Biophysical Journal, Volume *111*

Supplemental Information

MEMLET: An Easy-to-Use Tool for Data Fitting and Model Comparison Using Maximum-Likelihood Estimation

Michael S. Woody, John H. Lewis, Michael J. Greenberg, Yale E. Goldman, and E. Michael Ostap

Supporting Material

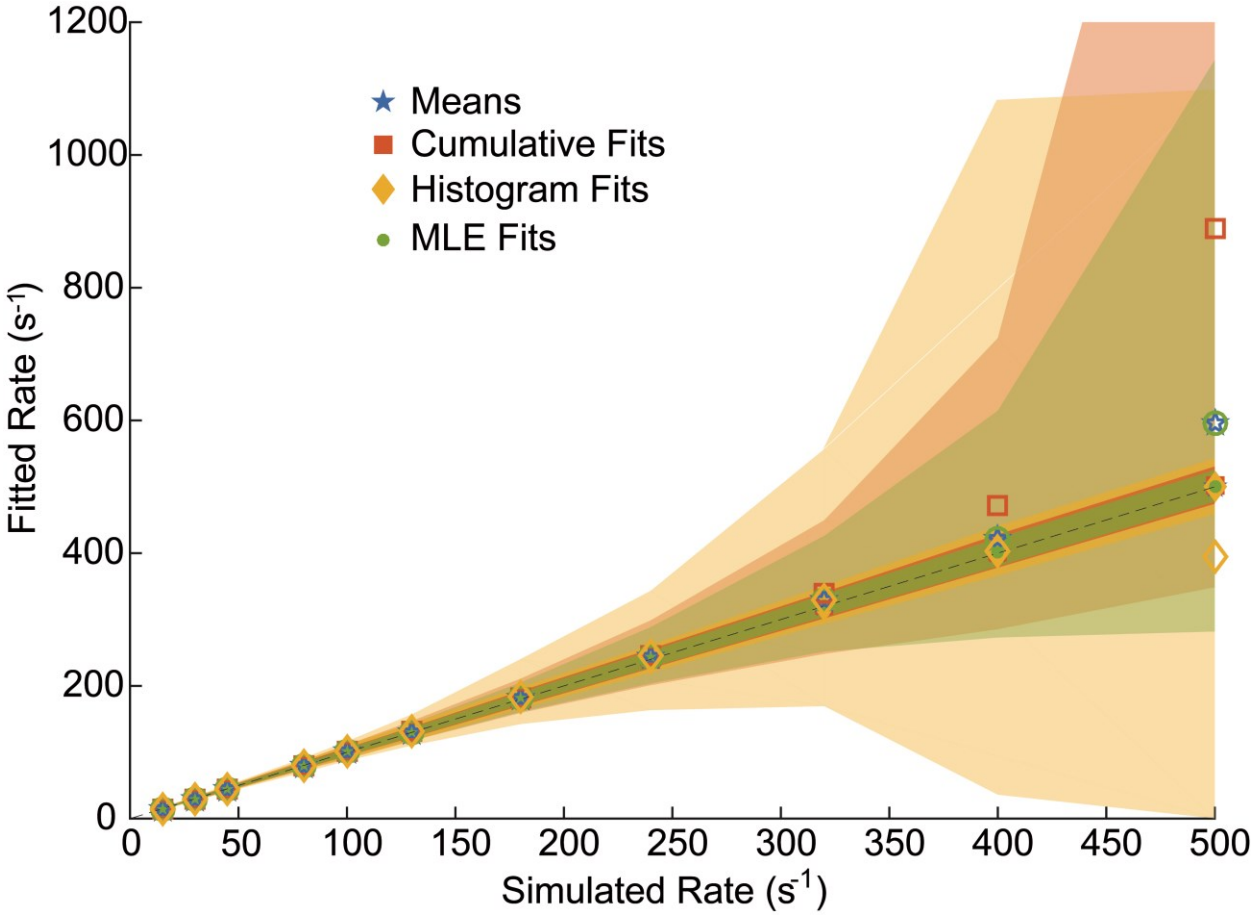
MEMLET: An Easy-to-use Tool for Data Fitting and Model Comparison Using Maximum Likelihood Estimation

Michael S. Woody[1], John H. Lewis[2], Michael J. Greenberg[1,3], Yale E. Goldman[1], E. Michael Ostap[1].

[1]Pennsylvania Muscle Institute, University of Pennsylvania, Philadelphia, PA, USA, [2]Department of Physiology, University of Pennsylvania, Philadelphia, PA, USA, [3]Current address: Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, MO, USA

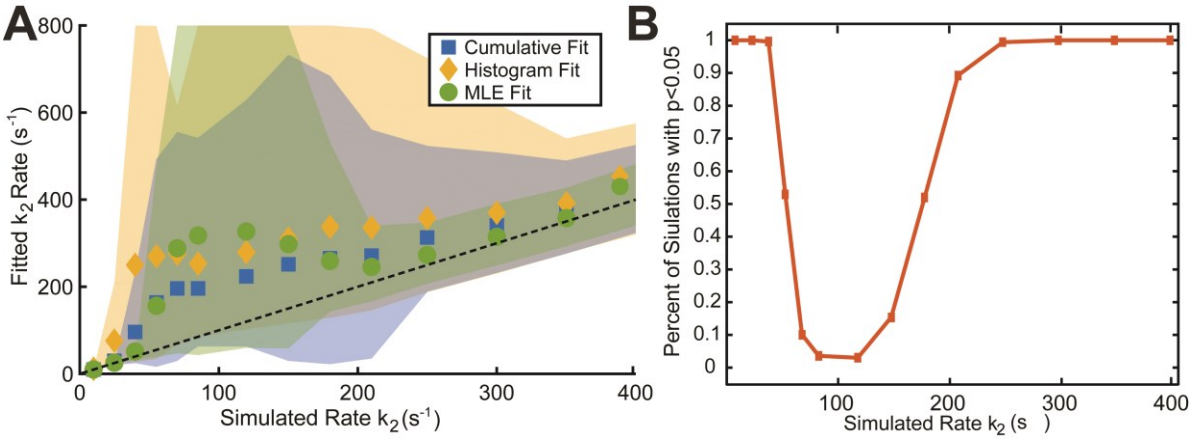
Supporting Figures and Tables

Figure S1. Comparison of fitting techniques with dead-time correction techniques used for all method for single exponential data.



Data were generated in the same way as in Figure 3, but dead-time corrections were used for all fitting methods. The inverse method was corrected as described in the text by subtracting the dead-time from the mean duration before calculating the inverse value. The fits to the binned data points were corrected by using the same dead-time corrected PDF used in the MLE fit and ignoring any bins whose range was less than or included the dead-time. Cumulative distributions were corrected by subtracting a floating amplitude term from the PDF given in Figure 3. The corrected mean method gives the same result as the MLE method (as expected and described in the main text), however this inverse mean method is not applicable to data with multiple exponential components. When the number of data points is kept constant at 1000 (closed symbols and dark shaded areas representing 90% confidence intervals from 1000 rounds of simulations), the fits all maintain accuracy, but the MLE and mean fits show the smallest confidence intervals (mean confidence intervals not shown for clarity, but are the same as the MLE). When the number of points being fit are allowed to decrease with increasing rate due to more events being shorter than the 10ms dead-time (open symbols and light shaded areas), the MLE (and mean) fit offers the more accurate fit and lowest confidence intervals, particularly at high rates.

Figure S2. The Effect of Rates on the Ability to Distinguish Two Phases of a Double Exponential Distribution.



Data sets composed of 1000 points each were simulated by drawing from a double exponential distribution with k_1 = 100 s^{-1} and varying k_2. 50% of events came from the variable rate process, and no dead time was imposed on the data. 1000 independent datasets were generated and fit using cumulative distributions, histograms, and MEMLET's MLE algorithm. (A) The fitted rate for k_2 is plotted, along with 90% confidence intervals. The variable rate is difficult to accurately fit when it is 2-fold (50-200 s^{-1}) of k_1. MLE fitting consistently yielded a more accurate rate with lower confidence intervals outside of this 2-fold range than the other methods tested. (B) Plot of the percentage of 1000 simulations in which the log-likelihood ratio test yields a p-value below 0.05, indicating a double exponential fit is justified over a single exponential fit. When the value of k_2 approaches k_1 (100 s^{-1}), the log-likelihood ratio test is unable to distinguish the two components.

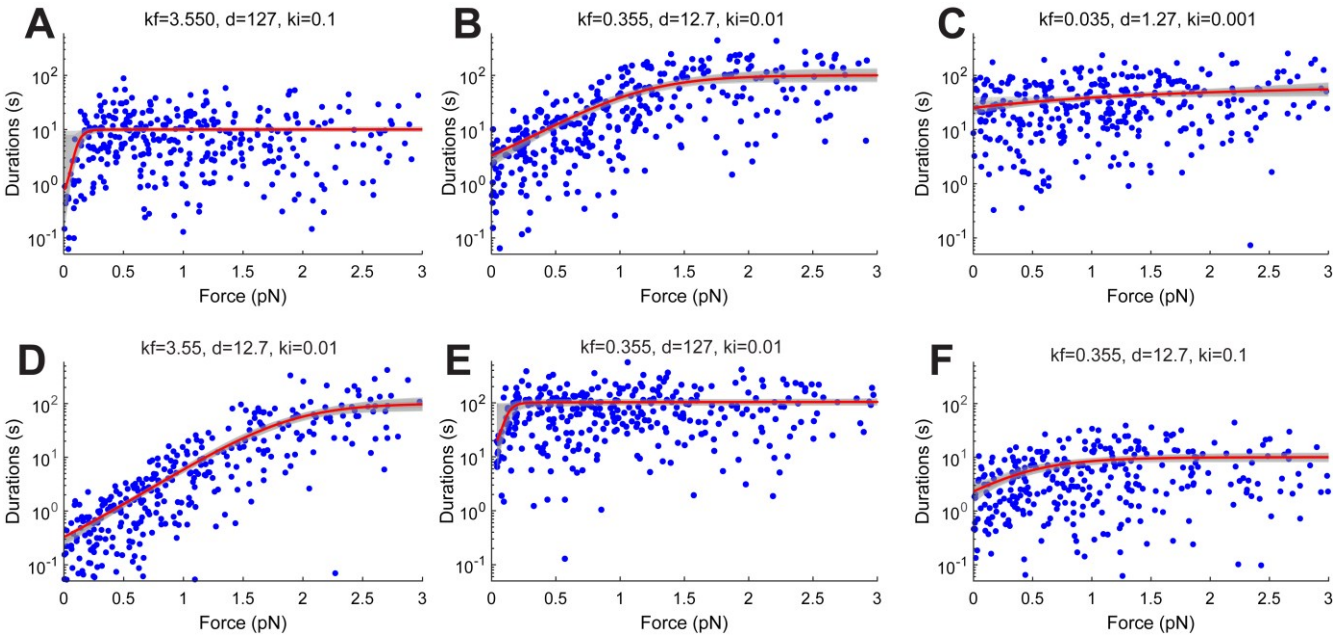
Table S1. Fit Results and Model Testing from Figure 7

Model	k_0	d	k_i	Log likihood value	p-value (cf. Model 1)	PDF used (not including dead-time renormalization)
One Force dependent phase and One force independent phase in parallel	0.348	12.23	0.0081	-1338	--	$(k_i + k_f)e^{-(k_i+k_f)t}$ where $k_f = k_0 e^{-(-\frac{Fd}{k_B T})}$
Bell equation	0.2295	7.956	0 (const)	-1350	$6\text{x}10^{-7}$	
Single Exponential	0 (const)	--	0.0260	-1515	$<1\text{x}10^{-16}$	
One Force dependent phase and One force independent phase in series	0.2205	7.81	50.18	-1350	See note below	$\frac{k_f * k_i}{k_i - k_f}\left(e^{-k_f t} - e^{-k_i t}\right)$ where $k_f = k_0 * e^{-(-\frac{Fd}{k_B T})}$

Results of fitting various models to the data show in Fig 7 in the main text. The Log-likelihood ratio test can be applied to determine p-values for whether the first equation (One Force dependent phase and One force independent phase in parallel) is statistically justified over other models.

Note: The independent series PDF can not be written as a simplified version of the parallel PDF, so it's not strictly possible to perform the log-likelihood testing to compare the two models. However, from the log-likelihoods and the fitted values, it can be seen that the series PDF fits no better than the Bell Equation, which is statsically less significant than the parallel fit and has one more degree of freedom. This shows that the series fit is inferior to the parallel fit for this dataset.

Figure S3. Simulated 2D Data Similar to that Shown in Figure 7 with Fits



Simulated data using the same kinetic model that was used to fit the data in Figure 7. 500 simulated datasets of 329 points each were generated using the "One Force dependent phase and One force independent phase in parallel" PDF from Table S1 and then were fit using the MLE fitting method. The simulated parameters are given at the top of each panel. 95% Confidence intervals (grey) were determined using the results of the 500 independently fit datasets. The program is able to accurately fit the parameters over a wide range of input parameters, including those similar to the data shown in Figure 7 (Panel B). The fitting performs worst when the distance parameter (d) is very low (Panel C), or very high (Panels A and E), because in such cases the force range being simulated is either too small or large, respectively, to show the effect of the force dependent rate. Tables S2 and S3 show the performance of the fits from each case.

Table S2. Parameters for simulated data from Fig. S3 and fits to data in Fig. 7

Panel	Fitted k_f (s^1)	Fitted d (nm)	Fitted k_i (s^1)	PDF used
A	4.20 (2.13-5.90)	125 (95.5-157)	0.0986 (0.0944-0.106)	$(k_i + k_f)e^{-(k_i+k_f)t}$ where $k_f = k_0 e^{-(-\frac{Fd}{k_B T})}$
B	0.359 (0.312-0.417)	12.6 (11.6-13.9)	0.0102 (0.00876-0.0115)	
C	0.0256 (0.0183-0.0331)	7.41 (2.53-5.87)	0.0160 (0.0120-0.0209)	
D	3.61 (3.19-4.06)	12.7 (12.1-13.4)	0.0103 (0.00851-0.0118)	
E	0.410 (0.133-0.596)	111 (71.6-155)	0.00956 (0.00890-0.0104)	
F	0.367 (0.271-0.459)	12.8 (9.39-16.4)	0.100 (0.0905-0.110)	
Fig 7 data	0.355 (0.204-0.684)	12.70 (9.63-16.28)	0.00920(0.00490-0.0110)	

Parameters of the simulation shown in Fig S3 that shows MEMLET is capable of accurately fitting data similar to that presented in Fig. 7. Number in parenthesis indicated 95% confidence intervals. Fits to the data in Figure 7 with 95% confidence intervals show in the last row.

Table S3. Relative errors in fits to simulated data in Fig. S3

Panel	**k_f (s^1) sim**	**k_f error**	**d (nm) sim**	**d error**	**k_i (s^1) sim**	**k_i error**
A	3.55	18.3 %	127	-1.84 %	0.1	-1.42 %
B	**0.355**	**1 %**	**12.7**	**-0.803 %**	**0.01**	**2.31 %**
C	0.0355	-27.8 %	1.27	483 %	0.001	1500 %
D	3.55	1.58 %	12.7	0.0101 %	0.01	2.96 %
E	0.355	15.5 %	127	-12.3 %	0.01	-4.41 %
F	0.355	3.27 %	12.7	0.769 %	0.1	0.347 %

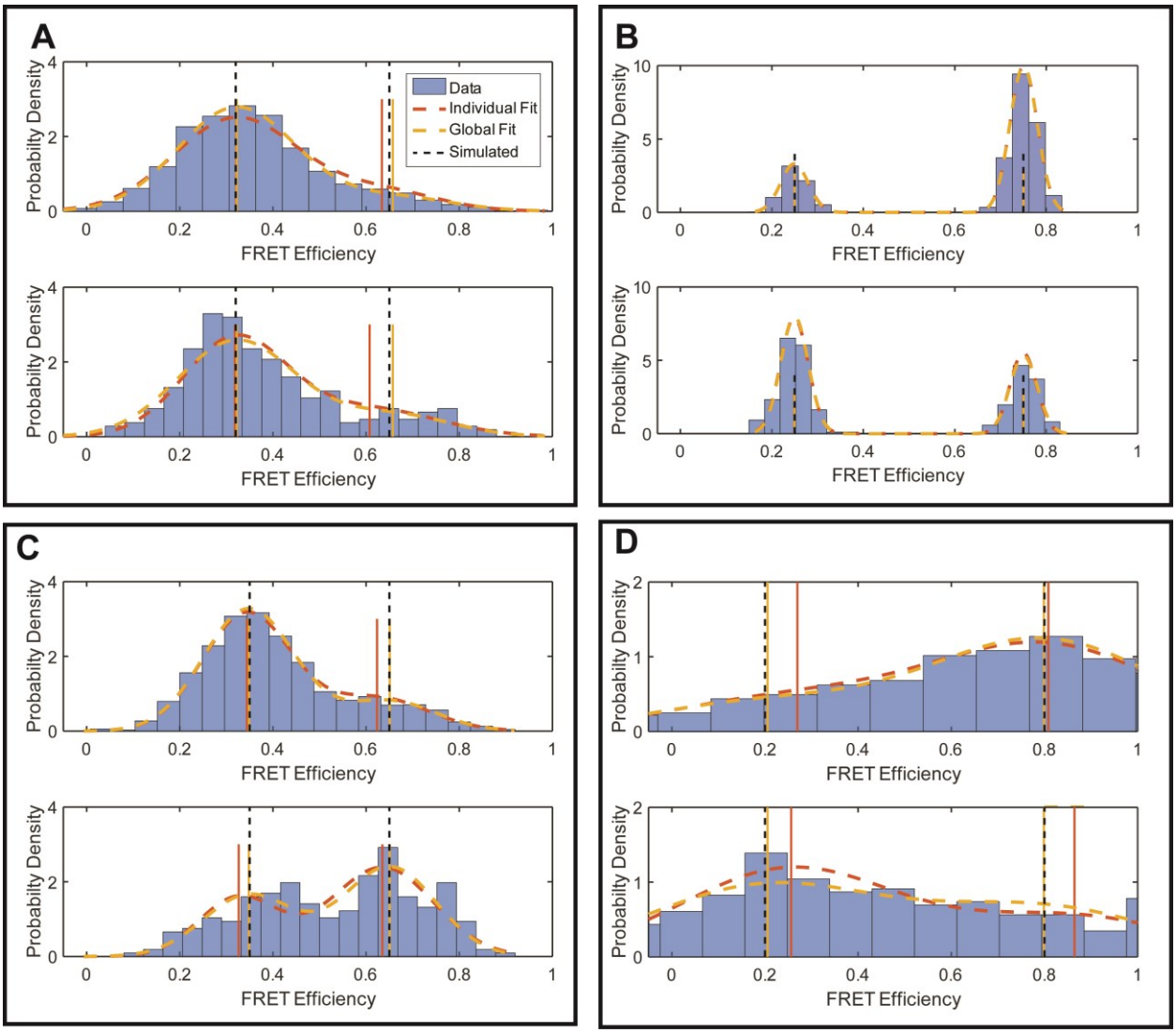
Values of the simulated parameters from Fig S3 and the percent error of the fits from the simulated values.

Table S4. Model Comparision and Fitted Values for Data in Figure 10

Individual Fits						
	A	mu1	sig1	mu2	sig2	Log-Likelihood
top	0.0598	0.734	0.109	0.332	0.153	467.603
bottom	0.2086	0.629	0.116	0.310	0.132	-77.49
				Sum of Individual LL		390.113
Global Fit (mu1,sig1,mu2,sig2 shared)						
glob top	0.1196	0.635	0.143	0.318	0.145	467.043
glob bot	0.1843	0.635	0.143	0.318	0.145	-76.169
				Sum of Global LL		390.874
				2* Ratio of Sum of Global & Indiv LL		1.522
				p value for 4 degrees of freedom		0.82

Table S3 shows the values of the fits from Figure 10, inlcuding the log-likelihoods for both the inidivudual and global fits, where the amplitude was allow to vary between datasets. The goodness-of-fit of the global fit compared to the individual fits can be compared using the log-likelihood ratio test by considering both datasets together. The sums of the log-likelihoods of the two datasets combined can be compared between the individual and global fits. There is a difference of four degrees of freedom between the global and inidivdual datasets (10 free fitting variables versus 6). This yields a p-value of 0.82 between the global and individual fits, indicating that the inidividual fits are not statistically justified over the global fit.

Figure S4. Simulated Double Gaussian Data with Global Fits vs. Individual fits



Simulated data with the same number of points as the data shown in Figure 10. Panel A shows a simulated dataset with similar values to that of Figure 10. Other panels show how the global fit can improve or match the accuracy of the fitted parameters compared to individual fits for a wide variety of parameters. Black dashed lines show the simulated peak positions, while the red line shows the peak positions from individually fitting the top and bottom datasets separately. Yellow lines show the peak positions when the top and bottom datasets were fit globally with only the amplitude of each of the two Gaussian components varying between the two datasets. Simulation and fitted parameters are given in Table S5.

Table S5. Simulation Parameters for Simulated Data in Figure S4.

	Panel A				Panel B			
	Simulated Value	**Individual top**	**Individual bottom**	**Global fit**	**Simulated Value**	**Individual top**	**Individual bottom**	**Global fit**
A_{top}	0.88	-4.59%	--	0.01%	0.25	-0.1%	--	-0.1%
A_{bot}	0.82	--	-11.29%	-0.67%	0.6	--	0.1%	0.1%
mu_1	0.32	0.34%	-0.44%	0.72%	0.25	-0.1%	-0.4%	-0.2%
sig_1	0.144	7.28%	-10.96%	0.96%	0.03	1.3%	1.3%	0.0%
mu_2	0.64	-2.48%	-6.54%	1.12%	0.75	0.1%	0.0%	0.1%
sig_2	0.144	-3.04%	17.60%	-0.16%	0.03	0.3%	-3.7%	0.3%

	Panel C				Panel D			
	Simulated Value	**Individual top**	**Individual bottom**	**Global fit**	**Simulated Value**	**Individual top**	**Individual bottom**	**Global fit**
A_{top}	0.8	-5.6%	--	0.0%	0.25	25.2%	--	-0.1%
A_{bot}	0.4	--	-15.1%	0.6%	0.6	--	15.8%	0.3%
mu_1	0.35	-1.7%	-6.7%	-0.5%	0.2	34.8%	28.1%	2.7%
sig_1	0.1	-4.4%	-14.2%	-2.6%	0.25	8.8%	-6.5%	0.5%
mu_2	0.65	-4.1%	-2.3%	0.1%	0.8	1.1%	8.1%	-0.2%
sig_2	0.1	9.8%	10.4%	-0.6%	0.25	-2.6%	-9.0%	-2.4%

Simulated values and the percent error of the individual fits to the top and bottom datasets from each Panel in Figure S4 compared to the percent error of the Global fit to each Panel. Panel A most closely resembles the data shown in Figure 10. In Panels A, C, and D, the global fits have a lower error than the individual fits.