Systematic assessment of burst impurity in confocal-based single-molecule fluorescence detection using Brownian motion simulations

Dolev Hagai [1] and Eitan Lerner [1,*]

[1] Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, Faculty of Mathematics & Science, The Edmond J. Safra Campus, The Hebrew University of Jerusalem, Jerusalem, Israel; Eitan.Lerner@mail.huji.ac.il

* Correspondence: Eitan.Lerner@mail.huji.ac.il; Tel.: +972-2-658-5457

1.1. Reasons for bias in parameters retrieved from fluorescence autocorrelation analysis of photons in SMFD

It is worth mentioning that fluorescence autocorrelation functions at the single molecule regime (at concentrations <100 pM), where most of the time no molecule crosses the effective excitation volume, might be inaccurate. Fluorescence correlation spectroscopy (FCS) assumes the signal has a well-defined mean, and that the information is found in the temporal fluctuations about that mean. However, at the single-molecule level, there is not one mean signal, but two, one of the BG process, and the other of the fluorescence process, that occurs only once in a while. As a result, Poisson statistics do not characterize it well, but rather a combination of two Poisson processes.

When focusing on photons of bursts, however, the signal should be analyzable by FCS approaches. That might be true as long as the shape of the part of the PSF from which photons are emitted is known and resembles the Gaussian approximation. The model used for analysis of fluorescence autocorrelation functions stems from a perfectly-shaped Gaussian PSF. We, however, have already shown that after burst analysis, the molecular positions form a shape that deviates from the Gaussian shape, even when performing the simulations using a Gaussian PSF model. Therefore, the estimation of the mean amount of molecules in the effective detection volume (EDV) at any given moment, <N>, from such fits to autocorrelation curves of burst photon timestamps, may introduce biased results. Additionally, using Poisson statistics to infer the probability of more than a single molecule, P(N>1), may carry with it an additional bias, because the statistics is not pure single Poisson anymore. The statistics may get closer to Poisson statistics when using high photon rate thresholds, relative to the BG rate. Finally, the process of retrieving P(N>1) involves so many steps that propagate errors, while the information is known from the ground truth of the simulation.

34 <u>Figure Legends</u>

35

36 **Figure S1.** The positions of diffusing molecules (in simulations using the gaussian PSF model) when they emitted
37 photons that were detected and selected by the burst analysis, either with minimal burst analysis parameter
38 values (m=5 & F=6; left panels) or with stringent burst analysis parameter values (m=10, F=6 & burst size
39 threshold, sz=40; right panels). In the top, central & bottom panels we show the 2D projections at the yz, xz & xy
40 planes, respectively. Each dot in the scatter plots is an emitted photon. These results are for the simulation of
41 molecules in a concentration of 62 pM, where the diffusion coefficient of the molecules was 90 μm2/s. The colors
42 of the points correspond to the burst number out of the overall number of bursts. In each panel, the 1D
43 projections are also shown as histograms. The black, brown and yellow contour lines align the position of the
44 gaussian PSF model (see shapes of PSF models in Figure S1).

45

46 **Figure S2.** histograms of the 1D projections shown in Figures 1 & 2 for the z , y & x coordinates (left, center &
47 right panels, respectively). From top to bottom, we assessed these histograms as a function of a sliding window
48 of m consecutive photons (m=5, 20, 15 & 20 in blue, orange, green & red, respectively), using a constant
49 instantaneous photon rate threshold of F=6; as a function of the instantaneous photon rate threshold F, (F=3, 6
50 ,11 ,16 & 21 in blue, orange, green, red & magenta, respectively), using a sliding window of constant m=10
51 consecutive photons; as a function of the minimal burst size threshold (10, 20, 40 & 80 in blue, orange, green &
52 red, respectively); and as a function of the minimal burst width threshold (0.0, 0.5 & 1.0 ms in blue, orange &
53 green, respectivey), for a constant m=10 & F=6. These results are for the simulation of molecules in a concentration
54 of 62 pM, where the diffusion coefficient of the molecules was 90 μm²/s. The colors of the points correspond to
55 the burst number out of the overall number of bursts.

56

57 **Figure S3.** The molecular position dispersion as a function of burst search criteria and experimental conditions.
58 Shown are the standard deviation of molecular positions in the z (left) & x (right) coordinates (the values in the
59 y coordinate are the same as the ones in the x coordinate, in within the error ranges), when they emitted photons
60 that were detected and selected by the burst analysis. The error values were calculated as the uncertainty of the
61 standard deviation. All values are reported in Table S1. The assessment of the molecular position dispersion
62 here is shown as a function of different concentrations for molecules diffusing with a constant diffusion
63 coefficient of 90 μm²/s, and in simulations using the gaussian PSF model.

64

65 **Figure S4.** The molecular position dispersion as a function of burst search criteria and experimental conditions.
66 Shown are the standard deviation of molecular positions in the z (left) & x (right) coordinates (the values in the
67 y coordinate are the same as the ones in the x coordinate, in within the error ranges), when they emitted photons
68 that were detected and selected by the burst analysis. The error values were calculated as the uncertainty of the
69 standard deviation. All values are reported in Table S1. The assessment of the molecular position dispersion
70 here is shown as a function of molecules diffusing with different diffusion coefficients, at a constant
71 concentration (62 pM), and in simulations using the gaussian PSF model.

72

73 **Figure S5.** Quantification of the level of impurity of single-molecule bursts using different burst analysis
74 parameter values. Each panel shows a histogram of all the bursts' level of impurity, calculated as the fraction of
75 photons arising from molecules other than the main one. The burst impurity histograms are of the simulation
76 results after burst search analysis using a constant photon rate threshold F=6 and varying value of m, for a sliding

77 window of m consecutive photons (left), constant m=10 and varying F values (center) and constant m=10, F=6
78 and varying burst size threshold values (right). The continuous and dashed vertical grey lines indicate the mean
79 and error range (as calculated using the mean and the standard error) impurity value for all bursts. These results
80 are for the simulation of molecules in a concentration of 62 pM, where the diffusion coefficient of the molecules
81 was 90 μm²/s.
82
83 **Figure S6**. The occurrence and level of impure bursts as a function of burst search criteria and concentrations –
84 numerical PSF model. Different burst analysis parameter values for different concentrations of molecules. The
85 relative occurrence of impure bursts (left) was calculated as the fraction of bursts with an impurity level larger
86 than 0 (error ranges calculated as the 95% confidence intervals), as the fraction of non-single-molecule bursts,
87 and hence as the fraction of impure bursts. The level of impurity (right) was calculated as either the mean of all
88 burst impurity levels (black; error ranges calculated as the standard error) or as the fraction of impure photons
89 from all bursts relative to all burst photons (red; no error ranges, as the calculation was performed over all
90 photons). The assessment is shown as a function of different diffusion coefficients at a constant concentration of
91 62 pM, and in simulations using the numerical PSF model.
92
93 **Figure S7**. The occurrence and level of impure bursts as a function of burst search criteria and concentrations –
94 gaussian PSF model. Different burst analysis parameter values for different concentrations of molecules. The
95 relative occurrence of impure bursts (left) was calculated as the fraction of bursts with an impurity level larger
96 than 0 (error ranges calculated as the 95% confidence intervals), as the fraction of non-single-molecule bursts,
97 and hence as the fraction of impure bursts. The level of impurity (right) was calculated as either the mean of all
98 burst impurity levels (black; error ranges calculated as the standard error) or as the fraction of impure photons
99 from all bursts relative to all burst photons (red; no error ranges, as the calculation was performed over all
100 photons). The assessment is shown as a function of different concentrations for molecules diffusing with a
101 constant diffusion coefficient of 90 μm2/s, and in simulations using the gaussian PSF model.
102 **Figure S8.** The occurrence and level of impure bursts as a function of burst search criteria and concentrations –
103 gaussian PSF model. Different burst analysis parameter values for different concentrations of molecules. The
104 relative occurrence of impure bursts (left) was calculated as the fraction of bursts with an impurity level larger
105 than 0 (error ranges calculated as the 95% confidence intervals), as the fraction of non-single-molecule bursts,
106 and hence as the fraction of impure bursts. The level of impurity (right) was calculated as either the mean of all
107 burst impurity levels (black; error ranges calculated as the standard error) or as the fraction of impure photons
108 from all bursts relative to all burst photons (red; no error ranges, as the calculation was performed over all
109 photons). The assessment is shown as a function of different diffusion coefficients at a constant concentration of
110 62 pM, and in simulations using the gaussian PSF model.
111
112 **Figure S9.** The molecular positions of pure & impure bursts photons, as a function of varying burst search
113 parameter, m (numerical PSF model) – shape and amplitude. We the histograms of molecular positions in the z
114 coordinate of impure photons (red), burst photons of impure bursts (yellow), of pure bursts (green) and of all
115 bursts (black), both un-normalized (left) to assess the weight of burst impurity, and normalized (right) to assess
116 the histogram shapes. These results refer to the simulations in concentration of 62 pM and diffusion coefficient
117 of 90 μm2/s, using the numerical PSF model.
118

119   **Figure S10.** The molecular positions of pure & impure bursts photons, as a function of varying burst size
120   threshold *sz* (numerical PSF model) – shape and amplitude. We the histograms of molecular positions in the z
121   coordinate of impure photons (red), burst photons of impure bursts (yellow), of pure bursts (green) and of all
122   bursts (black), both un-normalized (left) to assess the weight of burst impurity, and normalized (right) to assess
123   the histogram shapes. These results refer to the simulations in concentration of 62 pM and diffusion coefficient
124   of 90 μm2/s, using the numerical PSF model.

125

126   **Figure S11.** The molecular positions of pure & impure bursts photons, as a function of varying the instantaneous
127   photon rate threshold *F* (Gaussian PSF model) – shape and amplitude. We the histograms of molecular positions
128   in the z coordinate of impure photons (red), burst photons of impure bursts (yellow), of pure bursts (green) and
129   of all bursts (black), both un-normalized (left) to assess the weight of burst impurity, and normalized (right) to
130   assess the histogram shapes. These results refer to the simulations in concentration of 62 pM and diffusion
131   coefficient of 90 μm2/s, using the Gaussian PSF model.

132

133   **Figure S12.** The molecular positions of pure & impure bursts photons, as a function of varying burst search
134   parameter, *m* (Gaussian PSF model) – shape and amplitude. We the histograms of molecular positions in the z
135   coordinate of impure photons (red), burst photons of impure bursts (yellow), of pure bursts (green) and of all
136   bursts (black), both un-normalized (left) to assess the weight of burst impurity, and normalized (right) to assess
137   the histogram shapes. These results refer to the simulations in concentration of 62 pM and diffusion coefficient
138   of 90 μm2/s, using the Gaussian PSF model.

139

140   **Figure S13.** The molecular positions of pure & impure bursts photons, as a function of varying burst size
141   threshold *sz* (Gaussian PSF model) – shape and amplitude. We the histograms of molecular positions in the z
142   coordinate of impure photons (red), burst photons of impure bursts (yellow), of pure bursts (green) and of all
143   bursts (black), both un-normalized (left) to assess the weight of burst impurity, and normalized (right) to assess
144   the histogram shapes. These results refer to the simulations in concentration of 62 pM and diffusion coefficient
145   of 90 μm2/s, using the Gaussian PSF model.

146

147   **Figure S14.** Photon timestamp autocorrelations and their best fit results to a model of fluorescence correlation
148   of freely diffusing molecules. From top to bottom: autocorrelation of all photons, of burst photons with varying
149   the burst search parameter *m*, keeping the burst search parameter *F*=6, of burst photons with varying the burst
150   search parameter *F*, keeping the burst search parameter *m*=10, with varying burst selection parameter, burst size
151   threshold, keeping burst search parameters *m*=10 & *F*=6, and with varying burst selection parameter, burst width
152   threshold, keeping burst search parameters *m*=10 & *F*=6. One can observe the trend in the mean amount of
153   molecules in the effective detection volume at any given moment, *<N>*, as a function of burst analysis parameter
154   values, as the inverse of the change of the autocorrelation extrapolated to 0 lag time, τ=0.

155

156   **Figure S15.** The mean amount of molecules in the effective detection volume at any given moment is the same
157   in simulations of the same concentrations. The different panels show the values of *<N>*, the mean amount of
158   molecules in the effective detection volume at any given moment, retrieved as best fit values from fitting the
159   burst timestamp autocorrelation functions to a model of fluorescence autocorrelation of molecules freely
160   diffusing in 3D, in a confocal-based setup.

161

162 **Figure S16.** The correlation of the probability of more than a single molecule in the effective excitation volume
163 with molecular position dispersion. Different burst analysis parameter values for different concentrations of
164 molecules. The probability of more than a single molecule in the effective excitation volume, *P(N>1)*, (error
165 ranges were propagated from the values of the fitting error to the *<N>* parameter, after fitting the photon
166 timestamp autocorrelation functions to a model of fluorescence autocorrelation, as in figure S8) were compared
167 against the molecular position dispersion in the z coordinate (error ranges calculated as the uncertainty of the
168 standard deviation), as a function of different burst analysis parameter values (from left to right: varying *m*
169 values, varying *F* values, varying burst size threshold values & varying burst width threshold values), for
170 different simulation conditions (from top to bottom: different concentrations at a constant diffusion coefficient
171 value of 90 μm^2/s in simulations using the gaussian PSF model, different concentrations at a constant diffusion
172 coefficient value of 90 μm^2/s in simulations using the numerical PSF model, different diffusion coefficients at a
173 constant concentration of 62 pM in simulations using the gaussian PSF model, and , different diffusion
174 coefficients at a constant concentration of 62 pM in simulations using the numerical PSF model).

175

176 **Figure S17.** Two estimates of the diffusion time through the effective detection volume, and their usefulness.
177 Shown are the values of the mean of all burst widths (error calculated as standard error), the mean diffusion
178 time as was retrieved from best fits of fluorescence autocorrelation model of freely diffusing molecules in a
179 confocal-setup, to the photon timestamp autocorrelation functions, as in figure S8 (errors are fitting errors). The
180 figure shows the minimal error ranges of burst widths, the large error values of the diffusion times, and the lack
181 of correlation between the mean values, rendering the diffusion time values useless for accurate time estimates,
182 and the mean burst durations useful time estimates, when the simulation used either gaussian or numerical PSF
183 models (left or right, respectively). Top to bottom: varying the burst search parameter *m* and keeping the burst
184 search parameter *F*=6, varying the burst search parameter *F* and keeping the burst search parameter *m*=10,
185 varying burst size threshold and keeping burst search parameters *m*=10 & *F*=6, and varying burst width
186 threshold and keeping burst search parameters *m*=10 & *F*=6.

187

188 **Figure S18.** Estimating mean burst widths. Shown are the histograms of all burst widths after testing burst
189 analysis results with (from top to bottom) varying values of the burst search parameter *m* and a constant
190 instantaneous photon rate threshold *F*=6, varying values of the instantaneous photon rate threshold F and a
191 constant burst search parameter value of *m*=10, varying values of the burst size threshold *sz* and constant burst
192 search parameter values *m*=10 & *F*=6, and varying values of the burst width threshold *w* and constant burst
193 search parameter values *m*=10 & *F*=6. Continuous vertical lines indicate the mean burst widths and dashed
194 vertical lines indicate the error ranges, calculated from the mean and the standard error of the burst widths.
195 These results refer to the simulations in concentration of 62 pM and diffusion coefficient of 90 μm2/s, using the
196 numerical PSF model.

197

198

199 **Figure S19.** The correlation of the mean burst widths with the molecular position dispersion. The mean burst
200 widths, (error ranges were calculated as standard errors) were compared against the molecular position
201 dispersion in the z coordinate (error ranges calculated as the uncertainty of the standard deviation), as a function
202 of different burst analysis parameter values (from left to right: varying *m* values, varying *F* values, varying burst
203 size threshold values & varying burst width threshold values), the molecular dispersion in z and x coordinates
204 show as a pair of panels for different simulation conditions (from top to bottom: different concentrations at a

205   constant diffusion coefficient value of 90 $\mu m^2/s$ in simulations using the gaussian PSF model, different
206   concentrations at a constant diffusion coefficient value of 90 $\mu m^2/s$ in simulations using the numerical PSF model,
207   different diffusion coefficients at a constant concentration of 62 pM in simulations using the gaussian PSF model,
208   and , different diffusion coefficients at a constant concentration of 62 pM in simulations using the numerical PSF
209   model).
210

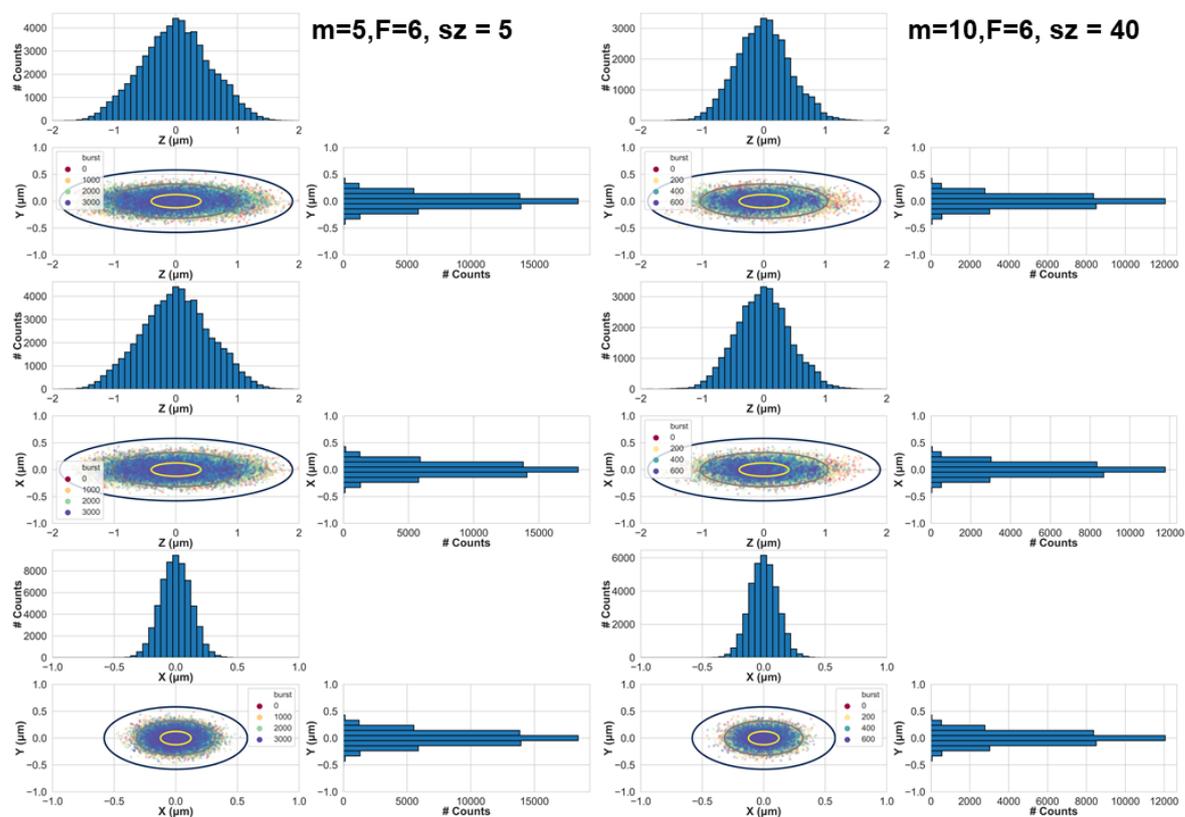
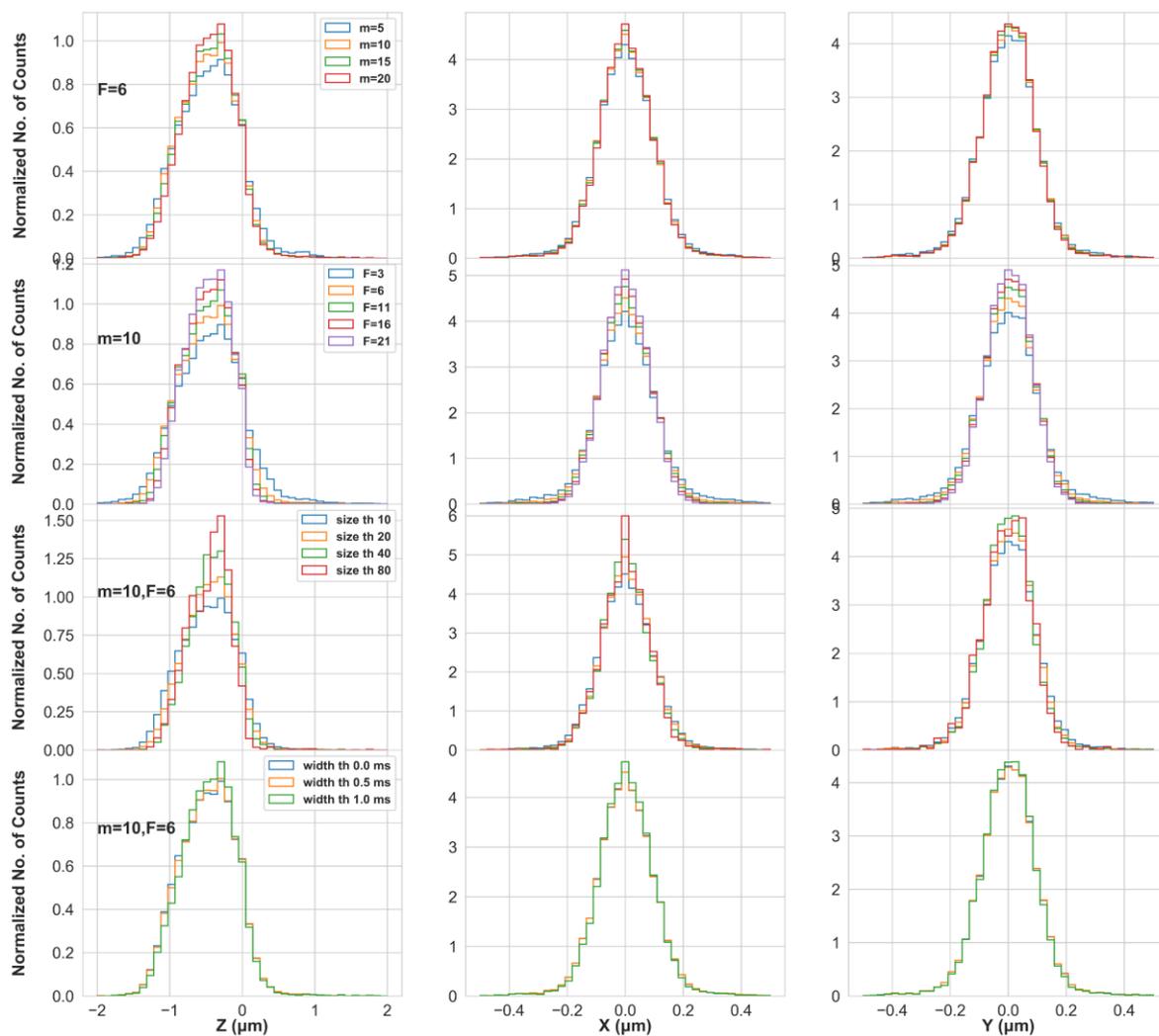
211   **Figure S20.** Simulations of smFRET with two FRET subpopulations, with best fit results to a sum-of-two-
212   gaussians model with a fixed fraction value f=0.6666. From top to bottom, each panel shows the resulting FRET
213   histogram (blue), the best fit sum of two-gaussians, with a fixed population fraction value f=0.6666 (red), the
214   best-fit mean FRET efficiencies (orange and cyan vertical lines; dimmer lines show the error ranges), and the
215   simulation ground-truth mean FRET efficiency values (dashed red and green vertical lines). These results are for
216   the 60 second simulation of molecules a concentration of 62 pM, where the diffusion coefficient of the molecules
217   was 90 $\mu m^2/s$, using the numerical PSF model, and the molecules were split to 10 with E=0.75 & 5 with E=0.5.
218   The number of bursts in each histogram is also reported in each panel. The best fit values and the fitting error
219   values are also reported in Table S2.
220

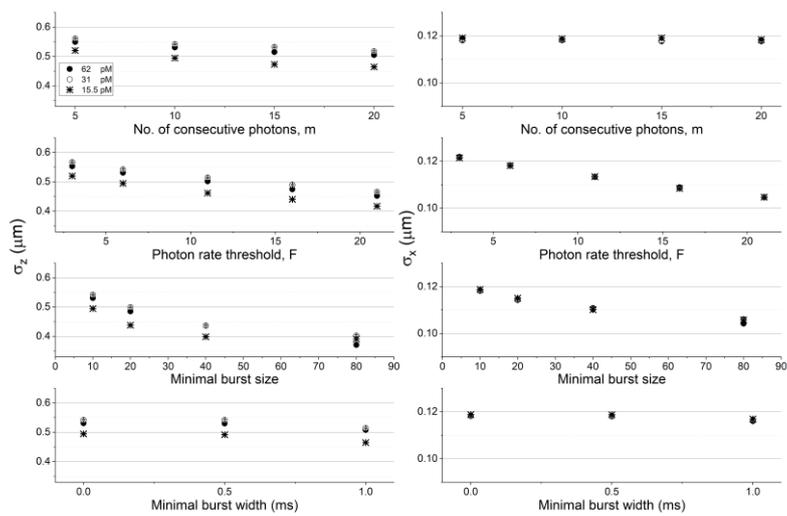
221   **Figure S21.** The values of the retrieved quantities are the same within the error ranges for 60 and 180 second
222   simulations. Simulations lasting either 60 or 180 seconds were performed and then analyzed using the different
223   burst analysis parameter values used over the whole work. The figure shows the values of the retrieved
224   quantities for the 60 seconds (black) and 180 seconds (red) simulations, testing them either for fast or slow
225   diffusion coefficients (90 or 5.625 $\mu m^2/s$, at a constant concentration of 62 pM), and against either gaussian or
226   numerical PSF models.
227
228

229

230
231   **Figure S1.**
232

233

234
235   **Figure S2.**

236

237

238

239 **Figure S3**

240

241

242

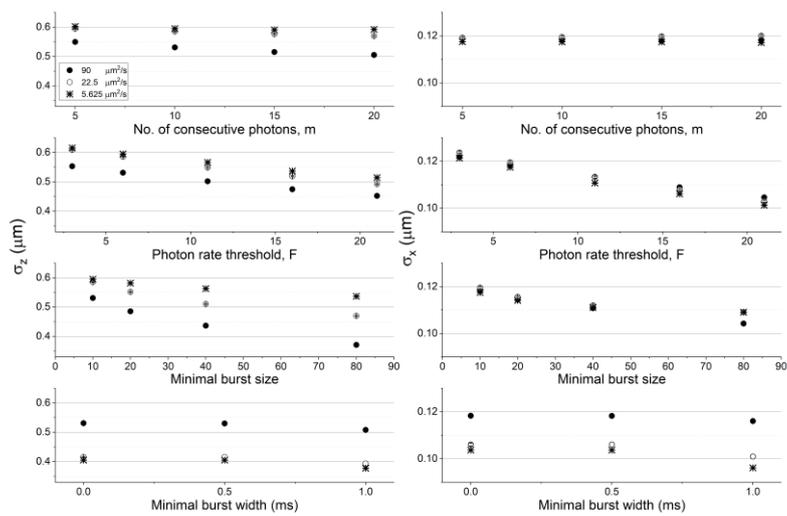



Figure S4

244

245

246

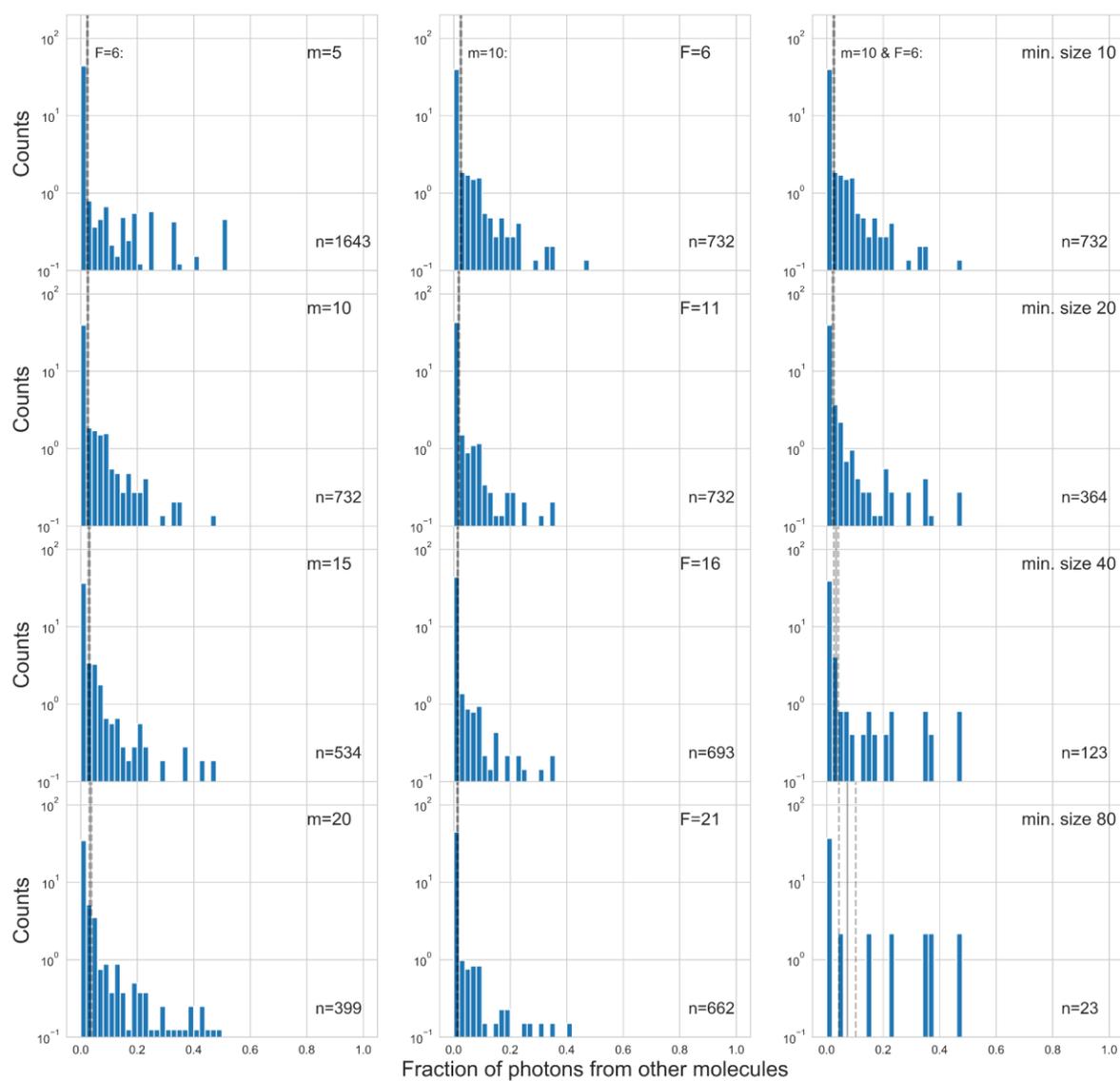
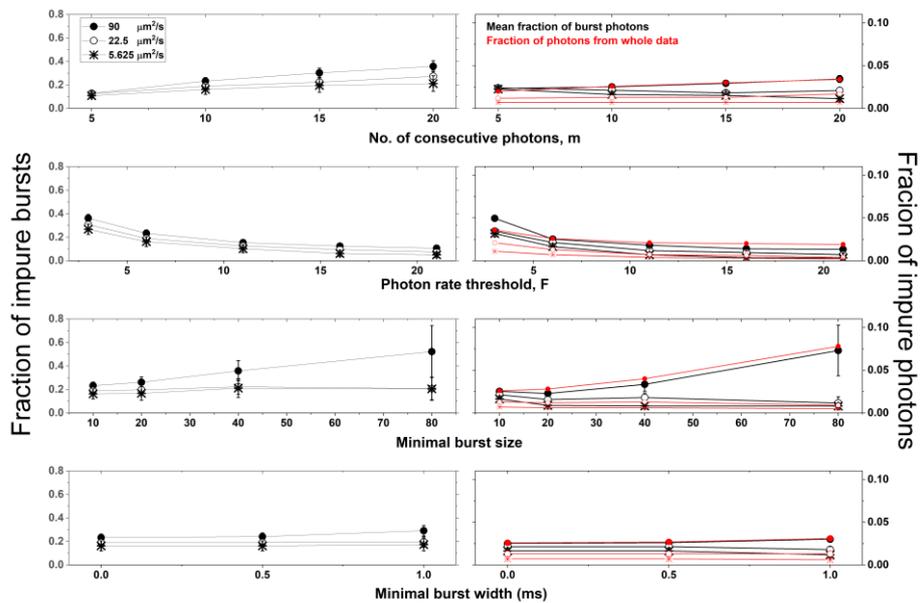



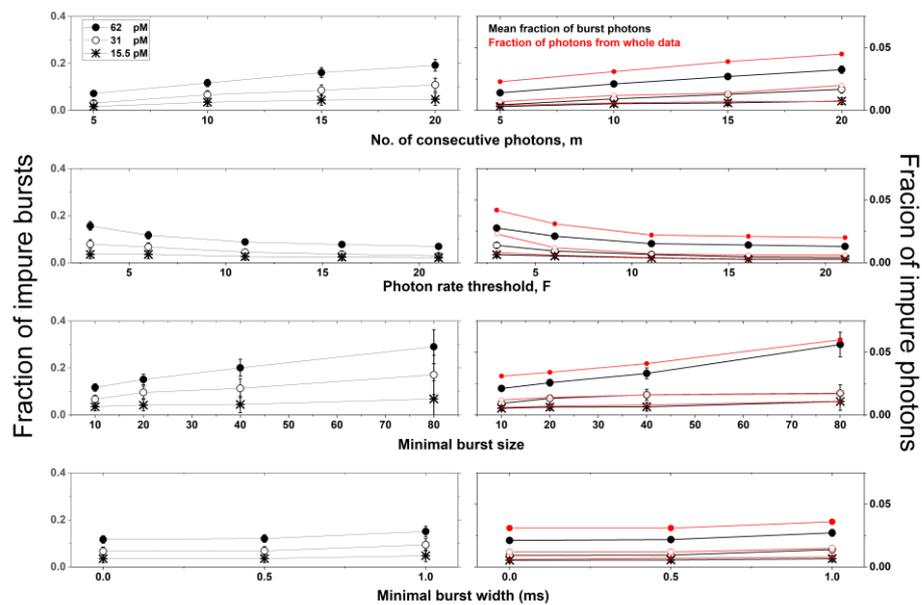
Figure S5

248

249

250

251 **Figure S6**

252

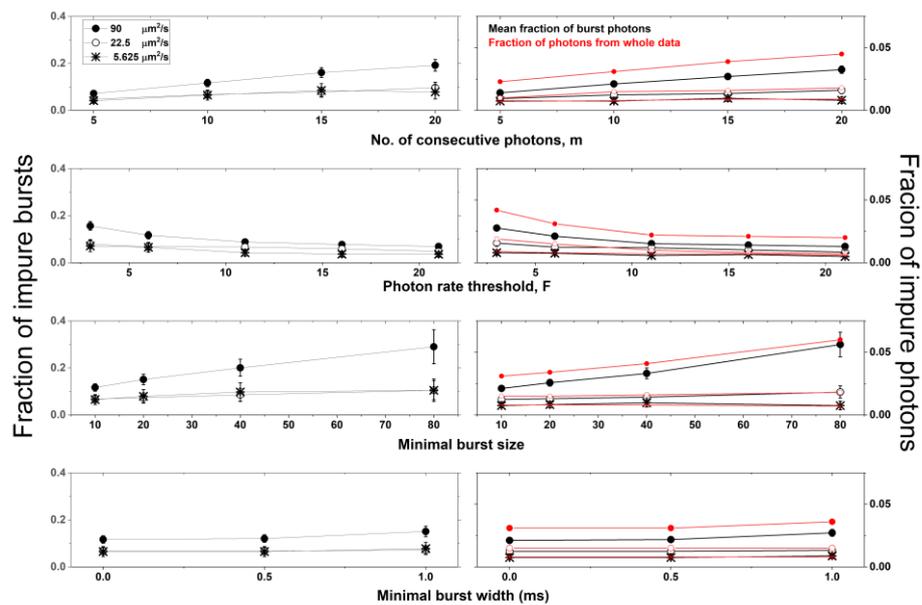
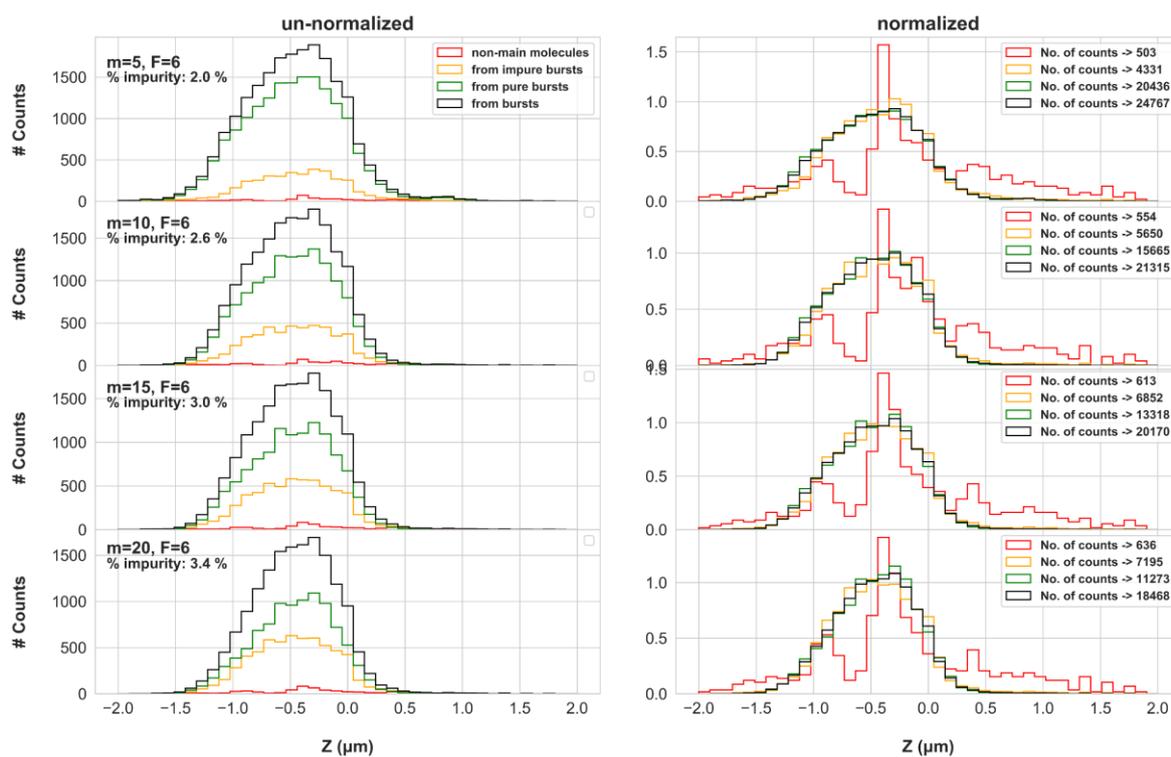
253

254

255 **Figure S7**

256

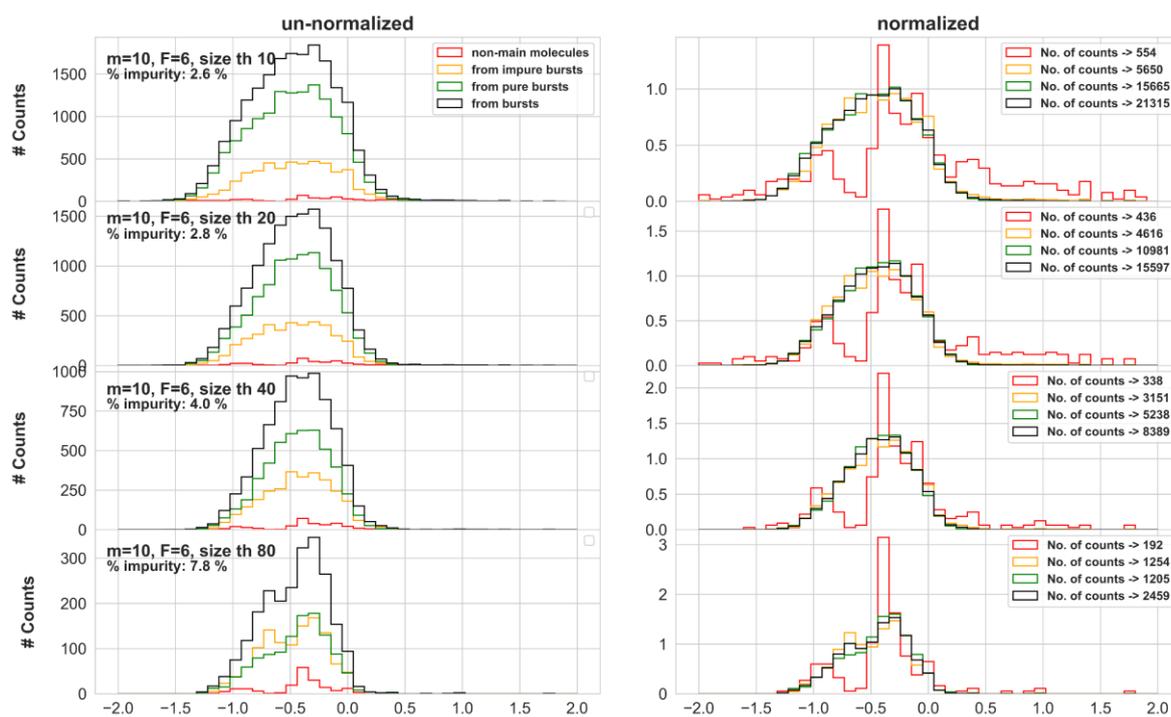
257

258
259   **Figure S8**

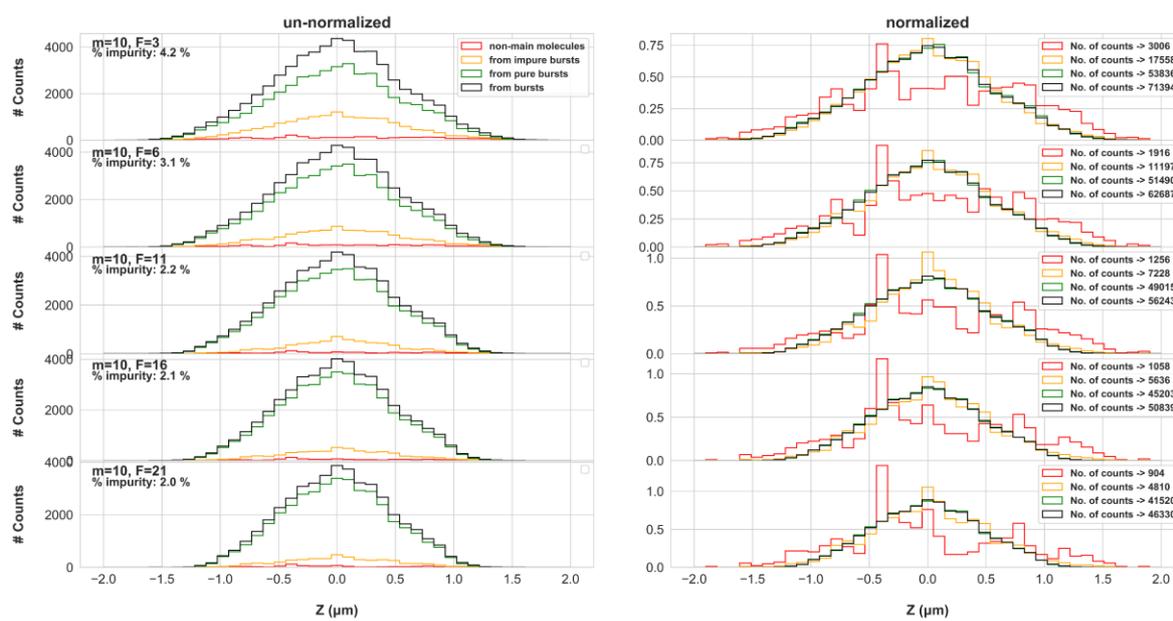
260

261

262

263   **Figure S9**

264

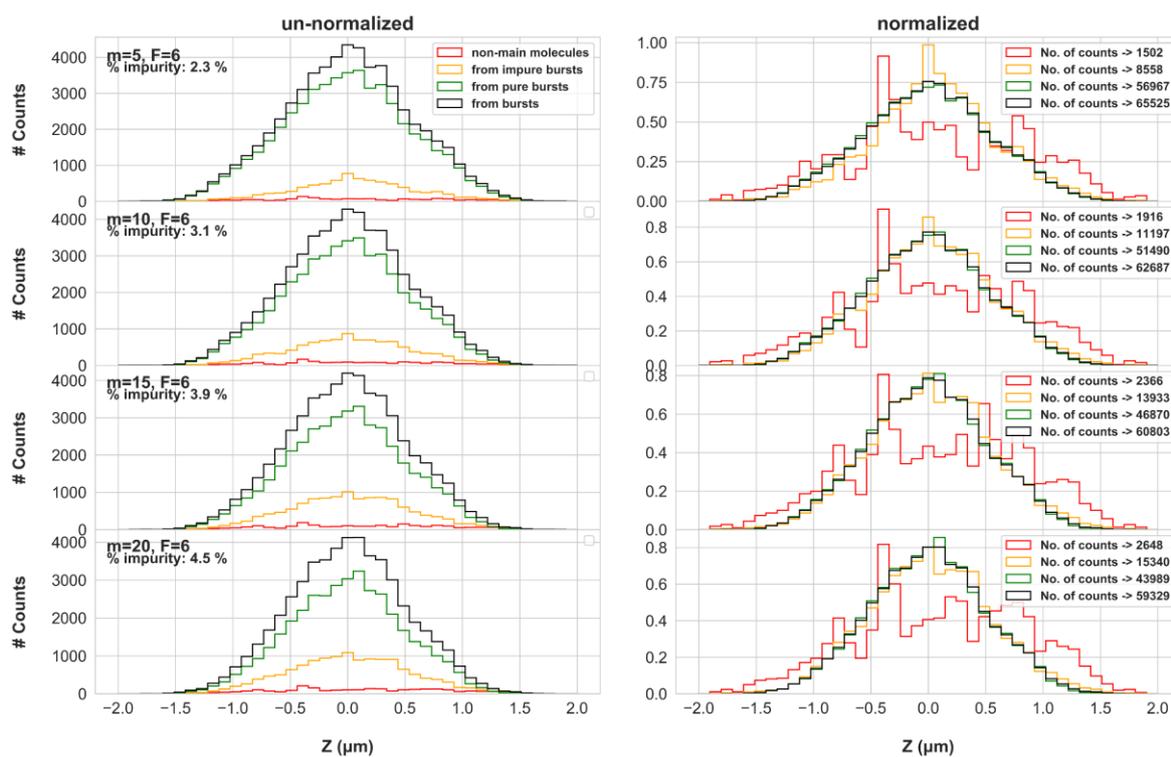
265

266
267 **Figure S10**

268

269

270

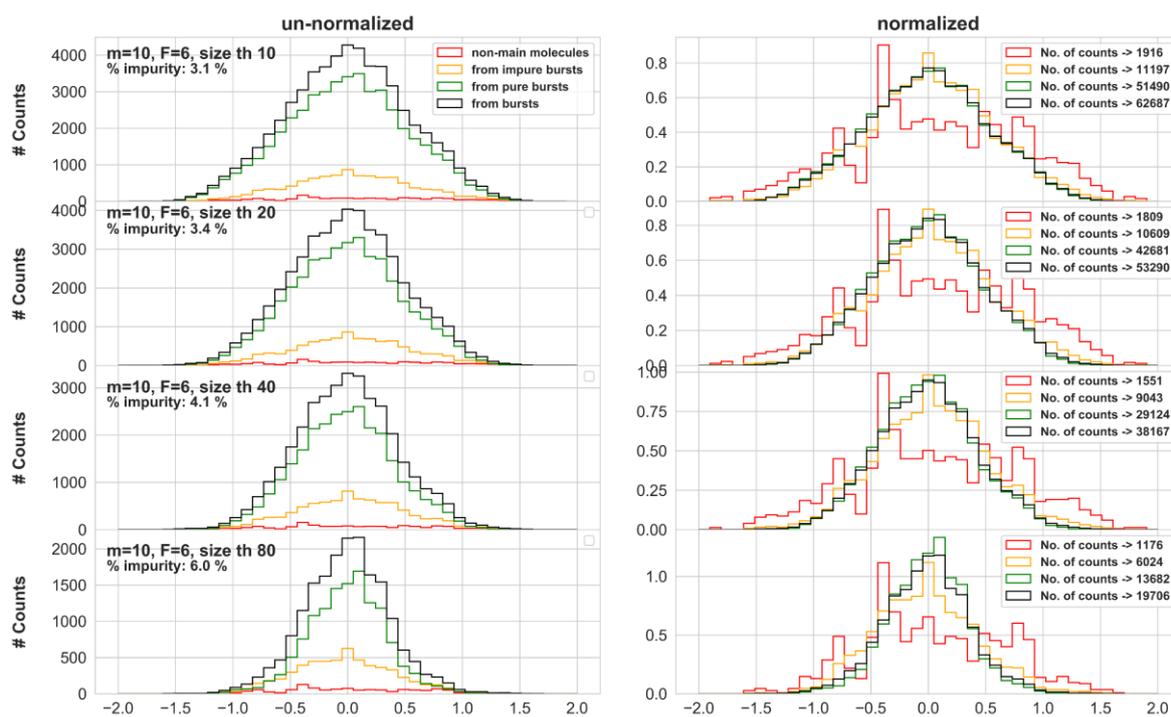
271 **Figure S11**

272

273

274

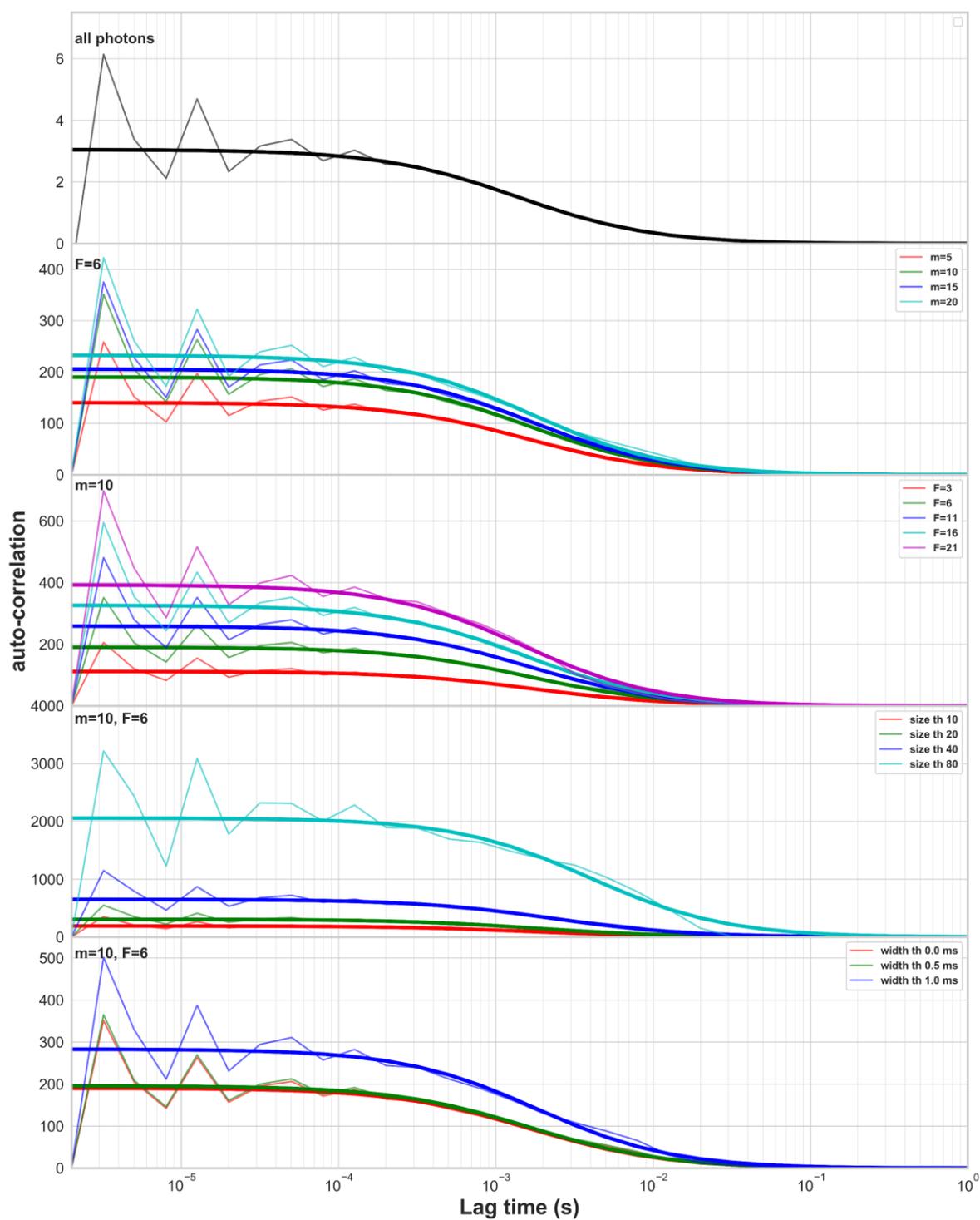
275   **Figure S12**

276

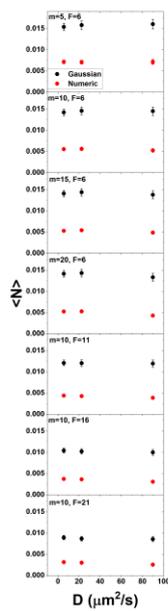
277

278

279   **Figure S13**

280

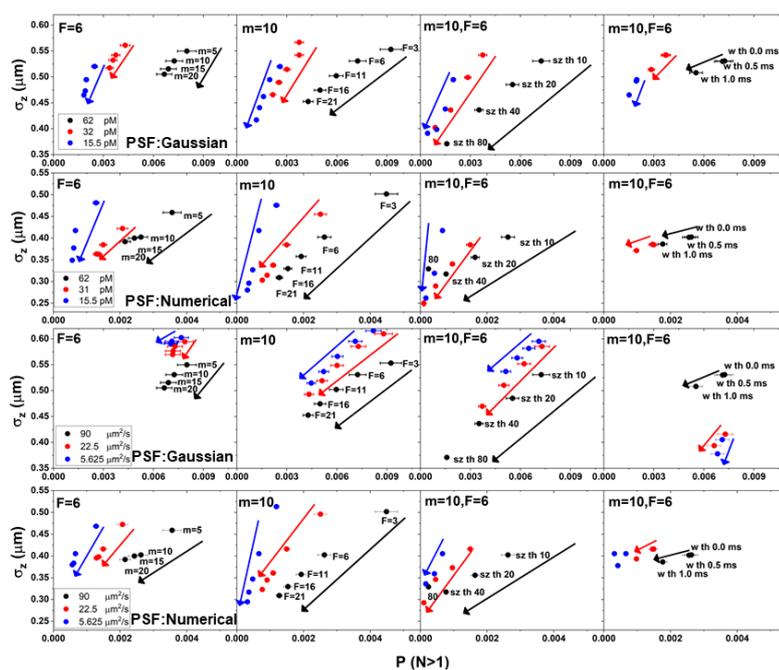
281

282
283 **Figure S14**

284

285

286

287   **Figure S15**

288

289

290

291 **Figure S16**

292

293

294

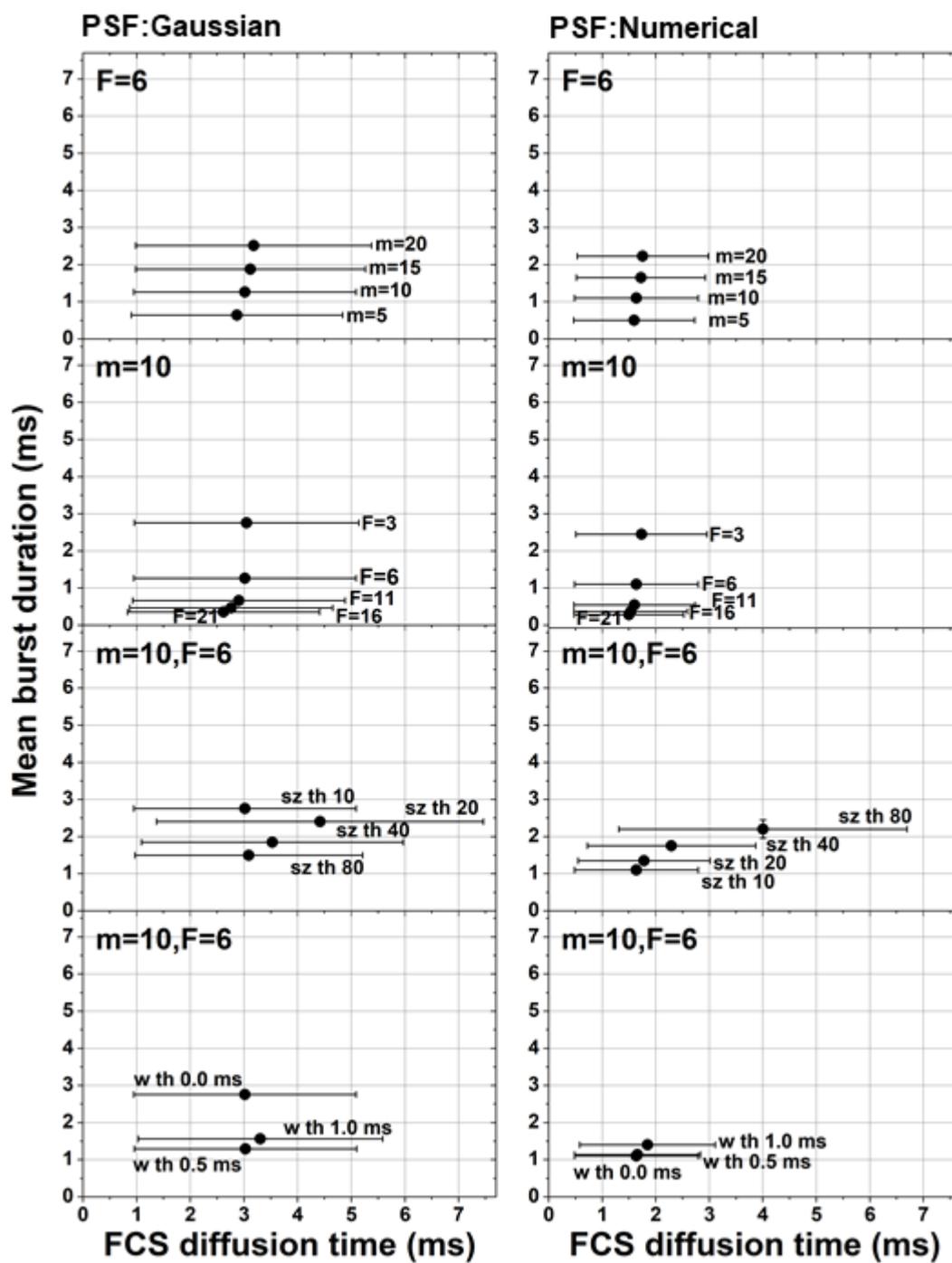



Figure S17

296

297

298

Figure S18

299

300

301

302

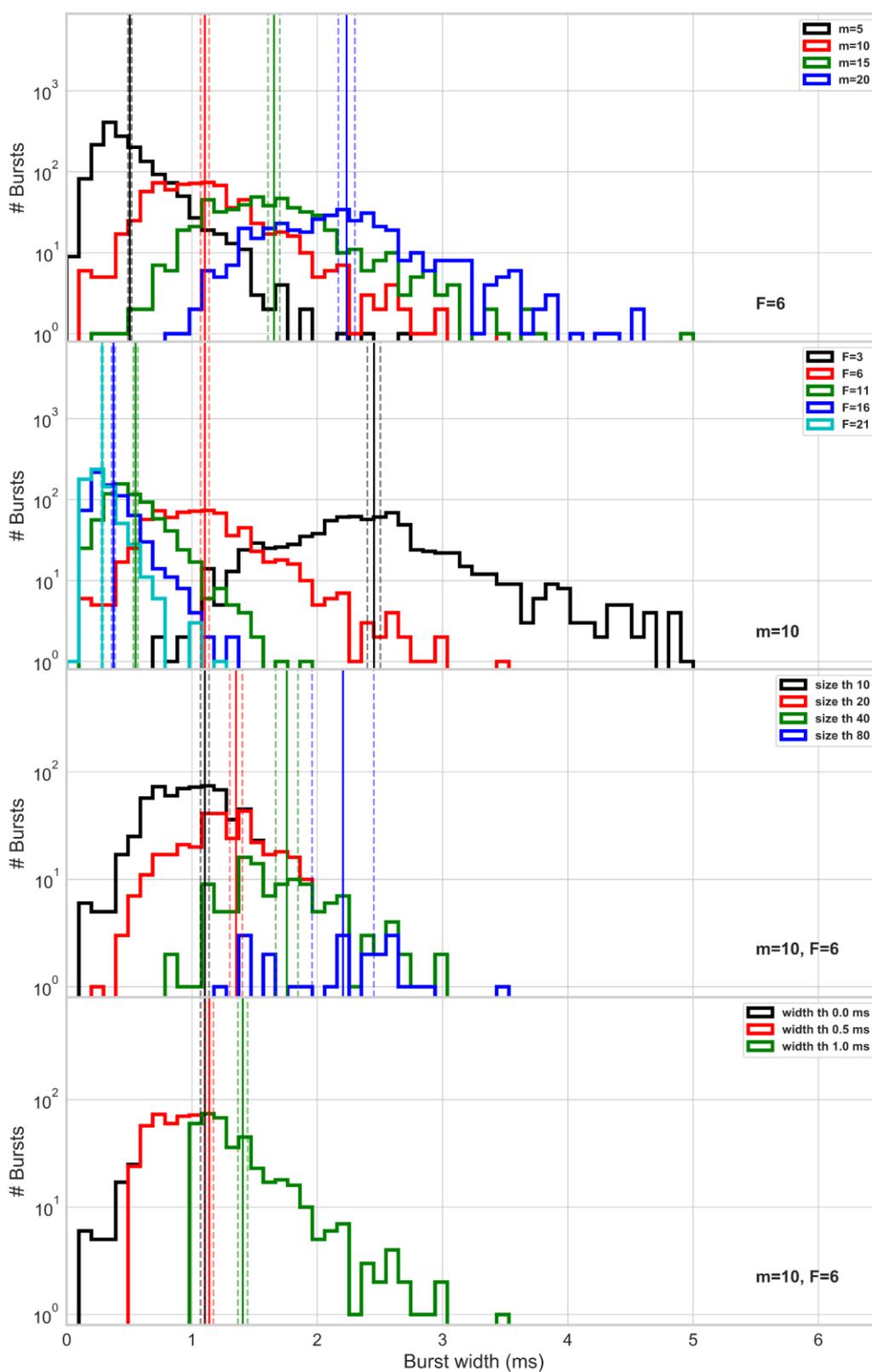
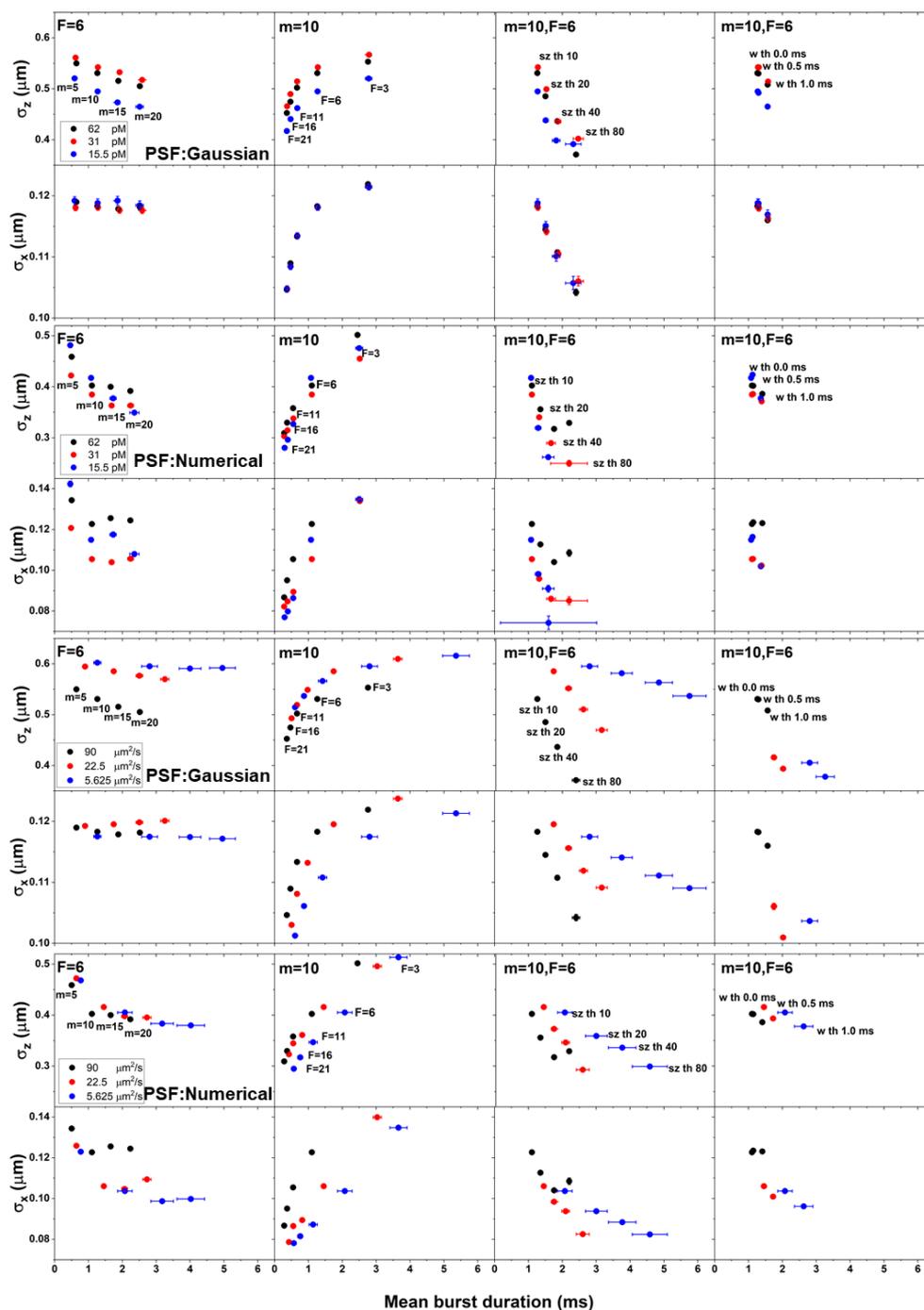
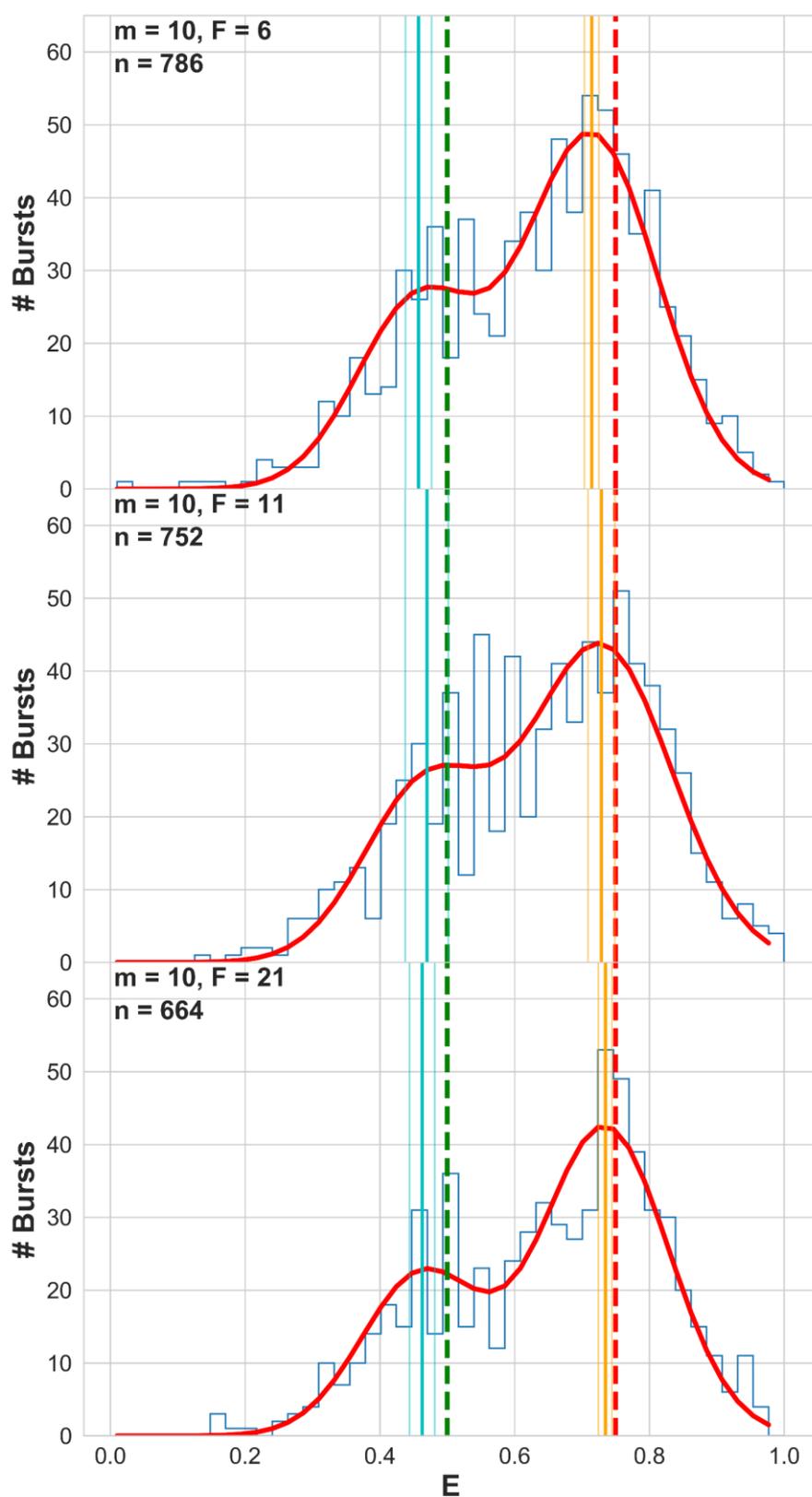



Figure S19

304

305

306
307 **Figure S20**

308

309

310
311 **Figure S21**

313

314