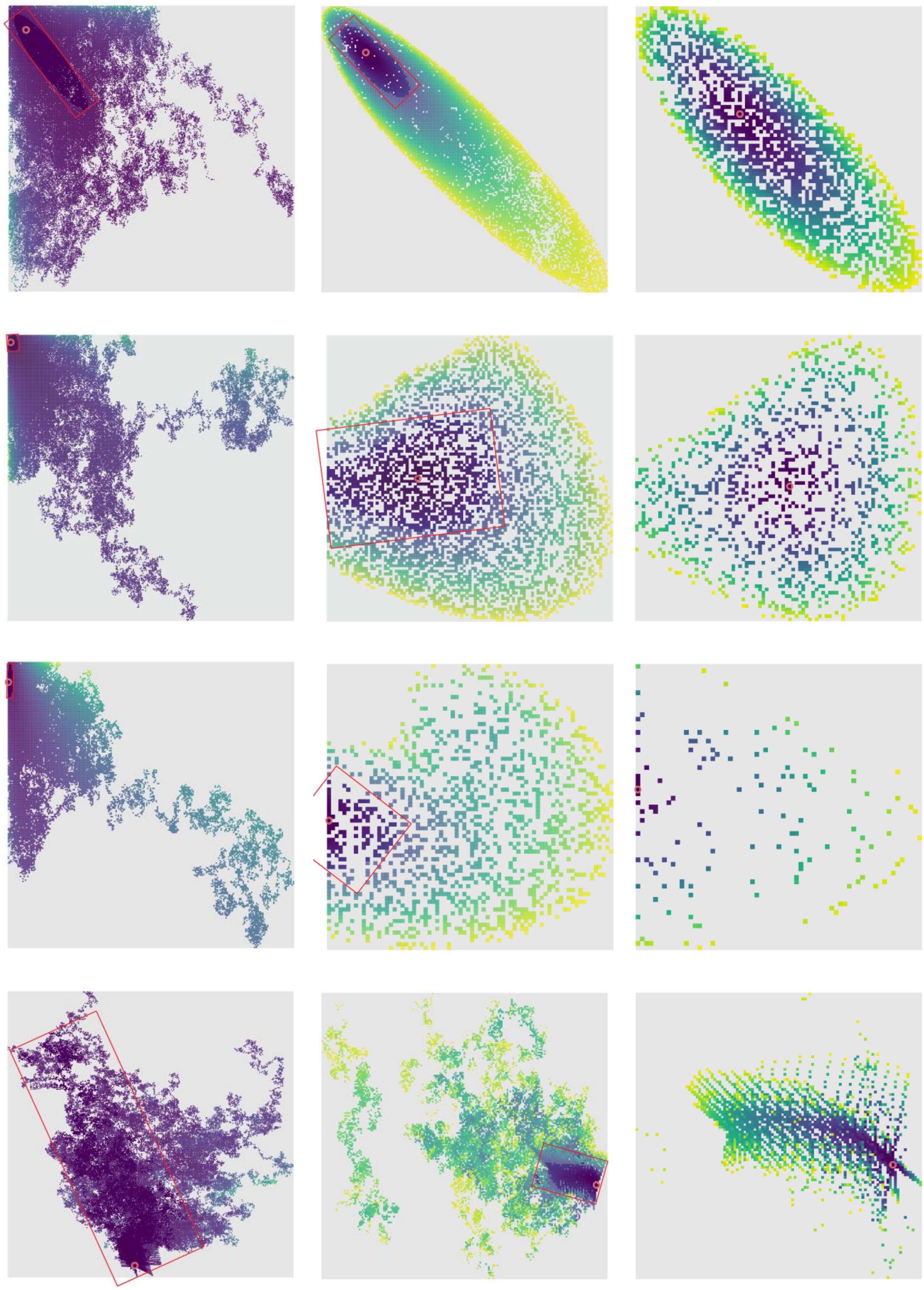
2 Examples of likelihood plots

2.1 Examples of "good" plots

In this paragraph, we present some examples of plots showing "good" maximum-likelihood estimate runs. In these plots, you will notice that the MLE is never in an isolated point and there are instead always a good number of high-likelihood points that have been sampled around the MLE.

Each row contains plots from the same analysis, the first column contains the plots obtained using all the sampled values (i.e. from the first page of the PDF produced by sMap); the second column contains only values within 5 log-units of the MLE (second page of the PDF) and the third column contains only values within 1 log-unit of the MLE (third page of the PDF).

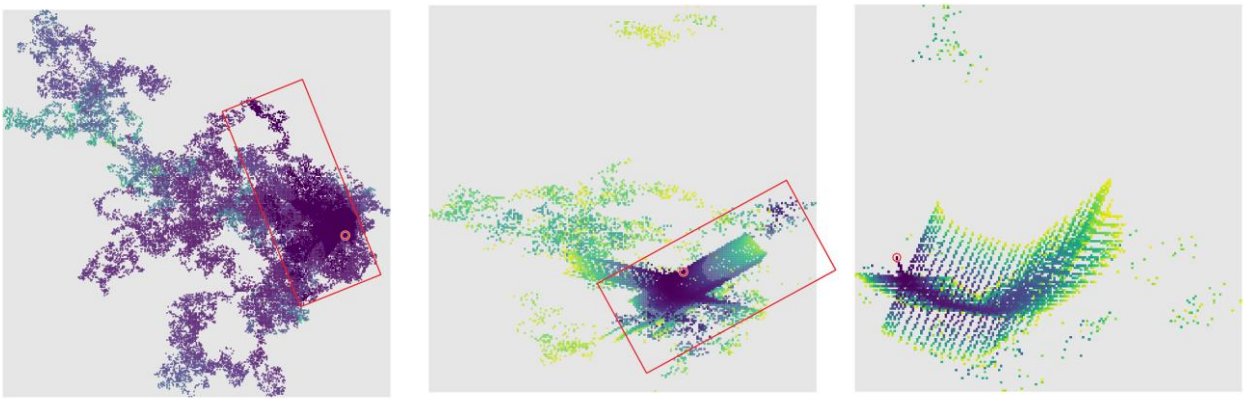




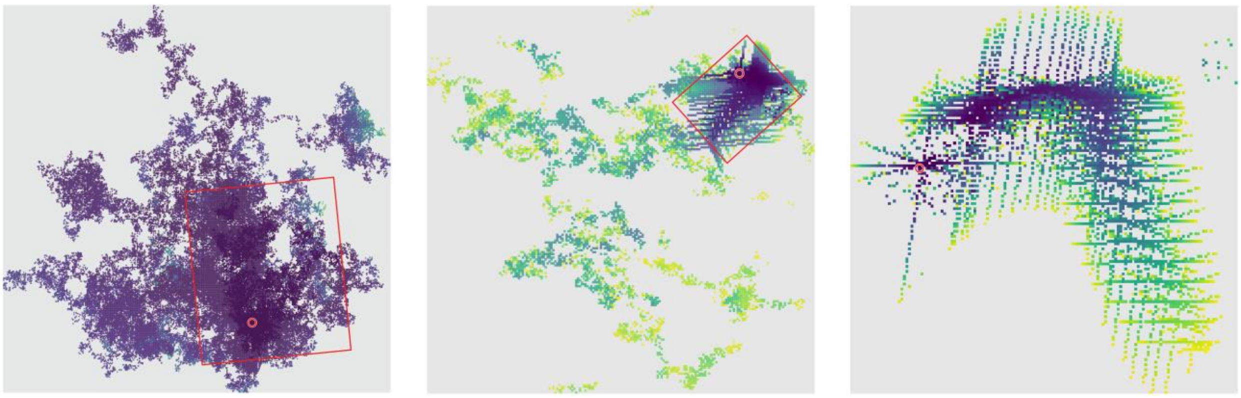
2.2 Examples of plots that need improvement

The plots in this section show examples of analyses that should be improved. Below each set of plots, there is a description of what is wrong with the plot and what changes to the analysis can be made to improve the reliability of the results, as well as the plots obtained by implementing these changes.

2.2.1 Maximum reached by hill-climbing

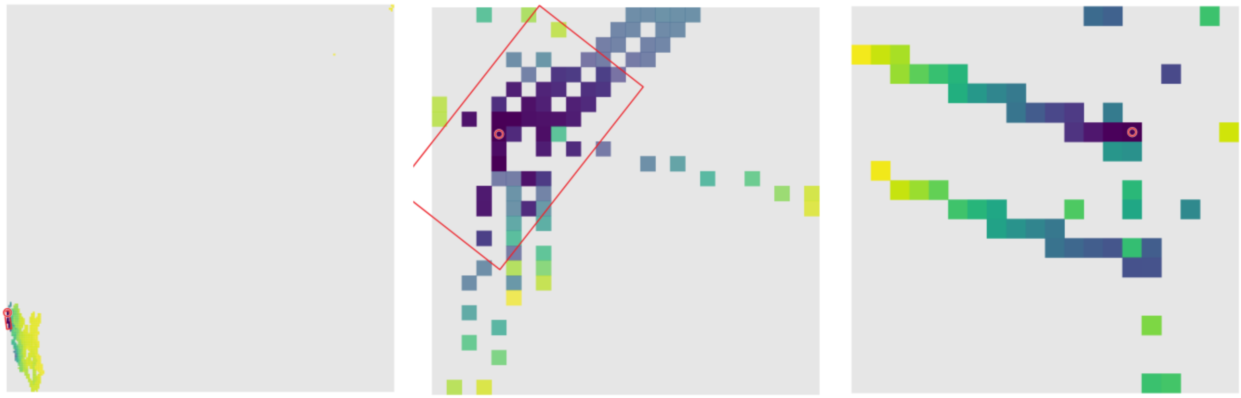


These plots are not actually extremely bad. The main issue, as is most visible in the last plot, is that the MLE was reached by hill-climbing, thus parameter values around the MLE have not been sampled adequately. This can be addressed by increasing the number of optimisation rounds with the --mr/ml-rounds option. The following plots were obtained by performing the same analysis with two consecutive rounds of ML optimisation:



In this case, there is an adequate number of samples that have been collected in the neighbourhood of the MLE.

2.2.2 Likelihood with a narrow peak



In the case of this plot, the likelihood function has a very narrow peak near low values of the rate parameters. This is apparent because the likelihood changes sharply over a small range of parameter values and because the plots on the second and third page contain very few points. This situation can be improved in multiple ways: one possibility is to normalise the tree length so that the MLE of the rates increases (if the length of the tree is divided by 10, the MLE of the rates is multiplied by 10); alternatively, it is also possible to tweak the ML optimisation strategy to increase the resolution at low rates. The following plots have been obtained by adding to the ML optimisation strategy two instances of “Iterative sampling”: one sampling with a resolution of 0.01 and one with a resolution of 0.001.