Chemical Measurement and Fluctuation Scaling

Quentin S. Hanley*

School of Science and Technology
Nottingham Trent University
Clifton Lane
Nottingham NG11 8NS
United Kingdom
quentin.hanley@ntu.ac.uk
Telephone: +44 (0) 115 848 3536

*** Corresponding author:** Q. S. Hanley

On first inspection, fluctuation scaling data may appear to approximate a straight line when log transformed. The data presented in figure 5 of the main text gives a reasonable approximation to a straight line and for many purposes this would be sufficient. The purpose of the study of fluorescence intensity was to determine whether adjusting the voltage of a photomultiplier tube while measuring a fluorescent sample changes the process (α), the dispersion (β) and/or the gain (G). In this regard, the linear model established that PMT setting affects more than the gain. However, a detailed analysis beginning with testing for model mis-specification provides additional information. Specifically, Poisson behavior is only seen over a limited wavelength range in the 600 V and 700 V data sets.

Tests for Mis-Specification:

Three tests for mis-specification were applied to the data in figure 5: the rainbow test,[1] the Harvey-Collier test,[2] and the Ramsey reset test[3] (table S1) using the lmtest package (version 0.9-34) for the R statistical program (version 3.3.1). As can be seen, although there is some disagreement among the tests, the HC and reset tests consistently found probable mis-specification.

Table S1: Mis-specification tests for the Power law representations in Figure 5. Tests deemed significant have been highlighted in bold. In the main text, the 600 V rainbow test was considered insignificant on its own on the basis of multiple comparisons.

PMT Voltage	Representation	Test	Test Statistic	p-value
400 V	Power law	Rainbow	0.95802	0.5771
400 V	**Power law**	**HC**	**2.4277**	**0.01628**
400 V	**Power law**	**Reset**	**4.4614**	**0.013**
600 V	**Power law**	**Rainbow**	**1.7099**	**0.008121**
600 V	**Power law**	**HC**	**2.8499**	**0.004938**
600 V	**Power law**	**Reset**	**20.343**	**1.312e-08**
700 V	Power law	Rainbow	0.80585	0.8211
700 V	**Power law**	**HC**	**3.4619**	**0.000705**
700 V	**Power law**	**Reset**	**10.569**	**5.215e-05**

Application of Piecewise Linear Modelling: Based on the finding of likely mis-specification, the data were then fit to a piecewise linear model with a single breakpoint, *c*. The treatment here is based on descriptions in Ryan and Porth.[4] The basic model is

$$\log(s) = \log\left(G_1^{(1-\alpha_1)}\beta_1\right) + \alpha_1 \log(\bar{M}) \text{ for } \log(\bar{M}) \le \log(c)$$
$$\log(s) = \log\left(G_2^{(1-\alpha_2)}\beta_2\right) + \alpha_2 \log(\bar{M}) \text{ for } \log(\bar{M}) > \log(c). \quad (1)$$

The number of variables in the fit can be reduced by assuming the two segments are continuous at the breakpoint.

$$\log\left(G_1^{(1-\alpha_1)}\beta_1\right)+\alpha_1\log(\bar{M})=\log\left(G_2^{(1-\alpha_2)}\beta_2\right)+\alpha_2\log(\bar{M}) \text{ when } \log(\bar{M})=\log(c) \quad (2)$$

The fit is then made to the 4 parameter ($\log\left(G_1^{(1-\alpha_1)}\beta_1\right), \alpha_1, \alpha_2$, and $\log(c)$) model

$$\log(s)=\log\left(G_1^{(1-\alpha_1)}\beta_1\right)+\alpha_1\log(\bar{M}) \text{ for } \log(\bar{M})\le\log(c)$$
$$\log(s)=\log\left(G_1^{(1-\alpha_1)}\beta_1\right)+c(\alpha_1-\alpha_2)+\alpha_2\log(\bar{M}) \text{ for } \log(\bar{M})>\log(c). \quad (3)$$

After fitting the data, the parameter $\log\left(G_2^{(1-\alpha_2)}\beta_2\right)$ can be obtained by the formula,

$$\log\left(G_2^{(1-\alpha_2)}\beta_2\right)=\log\left(G_2^{(1-\alpha_2)}\beta_2\right)+\log(c)(\alpha_1-\alpha_2). \quad (4)$$

This analysis (table S2, Figure S1) was done using the SiZeR package (version 0.1-4) in R using the piecewise.linear function.

Table S2: Piecewise regression coefficients. The high (H CI) and low (L CI) values of the 95% confidence interval are also provided. Only the confidence intervals for the high and low exponents and the position of the breakpoint are provided by the SiZeR package which uses a bootstrap procedure. The remaining confidence intervals were obtained by standard regression using the position of the breakpoint to segment the data sets.

PMT Voltage	$\log\left(G_1^{(1-\alpha_1)}\beta_1\right)$	$G_1^{(1-\alpha_1)}\beta_1$	$\log\left(G_2^{(1-\alpha_2)}\beta_2\right)$	$G_2^{(1-\alpha_2)}\beta_2$	α_1	α_2	$\log(c)$	c
400V	-1.25	0.056	-1.23	0.058	-0.02	0.15	-0.09	0.82
L CI	-1.31	0.048	-1.28	0.052	-1.32	0.07	-0.90	0.12
H CI	-1.18	0.066	-1.19	0.064	0.04	0.67	0.57	3.71
600V	-1.09	0.081	-1.61	0.025	0.46	0.79	1.53	33.72
L CI	-1.15	0.071	-1.80	0.016	0.03	0.65	0.49	3.07
H CI	-1.04	0.091	-1.42	0.038	0.51	0.94	1.83	67.61
700V	-0.87	0.135	-1.34	0.046	0.51	0.75	1.99	98.32
L CI	-0.95	0.112	-1.59	0.026	0.47	0.68	1.88	75.16
H CI	-0.80	0.158	-1.15	0.071	0.56	0.86	2.32	210.38

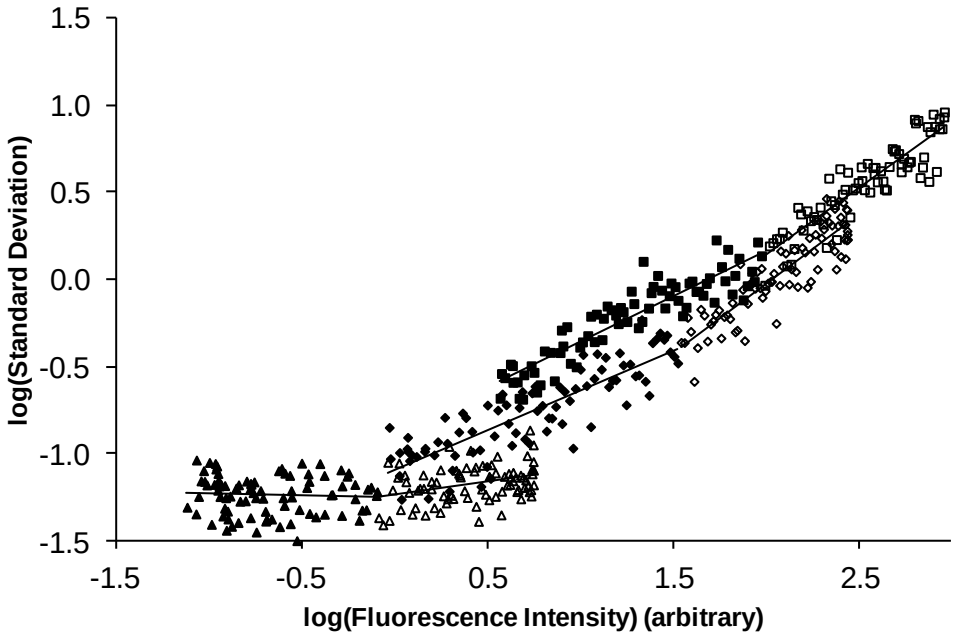

Figure S1: Piecewise model applied to the data. Solid markers represent the low portion of the model and open markers represent the high portion. The voltage settings are indicated by triangles (400 V), diamonds (600 V), and squares (700 V).

A number of points related to the exponents and breakpoints are worth noting. 1) The confidence interval for α_1 (the low portion) of the 400V data set includes 0 which means this is consistent with a Gaussian process. 2) The 95% confidence intervals overlap for both α_1 and α_2 in the 600V and 700V data sets indicating that we can consider these exponents the same (e.g. arising from the same process). 3) Although the value of the breakpoint differs significantly in the data sets, the position of the breakpoint in a value sorted data set can be considered the same in all three cases. 4) Only the low intensity segments of the 600 V and 700 V data sets are reasonably consistent with a Poisson process. For the higher intensity segments, both exponents (α_2 = 0.75 & 0.79) and their 95% confidence intervals include α = 0.85 making them more consistent with a Horwitz process than a Poisson process. 5) If the regime corresponding to α_1 = 0.51 in the 700 V data sets is considered to be Poisson, then

traditional mean-variance[5,6] gives the absolute Gain (0.02) which corresponds to 50 photons per arbitrary unit.

The G and β parameters can now be considered. From a consideration of the spectra, relative Gain parameters can be obtained (Figure S2). Using the 600V data as the reference, the G_R and β_R parameters may be obtained (Table S3). Because the absolute gain, G, is known to be 50 photons/unit for the 700 V low intensity segment, G can be computed for the low intensity portions of the 600 V (200 photons/unit) and 400 V (9560 photons/unit) data sets. β_R gives a somewhat more nuanced view than the presentation in the main text with best dispersion depending on which segment of the fluctuation scaling applies. What is clear from the results is that the regime corresponding to a Gaussian process (400 V) is not desirable in this instance.

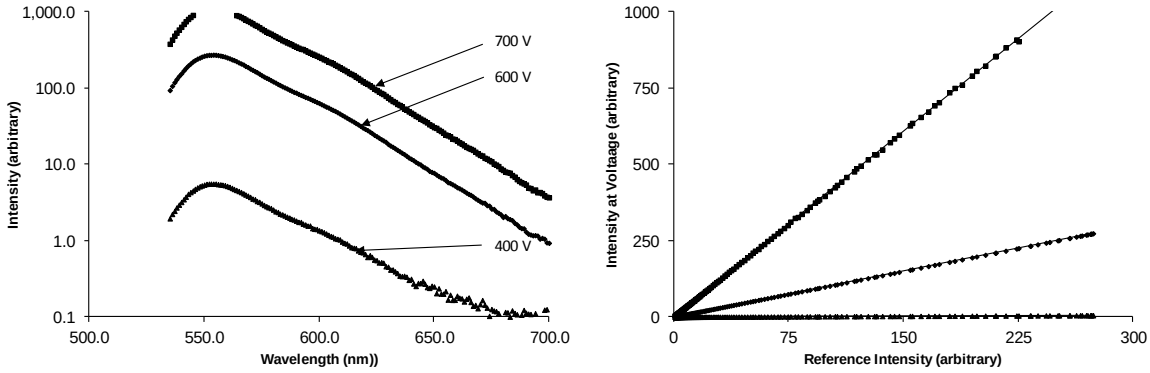

Figure S2: Relative Gain assessment. This data set represents spectra measured using the same sample at 400 V, 600 V, and 700 V (left panel). The position of the change point has been highlighted in each spectrum with an arrow. The gap in the 700 V data set is due to saturation of the instrument. Considering the 600 V data set to be the reference, relative gain parameters can be obtained from the slope of plots of intensities relative to the reference measured using the three conditions (right panel). The voltage settings are indicated by triangles (400 V), diamonds (600 V), and squares (700 V).

Table S3: Relative gain an dispersion parameters for the three PMT settings.

PMT Voltage	G_{R1}	β_{R1}	G_{R2}	β_{R2}
400 V	0.021	2.938	0.021	1.573

600 V	1.000	0.081	1.000	0.025
700 V	4.055	0.068	4.055	0.032

The final consideration for this data set is to try to understand the two portions of the piecewise fluctuation scaling curve. All three data sets exhibited complex scaling with the change point at roughly the same position in the intensity data set when sorted. For the Rhodamine 6G sample measured, the low intensity scaling regime appears entirely in the portion of the spectrum below approximately 615 nm. The reasons for this are unclear but may be due to the characteristics of the photo-emissive material in the PMT or a related wavelength dependent process in the instrument. In the absence of more detailed studies, this remains speculative.

This approach represents an automatable strategy allowing rigorous optimization of instrument parameters and once established the health of instrumentation and methods may be checked to see if process ($\alpha_{observed} \neq \alpha_{expectation}$), dispersion ($\beta_{observed} \neq \beta_{expectation}$), and/or gain ($G_R \neq 1$) have changed.

References:

(1) Utts, J. M. *Commun. Stat. Theory Methods* **1982**, *11*, 2801-2815.
(2) Harvey, A. C.; Collier, P. *J. Econometrics* **1977**, *6*, 103-119.
(3) Ramsey, J. B. *J. Roy. Stat. Soc. B* **1969**, 350-371.
(4) Ryan, S. E.; Porth, L. S. *A tutorial on the piecewise regression approach applied to bedload transport data, General Technical Report RMRS-GTR-189* **2007**.
(5) Sperline, R. P.; Knight, A. K.; Gresham, C. A.; Koppenaal, D. W.; Hieftje, G. M.; Denton, M. *Appl. Spectrosc.* **2005**, *59*, 1315-1323.
(6) Mortara, L.; Fowler, A. *Proc. SPIE, 0290* **1981**, 28-33.