# spotMAX saved numerical features

For each analysis run spotMAX saves 4 .csv files and 4 .h5 files. Note that if you choose to not perform the gaussian fit then the files are 3 instead of 4. Each .csv file is named with the pattern “r\_n\_test\_Summary\_v1.csv” while each .h5 files is named with the pattern “r\_n\_test\_data.h5”. In the filename, “r” is the run number, “n” is a number indicating at which stage of the analysis the file was saved (see below), and “test” indicates filtering test was performed on the saved data.

The “…Summary\_v1.csv” file contains spots data aggregated for each segmented object, while the .h5 files contains data for each single spot.

As an example, the file “2\_4\_spotfit\_data\_Summary\_v1.csv” is the aggregated data for run number 2 after the 4th filtering step. This filtering step is called “spotfit” because it is after the filtering by size determined with a 3D gaussian fit.

The 5 filtering steps are:

1. “0\_Orig” 🡪 these are ALL the spots detected before any filtering
2. “1\_ellip\_test” 🡪 these are the spots after removing spots that are too close to each other. Two spots are considered too close if the both lie within spheroid whose dimensions are determined from the diffraction limit, the z-resolution, the z-multiplier and the yx-multiplier.
3. “2\_p-\_test\_data” 🡪 these are the spots from step 1. filtered by either the t-test or the effect size, depending on the user choice
4. “3\_p-\_ellip” 🡪 these are the spots filtered by both step 2. And step 3. Typically, these is the final count.
5. “4\_spotFIT” 🡪 these are the spots from step 4. Filtered by minimum and maximum size. The size range is set by the user.

## Column meaning in the Summary.csv files

|  |  |
| --- | --- |
| Cell\_ID | The ID of the segmented objects. The segmented objects are those labels saved in the file ending with “segm.npy” |
| frame\_i | Frame index. Relevant only for time-lapse data.  *NOTE: Timelapse data is not supported yet.* |
| cell\_area\_pxl | Area of the segmented object in pixel. This is 0 for 3D segmentation data. |
| cell\_area\_um2 | Area of the segmented object in µm2. This is 0 for 3D segmentation data. |
| cell\_vol\_vox | Volume of the segmented object in voxels. For 3D segmentation data this is the sum of all the voxels in the object. For 2D segmentation data, the volume is estimated from the revolution of the 2D section. |
| cell\_vol\_vox | Volume of the segmented object in femtoliters. The conversion is through the pixel size provided by the user. |
| num\_spots | The total number of spots per segmented object |
| ref\_ch\_vol\_vox | The volume of the reference channel in voxels. |
| ref\_ch\_vol\_um3 | The volume of the reference channel in µm3. |
| ref\_ch\_num\_fragments | The number of separated objects as determined from segmentation of the reference channel. |
| spots\_ch\_norm\_val | The normalization value used to normalise the spots signal prior comparison with the reference channel signal. It is the median of the signal outside of the spots, but inside the segmented object. |
| ref\_ch\_norm\_val | Same as above, but for the reference channel. |
| spotsize\_tot\_fluoresc | The total fluorescence intensity (i.e., sum) inside the spots. The spots footprint is determined by the spotSIZE step. |
| spots\_INref\_tot\_fluoresc | The total fluorescence intensity (i.e., sum) of the spots signal inside the reference channel mask. |
| spotfit\_sum\_foregr\_integral | The sum of each spot’s “I\_foregr”. See data.h5 columns below. If spotFIT was successful, this is a good proxy for the total amount of “molecules” in each segmented object. |
| spotfit\_sum\_tot\_integral | The sum of the spot’s “I\_tot”. See data.h5 columns below. |
| mean\_sigma\_z | The mean of the spot’s “sigma\_z\_fit”. See data.h5 columns below. |
| mean\_sigma\_y | The mean of the spot’s “sigma\_y\_fit”. See data.h5 columns below. |
| mean\_sigma\_x | The mean of the spot’s “sigma\_x\_fit”. See data.h5 columns below. |
| std\_sigma\_z | The standard deviation of the spot’s “sigma\_z\_fit”. See data.h5 columns below. |
| std\_sigma\_y | The standard deviation of the spot’s “sigma\_y\_fit”. See data.h5 columns below. |
| std\_sigma\_x | The standard deviation of the spot’s “sigma\_x\_fit”. See data.h5 columns below. |
| sum\_A\_fit | The sum of the spot’s “A\_fit”. See data.h5 columns below. |
| mean\_B\_fit | The mean of the spot’s “B\_fit”. See data.h5 columns below. |
| solution\_found | The mean of the spot’s “solution\_found”. See data.h5 columns below. |
| mean\_reduced\_chisq | The mean of the spot’s “reduced\_chisq”. See data.h5 columns below. |
| mean\_NRMSE | The mean of the spot’s “NRMSE”. See data.h5 columns below. |
| mean\_F\_NRMSE | The mean of the spot’s “F\_NRMSE”. See data.h5 columns below. |
| mean\_ks | The mean of the spot’s “KS\_stat”. See data.h5 columns below. |
| combined\_p\_ks | The combined “p\_KS” (Fisher’s method) of the spots. See data.h5 columns below. |
| mean\_ks\_null | The mean of the spot’s “null\_ks\_test”. See data.h5 columns below. |
| mean\_chisq\_null | The mean of the spot’s “null\_chisq\_test”. See data.h5 columns below. |
| mean\_QC\_passed | The mean of the spot’s “QC\_passed”. See data.h5 columns below. |

## Column meaning in the data.h5 files

|  |  |
| --- | --- |
| vox\_spot | The raw intensity at the spot’s (z, y, x) centre coordinates. |
| vox\_ref | The raw intensity of the reference channel at the spot’s (z, y, x) centre coordinates. |
| |abs|\_spot | The mean of the spot’s signal from the minimum spot volume determined by the user. |
| |abs|\_ref | The mean of the reference channel signal from the minimum spot volume determined by the user (same voxels’ coordinates as above). |
| |norm|\_spot | The mean of the normalised signal from the same voxels as in |abs|\_spot column. |
| |norm|\_ref | The mean of the normalised signal from the same voxels as in |abs|\_ref column. |
| |spot|:|ref| t-value | The t-statistic from the t-test between spot and reference channel |
| |spot|:|ref| p-value (t) | The p-value of the t-test between spot and reference channel |
| z | The z-coordinate of the spot’s centre |
| y | The y-coordinate of the spot’s centre |
| x | The x-coordinate of the spot’s centre |
| peak\_to\_background ratio | The ratio between |abs|\_spot and the median of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| effsize\_cohen\_s | The Cohen’s effect size between the spot and the background or the reference channel in the same spot’s coordinates. |
| effsize\_hedge\_s | The Hedge’s effect size between the spot and the background or the reference channel in the same spot’s coordinates. |
| effsize\_glass\_s | The Glass’ effect size between the spot and the background or the reference channel in the same spot’s coordinates. |
| effsize\_cliffs\_s | The Cliff’s Delta effect size between the spot and the background or the reference channel in the same spot’s coordinates. |
| effsize\_cohen\_pop | The Cohen’s effect size between the spot and the background or the entire reference channel (population) |
| effsize\_hedge\_pop | The Hedge’s effect size between the spot and the background or the entire reference channel (population) |
| effsize\_glass\_pop | The Glass’ effect size between the spot and the background or the entire reference channel (population) |
| backgr\_INcell\_OUTspot\_mean | The mean of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| backgr\_INcell\_OUTspot\_median | The median of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| backgr\_INcell\_OUTspot\_75p | The 0.75 quantile of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| backgr\_INcell\_OUTspot\_25p | The 0.25 quantile of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| backgr\_INcell\_OUTspot\_std | The standard deviation of the background signal. The background is determined as the region outside of the spots, but inside the segmented object. |
| is\_spot\_inside\_ref\_ch | 1 or 0 depending on whether the spot is inside or outside of the reference channel segmentation mask. |
| spotsize\_yx\_radius\_um | The radius (in µm) of the spot in y or x direction as determined by spotSIZE. |
| spotsize\_z\_radius\_um | The radius (in µm) of the spot in z direction as determined by spotSIZE. |
| spotsize\_yx\_radius\_pxl | The radius (in pixel) of the spot in y or x direction as determined by spotSIZE. |
| spotsize\_z\_radius\_pxl | The radius (in pixel) of the spot in z direction as determined by spotSIZE. |
| spotsize\_limit | Background mean + 3 \* background std. This is used to determine when to stop the spotSIZE process. |
| spot\_surf\_50p | The median of the spot’s outer surface intensities |
| spot\_surf\_5p | The 0.05 quantile of the spot’s outer surface intensities |
| spot\_surf\_mean | The mean of the spot’s outer surface intensities |
| spot\_surf\_std | The standard deviation of the spot’s outer surface intensities |
| spot\_B\_min | The value used as initial guess for the parameter “B” in the 3D gaussian equation. |
| QC\_passed | 1 or 0 depending on whether the NRMSE of the spot is below a maximum limit or not. The limit is determined as the 0.75 quantile + 1.5 \* interquartile range (IQR) of all the spots’ NRMSE. |
| null\_ks\_test | 1 or 0 depending on whether the Kolmogorov–Smirnov ‘s test between data and gaussian fit was null or not. |
| null\_chisq\_test | 1 or 0 depending on whether the Chi-square test between data and gaussian fit was null or not. |
| solution\_found | 1 or 0 depending on whether the non-linear regression algorithm in the scipy library determined if the a solution was found or not. |
| z\_fit | The z-coordinate of the 3D gaussian fit. |
| y\_fit | The y-coordinate of the 3D gaussian fit. |
| x\_fit | The x-coordinate of the 3D gaussian fit. |
| sigma\_z\_fit | The z-sigma of the 3D gaussian fit. |
| sigma\_y\_fit | The y-sigma of the 3D gaussian fit. |
| sigma\_x\_fit | The x-sigma of the 3D gaussian fit |
| sigma\_yx\_mean | The mean between sigma\_y\_fit and sigma\_x\_fit. |
| spotfit\_vol\_vox | The volume of the ellipsoid with dimensions sigma\_z\_fit, sigma\_y\_fit and sigma\_x\_fit. |
| A\_fit | The A parameter of the 3D gaussian fit. Correlated to the amplitude of the peak. |
| B\_fit | The B parameter of the 3D gaussian fit. Correlated to the background lever of the peak. |
| I\_tot | The total integral of the 3D gaussian fit. |
| I\_foregr | The foreground integral of the 3D gaussian fit. |
| reduced\_chisq | The reduced Chi-square of the Chi-square test between data and gaussian fit. |
| p\_chisq | The p-value of the Chi-square test between data and gaussian fit. |
| KS\_stat | The statistic of the Kolmogorov–Smirnov ‘s test between data and gaussian fit. |
| p\_KS | The p-value of the Kolmogorov–Smirnov ‘s test between data and gaussian fit. |
| RMSE | The root mean squared error between data and gaussian fit. |
| NRMSE | The normalized room mean squared error between data and gaussian fit. Normalization is with the mean of the data. |
| F\_NRMSE | The normalized room mean squared error between data and gaussian fit, rescaled between 0 and 1. |