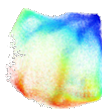# "How To EmbedSOM"

EmbedSOM plugin for FlowJo

EmbedSOM is a dimensionality reduction algorithm that helps to quickly visualize your high-dimensional data in two dimensions. The design is based off FlowSOM, which is another popular unsupervised manifold learning algorithm that uses self-organizing maps (SOMs) to find clusters in the dataset. EmbedSOM uses the same manifold learning method to gain information about an approximate manifold that describes the high-dimensional space, and uses it to quickly compute a low-dimensional embedding to help visualize the dataset.
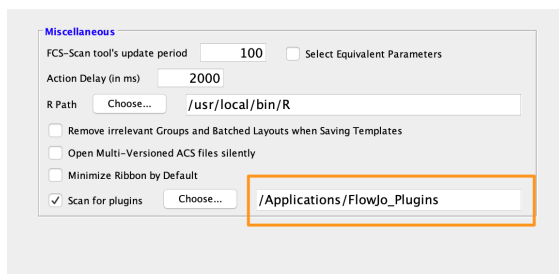
This updated plugin adds new options for the EmbedSOM algorithm in addition to its previous functions which allow users to load in a previously computed SOM from the FlowSOM clustering plugin. EmbedSOM will now compute its own embedding to generate a SOM as well as allow for the use of tSNE or UMAP landmark generation which can be used in place of a SOM to quickly generate the low-dimensional visualization.

The EmbedSOM algorithm has been developed and implemented as a R package by Miroslav Kratochvíl & Petr Ryšavý and the source code is available here: https://github.com/exaexa/EmbedSOM. You may read more about EmbedSOM in the research paper here:

Miroslav Kratochvíl, Abhishek Koladiya, and Jiří Vondrášek. "Generalized EmbedSOM on quadtree-structured self-organizing maps" F1000Research 8 (2019). doi:10.12688/f1000research.21642.2

Download and installation

1. Place the plugin .jar file in your Plugins folder, and direct FlowJo to that folder using the Diagnostics section of the Preferences.

2. Make sure you have R installed and the R path is specified in the R Path field of the Diagnostics section of the Preferences.
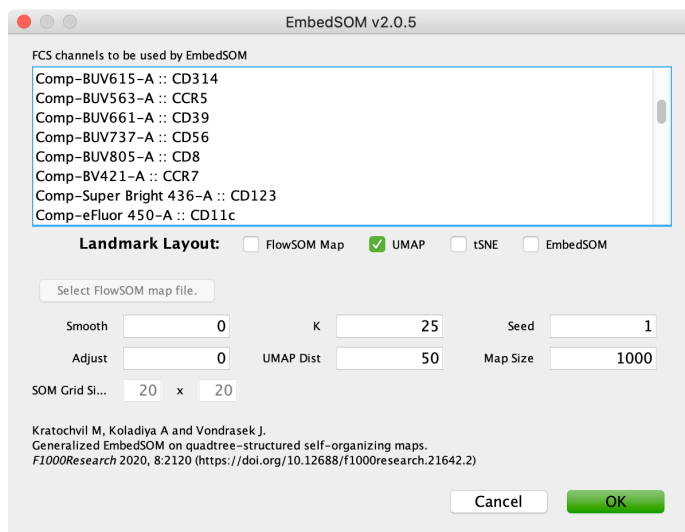

3. Running the plugin for the first time will install the needed R packages to allow the calculation to run in the R environment. Sometimes these installations can fail and you may need to manually install the R packages. To install the required R packages, use the following commands in R:

```
install.packages(c("ggplot2", "FNN", "igraph", "Matrix", "Rtsne", "umap",
                   "uwot", "utils", "devtools"))

devtools::install_github('exaexa/EmbedSOM')
```

Note This plugin was tested in R versions 4.1 & 4.2 and EmbedSOM version 2.1.1.

Usage

To run the EmbedSOM plugin on your FCS file, select the population of interest within the workspace. Go to the Workspace tab and select the EmbedSOM option from within the Plugins drop-down menu. This will bring up a dialog where researchers can choose which compensated parameters to use for dimensionality reduction. There are a variety of other options available to customize the embedding. Mousing over the fields will bring up a tool tip which describes the field's function.

Landmark Layout options in EmbedSOM Pick from different landmark generating functions to visualize the data in different ways. If you have previously used the FlowSOM plugin to calculate clustering results, you can use the SOM from FlowSOM to embed the same data or even map a new dataset to the SOM. Select the **FlowSOM Map** option then use the **Select FlowSOM map file** button to load in the FlowSOM .RData file that contains the SOM landmarks. This option will allow you to align multiple population positions across different files, as long as they have the exact same parameters. For example, we can concatenate multiple files together in FlowJo and run FlowSOM clustering, then use the FlowSOM .RData file to embed the same data into two dimensions. The same FlowSOM landmarks can be used again to embed new datasets to the same FlowSOM object and align these populations.
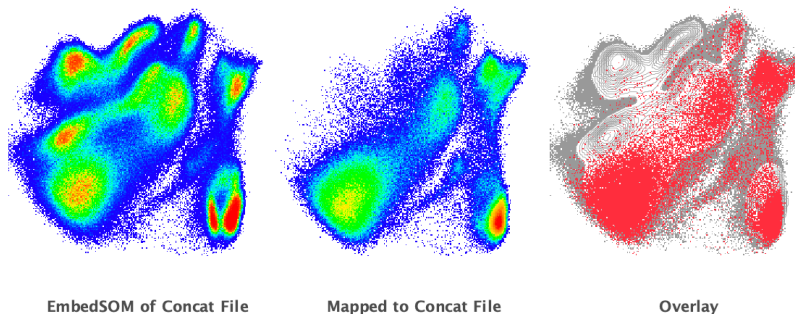

Figure 1 - Embedding on concatenated file (left-most image) consisting of different stim conditions is generated with FlowSOM .RData file and SOM landmarks. The middle image shows a new dataset mapped to the same FlowSOM landmarks and the results are overlaid in the right-most image.

Selecting between the **UMAP** and **tSNE** option will allow you to use these advanced dimensionality reduction methods to quickly generate the landmark positions using a subset of random events. The number of random

input points to use is entered in the **Map Size** input field in the plugin UI. **K** is the number of nearest points used for the approximation, the default is set to 25 but can be raised up to 100 when embedding larger datasets. When using UMAP to generate landmarks the **UMAP Dist** is used to specify the minimum distance between points in low-dimensional space. Smaller values will push points together, while larger values will push points further apart in the final embedding.

For any of the Landmark options selected, increase the **Smooth** parameter to produce "smoother" but possibly more convoluted embedding. Default value 0 is adjusted to be good for most datasets; lower values (-1, -2,. . .) produce sharper embeddings with less focus on projection. Higher values (1,2,. . .) produce smoother embeddings that better explains the large-scale structure of data, but some small-scale details may get smoothed out. The **Adjust** parameter is a negative power factor for reducing the effect of non-local relevance measure on the outcome. Use 0 for plain projection; values above 100 usually push cells closer to respective SOM vertex positions.

Selecting the **EmbedSOM** option will use this improved embedding enrichment method to produce a smooth low-dimensional embedding using the grid size specified in the **SOM Grid Size** input fields. The **Seed** value can be used to create reproducible results by using the same seeed value for the selected dataset.

Below we see embeddings created with each of the available methods in this plugin. The left-most plot was created by EmbedSOM SOM manifolds and shows a more smoothed-out embedding but still separates out the major lineages as shown by the colored islands. The next plot to the right was created using the FlowSOM .RData file and used the SOM landmarks created from the FlowSOM plugin and as map and embedded the new data. This embedding separates nicely and with the .RData file, can be applied to other sample files from the same experiment with the same parameter names. The UMAP result shows tighter grouping of islands, depicting more of the global structure of the data. The embedding on the right which used tSNE landmarks shows more local structure in the data but still separates the islands well thanks to the tunable parameters in this plugin.
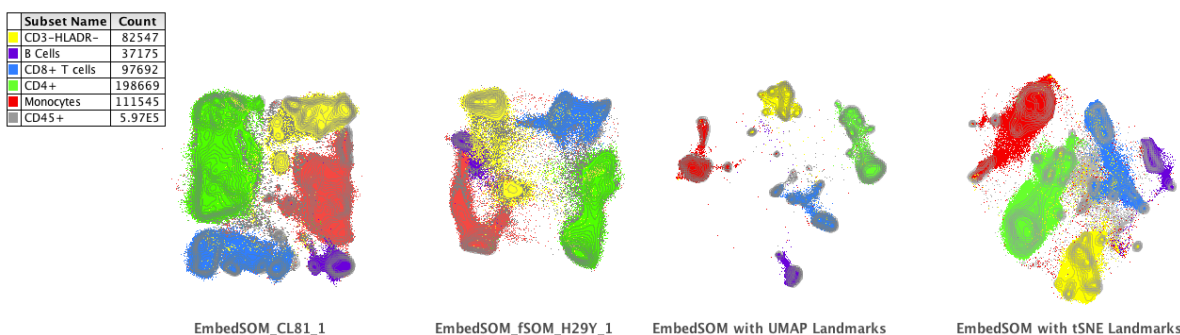

Subset Name	Count
CD3−HLADR−	82547
B Cells	37175
CD8+ T cells	97692
CD4+	198669
Monocytes	111545
CD45+	5.97E5

EmbedSOM_CL81_1   EmbedSOM_fSOM_H29Y_1   EmbedSOM with UMAP Landmarks   EmbedSOM with tSNE Landmarks

Figure 2 - Comparisons of the different embedding options on a CD45+ population is overlaid with major lineage populations. Each option is able to quickly produce embeddings to help visualize the data.

Leave us your feedback

Please write to flowjo@bd.com with any questions or concerns.