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 μm^2. 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 μm^3.
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.