# Chemical Measurement and Fluctuation Scaling

Quentin S. Hanley*

School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

S Supporting Information

ABSTRACT: Fluctuation scaling reports on all processes producing a data set. Some fluctuation scaling relationships, such as the Horwitz curve, follow exponential dispersion models which have useful properties. The mean-variance method applied to Poisson distributed data is a special case of these properties allowing the gain of a system to be measured. Here, a general method is described for investigating gain (G), dispersion (β), and process (α) in any system whose fluctuation scaling follows a simple exponential dispersion model, a segmented exponential dispersion model, or complex scaling following such a model locally. When gain and dispersion cannot be obtained directly, relative parameters, G_R and β_R, may be used. The method was demonstrated on data sets conforming to simple, segmented, and complex scaling. These included mass, fluorescence intensity, and absorbance measurements and specifications for classes of calibration weights. Changes in gain, dispersion, and process were observed in the scaling of these data sets in response to instrument parameters, photon fluxes, mathematical processing, and calibration weight class. The process parameter which limits the type of statistical process that can be invoked to explain a data set typically exhibited $0 < \alpha < 1$, with $\alpha > 4$ possible. With two exceptions, calibration weight class definitions only affected β. Adjusting photomultiplier voltage while measuring fluorescence intensity changed all three parameters ($0 < \alpha < 0.8$; $0 < \beta_R < 3$; $0 < G_R < 4.1$). The method provides a framework for calibrating and interpreting uncertainty in chemical measurement allowing robust comparison of specific instruments, conditions, and methods.
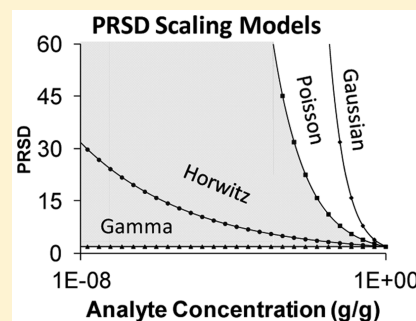
In his classic review of over 150 interlaboratory comparisons published in 1982,[1] Horwitz noted the coefficient of variation scaled inversely with sample concentration, C, and this scaling appeared to be limited by an empirical relationship producing a trumpet shaped curve. By 1997, the Horwitz curve was verified in nearly 10000 studies[2] and considered among the most intriguing relationships in Analytical Chemistry.[3] These classic studies[1,4] provide a view of the fluctuation scaling of interlaboratory comparisons. They also represent one of the greatest challenges to our understanding of uncertainty in chemical measurement and the overall effects of computational, procedural, and methodological practice in the field as a whole. The ubiquity of this trend led to the Horwitz ratio (HorRat) as a criterion for evaluating interlaboratory studies.[4−8]

The heteroscedasticity, nonconstant variance of standard deviation across an interval, made clear by the Horwitz function in interlaboratory comparisons has not been widely appreciated nor has the lesson of this curve been applied to chemical measurements at lower levels of aggregation. Chemical measurements are widely assumed to be normally distributed. This assumption underpins the application of t tests,[9] Q-tests,[10] Grubb's test,[11] and a range of other statistically based procedures for presenting and evaluating chemical data. Where a normal distribution is not explicitly required, our understanding is often framed in terms of a normal distribution (for example, detection limits).[12,13] Procedures used in statistical quality management[14] have particular meanings when data are assumed to follow a normal distribution which may not be correct when the data follow other distributions. The heteroscedasticity inherently

diagnosed by fluctuation scaling makes clearer when these assumptions are applicable.

When a single set of measurements gives reasonable correspondence to a normal distribution, it is easy to believe the process producing it is also Gaussian (e.g., error is independent of scale). This expectation is rarely tested. However, a simple plot of mean and standard deviation can quickly test this assumption and related plots have found wide utility in many areas of science. Urban traffic,[15] stock market trades,[16] crime,[17] measles cases,[18] wind energy,[19] and deviation of prime numbers from Riemann's counting formula[20] have all been shown to have deviations that scale with signal. In physics, scaling of standard deviation with the mean is referred to as fluctuation scaling. In biology, the relationship between mean and variance has become known as Taylor's law. Chemistry has not unified its terminology and, depending on the study, fluctuation scaling behavior might be referred to as Horwitz behavior,[6,7] characteristic functions,[7] uncertainty functions,[21] or not directly named.[22]

Concerns have been raised about the Horwitz curve. There has never been a satisfactory explanation for the relationship and its form suggests that as $C \to 0$ so does the standard deviation.[7] A variety of alternative uncertainty functions have been proposed (c.f., Thompson's review[21]); however, the bulk of these impose a mixed Gaussian-Gamma model (see below for details), with Gaussian behavior at low signal and Gamma behavior at high

A

signal. Thompson presented two alternatives to the Horwitz function including a segmented model composed of Gamma, Horwitz, and Poisson segments[23] and a mixed Gaussian-Gamma model.[7] These studies seek to explain an extremely high level of abstraction: aggregation across multiple interlaboratory studies at multiple concentrations using multiple methods and assessing multiple analytes. That level of abstraction begins with fundamental studies of units and measures, proceeds through basic measurements like mass, volume, and intensity in individual laboratories to be aggregated into results based on complex multistep procedures, and so forth until it reaches the Thompson-Horwitz level. Each level of abstraction can be assessed using fluctuation scaling methods.

This paper describes a generalized method that places these behaviors (e.g.: Horwitz, characteristic and uncertainty functions, etc.) in additional statistical context, applies exponential dispersion model (power law) scaling to measurements made in a single laboratory, introduces the concepts of defined scaling, measured scaling, derived scaling, generalized relative gain, G_R, and relative dispersion, β_R, compares presentations of fluctuation scaling using a set of methods central to chemical measurement (mass, emission intensity, and absorbance), and illustrates the effects of fluctuation scaling on derived results.

Theory. The classic Horwitz curve can be presented in terms of percent relative standard deviation (PRSD) or as a power law of the standard deviation:[24]

$$\text{PRSD} = 2C^{-0.15} \quad \text{or} \quad s = 0.02\overline{x}^{0.85} \tag{1}$$

A plot of the PRSD form produces the classic trumpet shaped Horwitz curve. Derivations based on the binomial distribution, Zipf's law,[3] and a heuristic approach[2] have been attempted; however, there is no general understanding of how it arises.

The Horwitz curve belongs to a family of power law fluctuation scaling relationships of the form

$$\sigma = \beta\overline{x}^{\alpha} \tag{2}$$

where $\overline{x}$ is the mean, σ is the standard deviation, and α and β are constants. This can be recast in terms of the PRSD to give Horwitz's presentation of the same fluctuation scaling relationship, $\text{PRSD} = 100\sigma/\overline{x} = 100\beta\overline{x}^{(\alpha-1)}$.

Particular statistical distributions produce specific fluctuation scaling behavior. The Tweedie family of distributions are exponential dispersion models following the form of eq 2, which exist for values of $\alpha \leq 0$ and ≥ 0.5.[25,26] This family includes the Gaussian distribution ($\alpha = 0$), the Poisson distribution ($\alpha = 0.5$), the Gamma distribution ($\alpha = 1$), and the compound Poisson-Gamma distribution ($0.5 < \alpha < 1$).[25,27] Classic Horwitz behavior ($\alpha = 0.85$) strongly suggests a Poisson-Gamma process (Figure 1). This might arise from a Poisson distributed number of laboratories sampling a Gamma distributed process. This provides statistical context for the coefficients of the Horwitz fluctuation scaling relationship, similar interlaboratory comparisons[28] and for fluctuation scaling of the form of eq 2. This could include measurements made at any level of abstraction: interlaboratory, interanalyst, and intralaboratory comparisons; basic measurements of mass, volume, and intensity; and direct measurements, derived measurements based on several parameters, and single- and multistep procedures. The literature of the fields of mathematics, physics, ecology, and statistics contain a wealth of information on processes leading to eq 2, which may be drawn on to assist interpretation.[6,16,25,29−31] Although scaling according to eq 2 is widely observed, more complex (e.g., segmented linear response after log transformation) scaling has
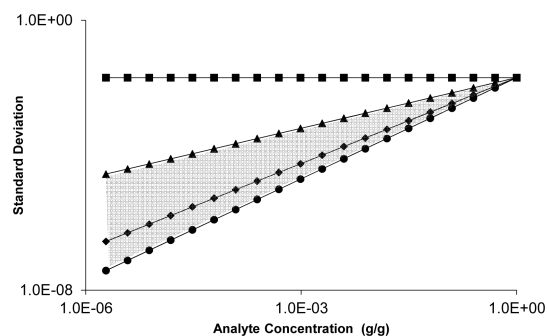
Figure 1. Power Law presentation of fluctuation scaling relationships for Gaussian ($\alpha = 0$, ■), Poisson ($\alpha = 0.5$, ▲), Horwitz ($\alpha = 0.85$, ◆), and Gamma ($\alpha = 1$, ●) processes, with shading highlighting the Poisson-Gamma region. For presentation, the pre-exponential factors have been kept constant.

been observed in measles cases,[18] crime,[17] and analytical chemistry.[22] In analytical chemistry, alternatives to the Horwitz function of the form $\sigma = \sqrt{a^2 + (\beta\overline{x})^2}$ have found utility.[21] In the current context, models of this type suggest a mixed Gaussian-Gamma model, with the Gaussian behavior dominating at low signal.

Fluctuation scaling of the form of eq 2 allows the gain or relative gain of a set of measurements to be defined. For the special case of Poisson distributed data ($\beta = 1$ and $\alpha = 0.5$), the mean-variance technique to measure the gain of charge coupled devices is well-known.[32,33] The approach is extended here to all fluctuation scaling processes following eq 2. If a process generates signals and gain, G, is applied such that the observed mean is $\overline{M} = G\overline{x}$ and the observed standard deviation is $s = G\sigma$, then substitution into eq 2 gives

$$s = G^{(1-\alpha)}\beta\overline{M}^{\alpha} \tag{3}$$

This expression provides a route to assess G, α, and β. If either G or β are known from other measurements, theory, or extensive experience, they may be interpreted directly.

When β is unknown but the processes are believed to be similar, a generalized relative gain may be invoked.[17] For example, two sets of measurements made under different conditions (e.g., in different laboratories or using different instrument parameters) could exhibit two pre-exponential factors in their respective fluctuation scaling plots, such as

$$s = G_1^{(1-\alpha_1)}\beta_1\overline{M}^{\alpha_1} \quad \text{and} \quad s = G_2^{(1-\alpha_2)}\beta_2\overline{M}^{\alpha_2} \tag{4}$$

where the subscripts refer to the first and second set of measurements. If $\beta_1 = \beta_2$ and $\alpha_1 = \alpha_2$, these will cancel from a ratio, R, of measured pre-exponential factors.

$$R = \frac{G_1^{(1-\alpha_1)}\beta_1}{G_2^{(1-\alpha_2)}\beta_2} = \left(\frac{G_1}{G_2}\right)^{(1-\alpha)} \tag{5}$$

Rearranging yields the generalized relative gain, G_R.

$$G_R = \frac{G_1}{G_2} = R^{1/(1-\alpha)} \tag{6}$$

G_R is a measure of amplification that occurs without changing the dispersion of the distribution from which it arises. Alternatively, when G_1 and G_2 are known or when the ratio G_1/G_2 is known this allows a relative dispersion parameter, β_R, to be defined.
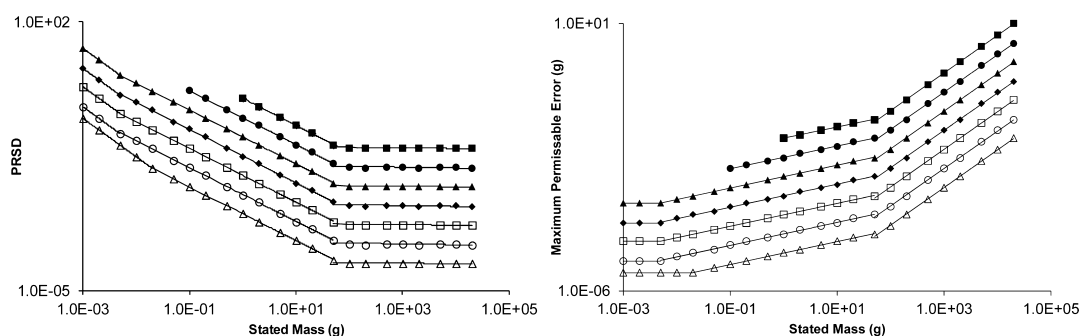
Figure 2. Comparison of PRSD (left panel) and power law scaling (right panel) of OIML defined maximum permissible errors (MPE) in weights of type E_1 ($\triangle$), E_2 ($\bigcirc$), F_1 ($\square$), F_2 ($\blacklozenge$), M_1 ($\blacktriangle$), M_2 ($\bullet$), and M_3 ($\blacksquare$) weights. PRSD was taken to be $100 \times$ MPE/mass.

$$\beta_R = \frac{\beta_1}{\beta_2} = \frac{R}{(G_R)^{(1-\alpha)}} \tag{7}$$

β_R is a measure of the extent that a change in a process stretches the width of the statistical distribution independent of any amplification. The parameters α, G_R, and β_R are related to the HorRat,[4−8] but provide additional details. The HorRat is the observed PRSD divided by the expectation from eq 1 (PRSD version). If the expected values are $\alpha = 0.85$, $G_R = 1$, and $\beta_R = 1$ for a process producing an observed PRSD, there is equivalence, and the HorRat defines the distance from the scaling law. When generalizing to fluctuation scaling following eq 2, the HorRat is insufficient. Specifically, it is possible for two fundamentally different scaling laws to intersect and produce HorRat = 1 at a particular point. It is possible that fundamentally different processes (α's not equal) are involved and it is also possible for G_R and β_R to scale such that there is no net change in HorRat.

The considerations represented by eqs 3−7 can also be used to understand segmented scaling such as that of the form $\sigma = \sqrt{a^2 + (\beta \bar{x})^2}$ provided the data conform to eq 2 over identifiable segments. In more complex cases where segmented scaling is not observed (e.g., absorbance), the derivative of log−log transformed uncertainty functions gives local estimates of α.

In addition to the generalized gain and dispersion parameters, three definitions are provided to describe types of scaling behavior. Defined fluctuation scaling is where allowable tolerances are defined by ASTM, IUPAC, and related agencies for classes of things such as calibration weights. These definitions may sometimes limit the comparability of results across laboratories using these items. Measured fluctuation scaling is the behavior obtained by measuring objects or materials over a wide range of values as reported directly by an instrument. Derived fluctuation scaling is scaling after computations are applied to measured results. This could be application of a calibration curve to the data, mathematical manipulation, filtering, and so on.

■ EXPERIMENTAL SECTION

The tolerances of calibration weights were obtained from published tables (OIML R 111−2). Mass measurements were made ($n = 10$) on a Sartorius Extend ED244S balance. Fluorescence measurements ($n = 10$) were made with a Cary Eclipse Fluorimeter. Absorbance measurements ($n = 10$) were made with an Agilent 8454 Diode Array Spectrophotometer. An ND4 filter (Nikon) was introduced to reduce the amount of light present for some of the absorbance measurements. Rhodamine 6G (Sigma-Aldrich) was used as both an absorber and a

fluorophore. Solutions were prepared by dissolving the rhodamine 6G in a small amount of methanol and diluting with water.

Fitting of scaling relationships was done using log transformed data and parameters were estimated by standard regression methods in MS Excel with two exceptions. Tests of the equivalence of α parameters at different PMT voltage settings were done using a regression model with categorical variables (Minitab 17). The rainbow test was applied to the emission intensity data series to assess mis-specification in a straight line model using R (Version 3.3.1) with the lmtest package (version 0.9−34). The piecewise fit to the fluorescence intensity data was done using R with the SiZeR package (Version 0.1−4) providing an estimate of the change point, slope, and intercept parameters with bootstrap confidence intervals.

■ RESULTS AND DISCUSSION

Mass. *Defined Scaling.* Two commonly applied standards for calibration weights are ASTM E617 and OIML R 111−2, which specify tolerances for varying classes of weights. These standards exhibit segmented power law scaling. Considering the recommendations in OIML R 111−2 for calibration weights of type E_1, E_2, F_1, F_2, M_1, M_2, and M_3, three regimes are observed (Figure 2). At low mass, the behavior is Gaussian with $\alpha = 0$. In the intermediate mass range, the scaling exhibited $\alpha = 0.3$, which is between that expected from Gaussian and Poisson distributions. For the largest masses (50−20000 g), $\alpha = 0.98$, closely matching the expectation from a Gamma distribution. Using these weights creates a complex segmented scaling model by definition. ASTM guidelines gives similar results except the regimes are not as clearly delineated.

Relative β_R. The segmented scaling in the defined characteristics is noteworthy because, with the exception of two E_1 class weights, changing the class of calibration weight does not change the scaling exponent. This indicates that weights having the same mass are created by a similar statistical process (α) with different pre-exponential factors. Since the masses have not changed, no gain has been introduced ($G_R = 1$), indicating the assumptions allowing β_R to be calculated apply (eq 7). The intermediate and high mass sets exhibit identical β_R (Table 1) with some variation in the lowest masses. These β_R factors define the increase in uncertainty for the same mass (e.g., the M_1 weight set represents a process with 67−100 times the variation of that producing the E_1 weight set). Changing weight type only changes the dispersion and has no impact on the defining process.
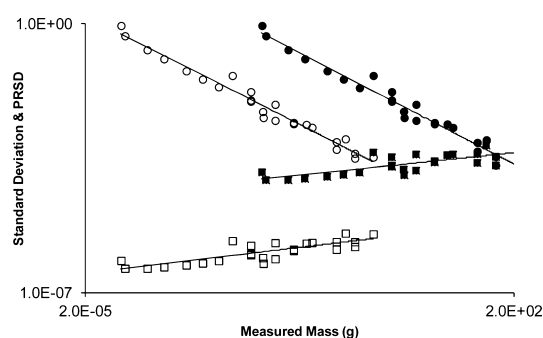
Measured Scaling. A set of calibration weights and laboratory materials covering a mass range from ~15 mg up to 200 g were repeatedly measured to determine the scaling behavior of an analytical balance (Figure 3, filled circles). The power law for the

Table 1. Comparison of Scaling Parameters across the Weight Sets Defined by OIML R 111-2[a]

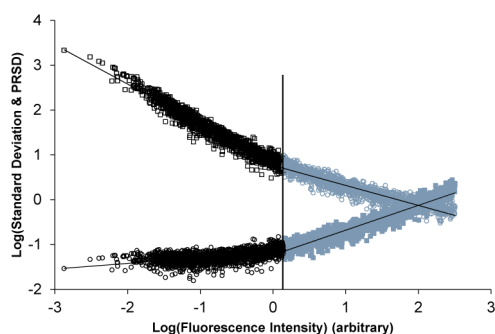| wt set	regime	$G^{(1-\alpha)}\beta$	α	R	β_R
M_1	gaussian	2.00×10^{-4}	0	67	67
F_2	gaussian	6.00×10^{-5}	0	20	20
F_1	gaussian	2.00×10^{-5}	0	6.7	6.7
E_2	gaussian	6.00×10^{-6}	0	2.0	2
E_1	gaussian	3.00×10^{-6}	0	1.0	1
M_1	intermediate	9.83×10^{-4}	0.3	100	100
F_2	intermediate	3.12×10^{-4}	0.3	31.7	31.7
F_1	intermediate	9.83×10^{-5}	0.3	10	10
E_2	intermediate	3.12×10^{-5}	0.3	3.2	3.2
E_1	intermediate	9.83×10^{-6}	0.3	1.0	1.0
M_1	approx. Γ	5.69×10^{-5}	0.98	100	100
F_2	approx. Γ	1.88×10^{-5}	0.98	33.0	33.0
F_1	approx. Γ	5.69×10^{-6}	0.98	10.0	10.0
E_2	approx. Γ	1.88×10^{-6}	0.98	3.3	3.3
E_1	approx. Γ	5.69×10^{-7}	0.98	1.0	1.0

[a]R and β_R parameters have been computed relative to the E_1 weight.

Figure 3. Comparison of PRSD (circles) and power law scaling (squares) presentation of raw mass data (filled symbols) and derived composition (open symbols).

Figure 4. Comparison of PRSD (circles) and power law (squares) scaling of fluorescence intensity. Two regimes are seen delineated by the vertical line. The low signal range (black symbols) approximated, but was not consistent with Gaussian scaling. The higher range (blue symbols) approximated but was also inconsistent with Poisson scaling.

model by the rainbow test ($p < 2.2e-16$), leaving two fluctuation scaling regimes corresponding to low intensity and high intensity, with a break at 1.41 arbitrary intensity units (95% CI 0.92–1.74). The low intensity side approximates Gaussian scaling and might be considered the noise floor of the detection system, except that $\alpha \neq 0$ ($s = 0.0674\overline{x}^{0.124}$). The high intensity portion approximated but was not quite Poisson scaling ($s = 0.0582\overline{x}^{0.554}$) with 0.5 outside the 95% confidence interval of the exponent.

Gain, β_R, and PMT Voltage. The amplification of a fluorescence signal in an instrument of the type used in this study may be adjusted using the voltage applied to a photomultiplier tube (PMT). A simple way to measure this is to record the emission spectrum of a single fluorescent solution while varying the PMT voltage. The variation in the intensities measured across a spectrum provides the data required for fluctuation scaling plots. A comparison of the scaling at PMT voltage settings of 400, 600, and 700 V indicated that adjustment of the PMT fundamentally changed the statistical behavior of the system (Figure 5). Although there is evidence that a linear model represents a mis-specification (see Supporting Information for details), the scaling behavior can be approximated by a function of the form of eq 2. Specifically, at the low setting, the exponent ($\alpha = 0.06$) was near that expected for a Gaussian process. At the higher PMT settings, a single exponent was found ($\alpha = 0.59$). Using the ratio of the measured spectra at 600 and 700 V gave G_R

PRSD curve was PRSD $= 0.018\overline{x}^{-0.83}$ and for s was $s = 0.00018\overline{x}^{0.17}$ conforming to expectation from the two presentations (e.g., $(\alpha - 1)$ and $100 \times \beta$ for PRSD vs α and β for standard deviation). With this confirmation, only representations following eq 2 will be discussed subsequently. These data indicate that the analytical balance data were not produced by a device with error independent of signal as expected for a Gaussian process.
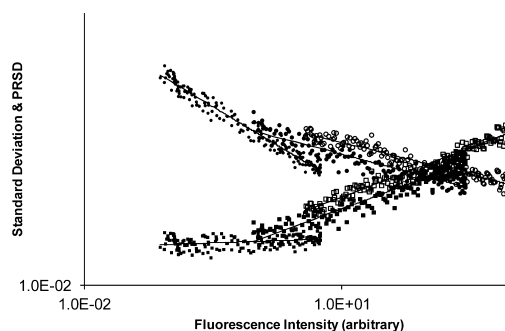
Derived Scaling. In the case of mass, a simple derived quantity might be a mass of analyte (x) per mass of sample (y). If samples consist of 200 g and the analyte can be anywhere from 15 mg up to 200 g, application of a division operation to data in Figure 3 yields modified scaling behavior (Figure 3, open squares). The division operation resulted in both the exponent and pre-exponential changing to $s = 2.4 \times 10^{-6}\overline{x}^{0.19}$. Although the bulk of the change is in the pre-exponential factor, both parameters were affected by the operation.

Fluorescence. Measurement of fluorescence emission intensity is directly related to the number of photons collected. As such, the fluctuations would be expected to be limited by a Poisson scaling law. To investigate the fluctuation scaling of a typical fluorescence measurement, replicate spectra of a rhodamine 6G dilution series were measured using a single photomultiplier voltage (600 V) over a wide range of concentrations and the entire data set replotted in the form of eq 2 (Figure 4). The data were inconsistent with a single line

Figure 5. Effect of PMT voltage on the scaling behavior of a single fluorescent sample. The PRSD presentation is indicated with circles and scaling of standard deviation with squares. The small filled markers correspond to 400 V, the large filled markers were measured at 600 V, and the large open markers were taken at 700 V. A fundamental change in the statistics takes place between 400 and 600 V, as seen in the nearly flat scaling at 400 V and consistent slope at 600 and 700 V.
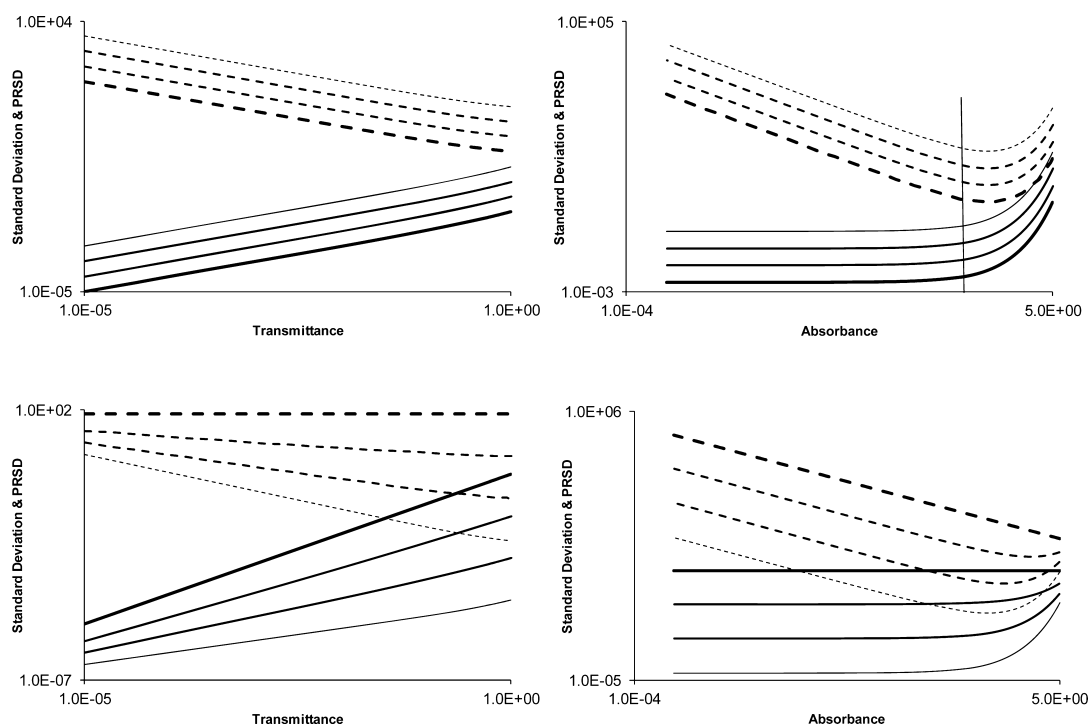
Figure 6. Comparison of simulated PRSD (dashed lines) and power law scaling (solid lines) of transmittance (left panels) and absorbance (right panels). In the upper panels the number of photons in I_0 varied with the narrowest line corresponding to $I_0 = 100$ with each increase in line width representing a factor of 10 increase in I_0. The scaling behavior of transmittance is close to that expected from Poisson statistics while absorbance is more complex but approximates Gaussian scaling below 0.5 (indicated with the vertical line). Above this as I goes to zero, its relative error begins to dominate creating the hockey stick shape. In the lower panels, α is set to 1.0 (widest lines), 0.8, 0.6, and 0.4 (narrowest lines) while holding $G = 1$ and $\beta = 0.5$.

= 4.06, allowing β_R to be found ($\beta_R = 0.88$). These considerations indicate that when going from 600 to 700 V, the width of the distribution compressed slightly. Changing the PMT voltage increased the gain by a factor of 4 while improving the dispersion. Based on these considerations, provided the instrument does not reach saturation, it would be better to measure at 700 V.

In this context, it is worth considering the meaning of a generalized HorRat for a fluorescence intensity measurement. If a community decides that accepted scaling behavior is given by the 700 V data, an ill-judged sample used in a comparison set might appear at the intersection of the 400 and 700 V data sets. This will not diagnose the very different results produced by setting the PMT to 400 V and the HorRat would be one at the intersection. Similarly, the interplay between gain and dispersion cannot be appreciated by a simplistic ratio of scaling laws.

Transmittance and Absorbance. *Defined Scaling.* Although standards are available for calibrating the photometric accuracy of spectrophotometers, standards organizations do not appear to have defined classes of absorbing standards similar to those available for calibration of balances. IUPAC[34] provides guidance for best precision (PRSD) depending on whether the measurement is limited by detector noise or photon noise. It gives no guidance on spectrophotometer quality (e.g., the measurement can be made more precise by collecting more photons) or how to evaluate which type of noise (or a hybrid) dominates. As such, it is largely outdated in this regard. However, uncertainty functions[21] for transmittance and absorbance measurements are easily derived from error propagation rules for division and log transformations. The uncertainty functions for transmittance and absorbance are

$$s_T = T\sqrt{\left(\frac{s_I}{I}\right)^2 + \left(\frac{s_{I_0}}{I_0}\right)^2} \quad \text{and}$$

$$s_A = 0.4343\sqrt{\left(\frac{s_I}{I}\right)^2 + \left(\frac{s_{I_0}}{I_0}\right)^2} \tag{8}$$

If the photons are Poisson distributed and this is measured in the presence of Gaussian detector noise, s_d, s_I may be replaced by $\sqrt{I + (s_d)^2}$.

$$s_T = T\sqrt{\left(\frac{\sqrt{I + (s_d)^2}}{I}\right)^2 + \left(\frac{\sqrt{I_0 + (s_d)^2}}{I_0}\right)^2} \quad \text{and}$$

$$s_A = 0.4343\sqrt{\left(\frac{\sqrt{I + (s_d)^2}}{I}\right)^2 + \left(\frac{\sqrt{I_0 + (s_d)^2}}{I_0}\right)^2} \tag{9}$$

With $s_d = 0$, the simulated behavior (eq 8) of transmittance (Figure 6, top left panel) shows good correspondence to eq 2 with a pre-exponential dependent on the number of photons collected in I_0. The exponent, $\alpha = 0.466$, approximates that expected from a Poisson limited process (e.g., $\alpha = 0.5$) modified by a division operation. As the number of photons increases, no change in transmittance occurred indicating $G_R = 1$. Hence, dispersion changes such that a factor of 10 change in intensity gives $\beta_R = 3.16$.

Transformation of transmittance by the log function (Figure 6, top right panel) has the effect of making absorbance nearly Gaussian when the measured absorbance remains below 0.5 absorbance units. In this range, the exponent can be considered
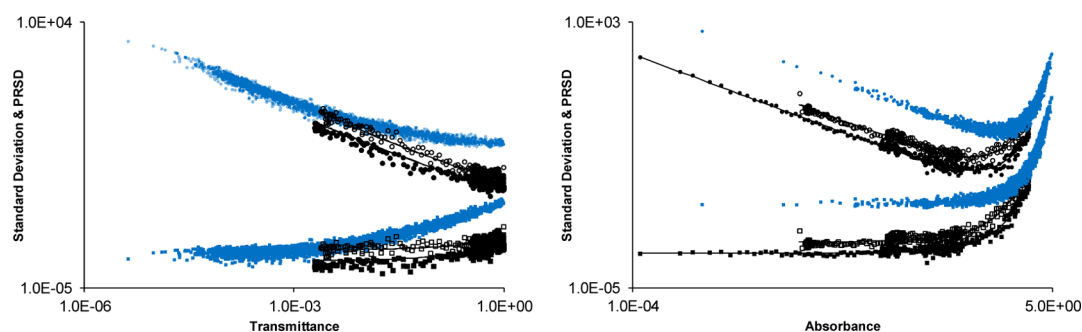
Figure 7. Comparison of PRSD (circles) and power law scaling (squares) of transmittance (left panel) and absorbance (right panel) as I_0 was adjusted (black symbols) with derived scaling from a non-Poisson system (blue symbols). Replicate measurements of a rhodamine 6G solution were made with (open black symbols) and without (filled black symbols) a neutral density (ND4) filter. The minima for the absorbance PRSD were found at 0.59 (with ND4) and 0.67 (without ND4) for the diode array data and ~1.5 for the non-Poisson system.

nearly constant and near zero ($\alpha = 0.027$) and the standard deviation depends on I_0. Similar to transmittance, absorbance exhibits $\beta_R = 3.16$ for a factor of 10 change in I_0. Above 0.5, the scaling behavior curves upward and deviates from the behavior expected from eq 2. This is due to the relative variance of the I term in eq 8 growing rapidly as $I \rightarrow 0$. Consideration of the derivative gives α values well in excess of 1 at high absorbance. The PRSD function reaches a minimum near 0.86 for this detector noise free case reinforcing IUPAC guidance, while making clear the number of photons must be specified for general good practice.

Alternatively, based on the data in Figure 5, s_I may have power law scaling of the form of eq 3. Substitution of the power law scaled uncertainty functions for intensity yields

$$s_T = T\sqrt{\left(\frac{G^{\alpha-1}\beta\overline{I}^\alpha}{I}\right)^2 + \left(\frac{G^{\alpha-1}\beta\overline{I_0}^\alpha}{I_0}\right)^2} \quad \text{and}$$

$$s_A = 0.4343\sqrt{\left(\frac{G^{\alpha-1}\beta\overline{I}^\alpha}{I}\right)^2 + \left(\frac{G^{\alpha-1}\beta\overline{I_0}^\alpha}{I_0}\right)^2} \tag{10}$$

Terms can be added to eq 10 to account for additional sources of variation such as readout error, "flicker", positioning, and so on,[35,36]

$$s_T = \sqrt{\left(\frac{G^{\alpha-1}\beta\overline{I}^\alpha + s_d + ...}{I}\right)^2 + \left(\frac{G^{\alpha-1}\beta\overline{I_0}^\alpha + s_d + ...}{I_0}\right)^2} \quad \text{and}$$

$$s_A = 0.4343\sqrt{\left(\frac{G^{\alpha-1}\beta\overline{I}^\alpha + s_d + ...}{I}\right)^2 + \left(\frac{G^{\alpha-1}\beta\overline{I_0}^\alpha + s_d + ...}{I_0}\right)^2} \tag{11}$$

The utility of eqs 10 and 11 is to consider uncertainty functions based on non-Poisson processes (Figure 6, bottom panels). As α approaches 1 in eq 10, absorbance scaling approaches eq 2 and the position of the minimum PRSD depends on the value of α. When $\alpha = 1$, no minimum is seen. The photomultiplier tube data (Figures 4 and 5) exhibited α from 0 up to 0.76 (see Supporting Information for additional details) and a range of G_R and β_R values. Eqs 10 and 11 provide context for these values and can also be used to consider the effects of number squeezed light[37] on this type of measurement.

Measured Scaling. To assess measured scaling in absorbance measurements, a single solution was measured in the presence and absence of a ND4 filter to reduce the amount of light reaching the diode array (Figure 7, black symbols). Although

qualitatively the data follow the complex scaling trends of Figure 6, there is a marked additional upward curvature at high absorbance attributable to detector noise. Fits to eq 9 gave estimates of I_0 (7.1×10^7 photons) and s_d (equivalent to 3700 photons) and consideration of the derivative found α reaching values in excess of 4 above absorbance values of 2. Transmittance and absorbance values were also generated from the intensities in Figure 4 (Figure 7, blue symbols). In these results, the effects of the complex segmented fluctuation scaling of the photomultiplier tube are clearly seen. The minimum of this system is found near 1.5 absorbance units, well outside the range 0.43–0.86 typically assumed.

While the general shape of absorbance PRSD curves has been known for some time, detailed studies on modern instrumentation are limited and the effects of light sources with non-Poisson scaling have not been extensively modeled. Constructing fluctuation scaling plots provides a simple way to compare spectrophotometers and to specify methods depending on them. To a trained eye, Figure 7 (right panel) reveals the number of photons used, the flatness of the light source across the wavelengths assessed, where the best measurements can be made, and information about the dominant noise sources.

CONCLUSION

This study used fluctuation scaling to calibrate the uncertainty and interpret the gain, dispersion, and process parameters obtained from a set of measurement techniques which underpin nearly all of analytical chemistry. The reported calibrations reflect the specific conditions at the time of measurement and make a number of things clear. (1) Horwitz scaling parameters may be characteristic of interlaboratory comparisons but do not apply to any of the underpinning measurements considered here. (2) Simple or segmented power law scaling (eq 2) was observed in nearly all cases and showed how acquisition parameters and computations affect the dispersion and gain of measurements. (3) When more complex scaling is observed (e.g., absorbance) fitting the fluctuation scaling to uncertainty functions provides a rigorous way to compare the performance of instruments and methods. (4) There is no justification for a HorRat approach to the measurements presented here and application of the HorRat may be misleading in the absence of a full fluctuation scaling study in each lab of an interlaboratory comparison. The issues highlighted here suggest that current applications of this metric need to be reviewed. (5) This study and previous work make clear that chemically relevant measurements typically exhibit exponents over the range $0 \leq \alpha \leq 1$ (Gaussian to Gamma). Any

value in this range is valid and reflects the process underlying the measurement. In some cases, like sections of an absorbance curve, $\alpha > 1$ is also possible.

The data presented here was intralaboratory but provides a number of insights that may assist analysts, instrumentation providers, and participants in interlaboratory comparisons. Specifically, analysts, professional bodies, and instrumentation providers have broad understanding of the concept of detection limits. There is limited appreciation of fluctuation scaling and the concepts of gain, dispersion, and process. Detection limits have a statistical basis, however, a detection limit is of limited value compared to a fluctuation scaling investigation. The former provides information about a regime best avoided for routine analysis. The latter provides a clear picture of the expected statistics of an instrument or method over its full range of operation. As an extreme example, two instruments could have the same very low detection limit. If one of them followed gamma scaling, its PRSD would never change over its entire dynamic range. Such an instrument would be vastly inferior to another having Poisson or Gaussian scaling and arguably would never reach a quantification limit. Further, because fluctuation scaling behavior is not well appreciated, it is possible for manufacturers to provide error estimates in their software that are unrealistic. Manufacturers should be encouraged to report and specify the scaling behavior of their products and services.

Fluctuation scaling gives a compact presentation of instrument and measurement behavior allowing comparisons between instrument settings, analysts, manufacturers, providers of analytical services, and laboratories. For example, consider an analyst, instrument, or laboratory returning a result deemed an outlier. Making a decision based on a single point is not as valuable as the overall scaling behavior. Scaling can determine if that outlier arises from a different process ($\alpha_{observed} \neq \alpha_{expectation}$), one that is differently dispersed ($\beta_{observed} \neq \beta_{expectation}$), and one exhibiting unexplained gain ($G_R \neq 1$). The additional insight this provides the analytical community should fundamentally change the way we view our instrumentation, methods, and uncertainty estimates.

■ ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b02335.

A detailed treatment of the data found in Figure 5, including additional tests for mis-specification, segmented analysis of the data from each voltage, G, and β computations, and a discussion of the results (PDF).

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: quentin.hanley@ntu.ac.uk.

ORCID ⓞ

Quentin S. Hanley: 0000-0002-8189-9550

Notes

The author declares no competing financial interest.

■ REFERENCES

(1) Horwitz, W. *Anal. Chem.* **1982**, *54*, 67A−76A.
(2) Albert, R.; Horwitz, W. *Anal. Chem.* **1997**, *69*, 789−790.
(3) Hall, P.; Selinger, B. *Anal. Chem.* **1989**, *61*, 1465−1466.
(4) Horwitz, W.; Britton, P.; Chirtel, S. J. *J. AOAC Int.* **1997**, *81*, 1257−1265.
(5) McClure, F. D.; Lee, J.-K. *J. AOAC Int.* **2003**, *86*, 1056−1058.
(6) Horwitz, W.; Albert, R. *J. AOAC Int.* **2006**, *89*, 1095−1109.
(7) Thompson, M. *J. AOAC Int.* **2012**, *95*, 1803−1806.
(8) Coucke, W.; Charlier, C.; Lambert, W.; Martens, F.; Neels, H.; Tytgat, J.; Van de Walle, P.; Vanescote, A.; Wallemacq, P.; Wille, S. *Clin. Chem.* **2015**, *61*, 948−954.
(9) Student. *Biometrika* **1908**, *6*, 1−25.
(10) Dean, R.; Dixon, W. *Anal. Chem.* **1951**, *23*, 636−638.
(11) Grubbs, F. E. *Ann. Math. Stat.* **1950**, *21*, 27−58.
(12) Long, G. L.; Winefordner, J. D. *Anal. Chem.* **1983**, *55*, 712A−724A.
(13) Currie, L. A. *Anal. Chim. Acta* **1999**, *391*, 127−134.
(14) Nelson, L. *J. Qual. Technol.* **1984**, *16*, 238−239.
(15) Petri, G.; Expert, P.; Jensen, H. J.; Polak, J. W. *Sci. Rep.* **2013**, *3*, 1−8.
(16) Eisler, Z.; Bartos, I.; Kertész, J. *Adv. Phys.* **2008**, *57*, 89−142.
(17) Hanley, Q.; Khatun, S.; Yosef, A.; Dyer, R.-M. *PLoS One* **2014**, *9*, e109004.
(18) Keeling, M.; Grenfell, B. *Philos. Trans. R. Soc., B* **1999**, *354*, 769−776.
(19) Calif, R.; Schmitt, F. G. *Resources* **2015**, *4*, 787−795.
(20) Kendal, W. S.; Jørgensen, B. *Computation* **2015**, *3*, 528−540.
(21) Thompson, M. *TrAC, Trends Anal. Chem.* **2011**, *30*, 1168−1175.
(22) Rocke, D. M.; Lorenzato, S. *Technometrics* **1995**, *37*, 176−184.
(23) Thompson, M. *Analyst* **2000**, *125*, 385−386.
(24) Thompson, M. *Analyst* **1999**, *124*, 991−991.
(25) Kendal, W. S.; Jørgensen, B. *Phys. Rev. E* **2011**, *83*, 066115.
(26) Jørgensen, B.; Kokonendji, C. C. *AStA Adv. Stat. Anal.* **2016**, *100*, 43−78.
(27) Kendal, W. S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 837−841.
(28) Kodama, T.; Kurosawa, Y.; Kitta, K.; Naito, S. *J. AOAC Int.* **2010**, *93*, 734−749.
(29) Cohen, J. E.; Xu, M.; Schuster, W. S. *Proc. R. Soc. London, Ser. B* **2013**, *109*, 15829−15834.
(30) Taylor, L. *Nature* **1961**, *189*, 732−735.
(31) El-Shaarawi, A. H.; Zhu, R.; Joe, H. *Environmetrics* **2011**, *22*, 152−164.
(32) Mortara, L.; Fowler, A. *Proc. SPIE* **1981**, *0290*, 28−33.
(33) Sperline, R. P.; Knight, A. K.; Gresham, C. A.; Koppenaal, D. W.; Hieftje, G. M.; Denton, M. *Appl. Spectrosc.* **2005**, *59*, 1315−1323.
(34) Laqua, K.; Melhuish, W.; Zander, M. *Pure Appl. Chem.* **1988**, *60*, 1449−1460.
(35) Rothman, L.; Crouch, S.; Ingle, J., Jr *Anal. Chem.* **1975**, *47*, 1226−1233.
(36) Galbán, J.; Marcos, S. d.; Sanz, I.; Ubide, C.; Zuriarrain, J. *Anal. Chem.* **2007**, *79*, 4763−4767.
(37) Vahlbruch, H.; Mehmet, M.; Chelkowski, S.; Hage, B.; Franzen, A.; Lastzka, N.; Gossler, S.; Danzmann, K.; Schnabel, R. *Phys. Rev. Lett.* **2008**, *100*, 033602.