

Statistical Analysis of Physical Chemistry Data: Errors Are Not Mistakes

Cite This: *J. Phys. Chem. A* 2020, 124, 2109–2112

Read Online

ACCESS |

Metrics & More

Article Recommendations

The goal of a physical chemistry experiment is to measure an interesting property. This property could be the lifetime of an excited electronic state, the exciton diffusion length within a solar cell, or the coupling strength between two optical transitions. In reporting the results of experiments it is extremely important to quantify the accuracy of the measurements. Errors are inevitable in experiments; however, they can be hard to work out. This is especially true when complex fitting functions are used to analyze the data, or when the data are compared to simulations. Nevertheless, a good analysis of errors can elevate the quality of your paper and help convince readers that the data have been carefully analyzed and are reliable.

The goal of this editorial is to provide a basic practical guide for reporting errors in physical chemistry experiments. Unlike experiments in biological sciences which involve controls and sometimes complex hypothesis testing, errors in physical chemistry are relatively straightforward. There are two common situations: (i) the case where a series of experimental measurements are averaged together to determine an expectation value and (ii) curve fitting where experimental data are fit to a function. In both cases the goal is to provide a best estimate for the quantity being measured, as well as an estimate of the range of possible values. This discussion will assume that we are only dealing with random errors; that is, that there are no systematic errors in the measurements.

Averages and Standard Deviations. When a series of nominally identical measurements have been performed, the best estimate of the true value of the quantity being measured is the sample mean $\bar{x}$, and the range of the values obtained from the experiments is characterized by the standard deviation σ_x :

$$\bar{x} = \sum_i x_i / N \quad (1a)$$

$$\sigma_x = \sqrt{\sum_i (x_i - \bar{x})^2 / (N - 1)} \quad (1b)$$

where x_i are the values from the individual measurements and N is the total number of measurements. The standard deviation tells us the range of values we expect in a measurement, but it is not the error in $\bar{x}$. The uncertainty in $\bar{x}$ depends on the number of measurements and is estimated by the standard error (also referred to as the standard deviation of the mean) $\sigma_x / \sqrt{N}$. For a large number of measurements and random errors, there is a 68% chance that the true value of the

quantity you are trying to measure lies in the range $\bar{x} \pm \sigma_x / \sqrt{N}$. The 95% confidence limits for the measurements (which is what should really be reported) is $\pm 1.96 \sigma_x / \sqrt{N}$. It is important to note that the numerical factor in front of σ_x for the 95% confidence limits depend on the number of measurements. For example, the 95% confidence limit for six measurements is $\pm 2.57 \sigma_x / \sqrt{6}$. The values for this numerical factor (which are obtained from the Student's t -distribution) are given in Table 1 for a range of N values (different degrees-

Table 1. Values of Numerical Factor t for the 95% Confidence Limit ($t\sigma_x / \sqrt{N}$) in Sample Averages versus the Number of Degrees of Freedom ($N - 1$)

$N - 1$	3	5	7	10	20	∞
t	3.18	2.57	2.36	2.23	2.09	1.96

of-freedom). This table shows that there is a significant advantage to performing more than five measurements in terms of reducing uncertainty. Also, note that for $N < 10$ the value of $\sqrt{N}$ is similar to the numerical factor from the Student's t -distribution, so that the 95% confidence limits are close to the standard deviations.

Figure 1 shows a plot of some experimental data obtained in the author's laboratory (measurements of the quality factors for the breathing vibrational modes of single gold nanowires), along with a table of the quality factors and the errors. The average value of the quality factor is $\langle Q \rangle = \sum_i Q_i / 16 = 62.0$ and the standard deviation is $\sigma_Q = \sqrt{\sum (Q_i - \langle Q \rangle)^2 / (15)} = 15.9$. This gives 95% confidence limits for the data of $\pm t\sigma_Q / \sqrt{N} = \pm 2.13 \times \frac{15.9}{\sqrt{16}} = \pm 8.5$. Thus, the average quality factor should be reported as $\langle Q \rangle = 62 \pm 8$, where the error is the 95% confidence limit. The horizontal line in the graph shows the average value of the measurements, and the shaded area corresponds to the standard deviation for the data. The standard deviation is similar to the error bars for the measurements. This is intuitive as the spread of measured Q values arises from the errors in the experiments. Note that

Published: March 19, 2020

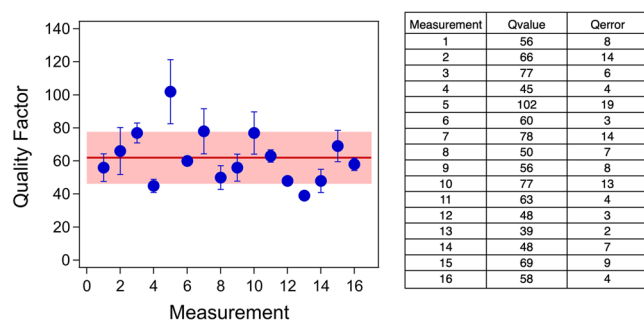


Figure 1. Measured quality factors (Q) and errors in Q for the breathing modes of suspended Au nanowires. The red line in the figure shows the average Q value, and the shaded area shows the standard deviation ($\langle Q \rangle \pm \sigma_Q$).

because there is a reasonably large number of measurements, the 95% confidence limit for $\langle Q \rangle$ is significantly smaller than the standard deviation. Also, because the spread in the Q values arises from the error in the individual measurements, there is no need to include the errors for the individual measurements in the final estimate of the error for $\langle Q \rangle$ —this would be overcounting. However, if the errors for the individual measurements were much larger than the standard deviation, then it would be better to use the errors for the individual measurements to determine the final error.

Curve Fitting. A common situation in physical chemistry studies is to fit the results of an experiment to a function. An example is shown in Figure 2. In this plot a signal versus time

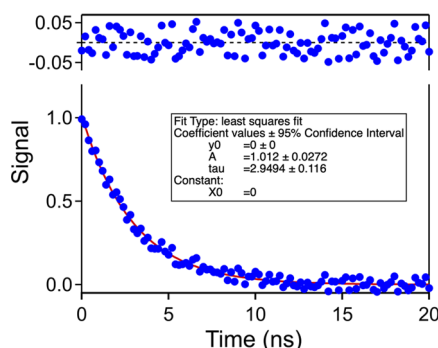


Figure 2. Simulated lifetime data (blue markers) and single exponential fit (red line). The inset shows the results from the least-squares fit to the data, and the residuals are shown at the top of the figure.

trace for a single exponential has been simulated with a lifetime of 3 ns and 10% added noise (a random number between -0.05 and $+0.05$). The trace has been fit to a single exponential function $y = y_0 + Ae^{-(x-x_0)/\tau}$ using a commercial data analysis software program. There are a number of programs with excellent built-in least-squares fitting routines that report expectation values along with error estimates. The results from the fit are shown in the inset (in this fit y_0 and x_0 were held constant), and a plot of the residuals (data minus the fit function) is shown above the graph (plotting residuals can be a good way of quickly checking if your model is a good match for the experiments). The least-squares fitting routine spits out more significant figures than is justified by the error. For this example, the lifetime should be reported as $\tau = 2.9 \pm 0.1$ ns. It is common in some fields to report two significant

figures for errors (that is, $\tau = 2.95 \pm 0.12$ ns); however, this is simply wishful thinking and is not justified from the data analysis!

Figure 3 shows simulated Stern–Volmer data (emission from a sample with different amounts of added quencher) that

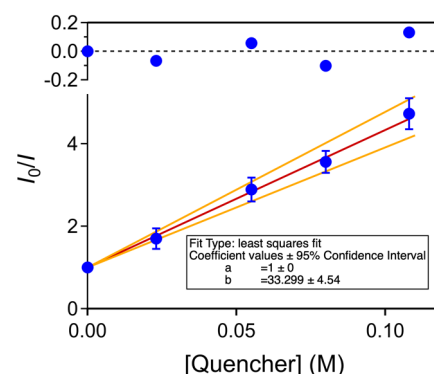


Figure 3. Simulated Stern–Volmer plot. The blue markers are the data, and the red line is a fit to the data using the Stern–Volmer equation.

have been fit to a linear function $y = a + bx$. The Stern–Volmer equation is $I_0/I = 1 + K_{SV}[Q]$, where I is the emission intensity, K_{SV} is the Stern–Volmer constant, and $[Q]$ is the concentration of quencher. Thus, the value of a in the fitting function was constrained to $a = 1$ (allowing a to vary would naturally give a better fit to the data but would not give the correct value for the Stern–Volmer constant). The Stern–Volmer constant from the fit is $K_{SV} = 33 \pm 4 \text{ M}^{-1}$. However, the error estimates determined from the fit do not account for the error bars on the data. A simple way to include the uncertainty from the error in the individual points is to draw lines through the data points that give estimates of the maximum and minimum values of K_{SV} consistent with the error bars. This is shown as the orange lines in Figure 2. This analysis gives a more conservative (and safer) estimate of the error of $K_{SV} = 33 \pm 5 \text{ M}^{-1}$.

Error Propagation. In many cases the results from different experiments are combined together. For example, in discussing resonances it is common to talk about quality factors $Q = \omega/\Gamma$, where ω is the resonance frequency and Γ is the line width (see Figure 1 above). When two quantities are added or subtracted together, then we add the errors using a quadratic sum. In contrast, if two quantities are multiplied or divided, then we add the relative errors. For more complex functions partial derivatives have to be used.

$$f = a \pm b \Rightarrow \delta f = \sqrt{(\delta a)^2 + (\delta b)^2} \quad (2a)$$

$$f = a \times b \text{ or } f = \frac{a}{b} \Rightarrow |\delta f/f| = \sqrt{(\delta a/a)^2 + (\delta b/b)^2} \quad (2b)$$

$$f = f(x, \dots, z) \Rightarrow \delta f = \sqrt{(\partial f/\partial x)^2(\delta x)^2 + \dots + (\partial f/\partial z)^2(\delta z)^2} \quad (2c)$$

where $\partial f/\partial x$ and $\partial f/\partial z$ are the partial derivatives of f with respect to the different measured quantities ($x, \dots, z$).

To give a concrete example of error propagation, Figure 4 shows a simulated time-resolved experiment (similar to Figure 2) that has a double exponential decay (exponentials with time

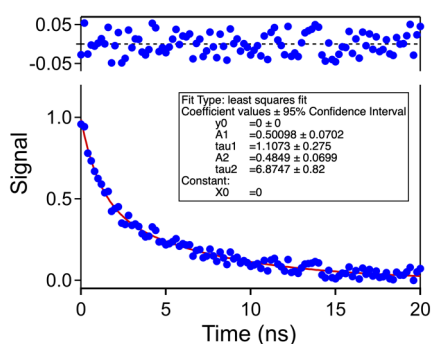


Figure 4. Simulated lifetime data (blue markers) and double exponential fit (red line). The inset shows the results from the least-squares fit to the data.

constants of 1 and 7 ns with equal amplitudes) with 10% added noise. The data were fit to a double exponential decay function $y = y_0 + A_1 e^{-(x-x_0)/\tau_1} + A_2 e^{-(x-x_0)/\tau_2}$ where y_0 and x_0 were fixed to zero. A least-squares fit to the data gives values of the amplitudes and time constants of $A_1 = 0.50 \pm 0.07$, $\tau_1 = 1.1 \pm 0.3$ ns and $A_2 = 0.48 \pm 0.07$, $\tau_2 = 6.9 \pm 0.8$ ns. Now let us say you have a number of these traces and you want to compare them. One way to do this is to calculate an average lifetime by $\langle \tau \rangle = (A_1 \times \tau_1 + A_2 \times \tau_2) / (A_1 + A_2) = 3.9$ ns. The error for this quantity is determined by the errors in both the amplitudes and the lifetimes (the amplitude part is often ignored in analysis of time-resolved measurements). Working out the partial derivatives and adding the errors according to the formula above gives $\langle \tau \rangle = 3.9 \pm 0.5$ ns. Note that if we had only used the error in the lifetimes, then our error would be ± 0.4 ns, which is actually close to the correct answer.

Sensitivity Analysis. In some cases the standard least-squares routines in data analysis programs cannot be used to fit the data. Examples include when nonanalytic functions are used to model the experiments, or where the experiments are being compared to simulations. Figure 5 shows an example of

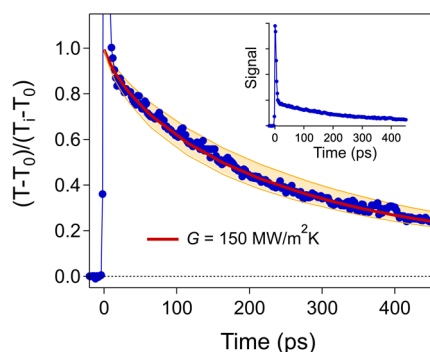


Figure 5. Experimental measurements (markers) and simulations (lines) of heat dissipation for 20 nm radius Au nanoparticles in water. The inset shows the full transient absorption trace.

the latter case. The experimental data are a transient absorption trace for 20 nm radius Au nanoparticles in water. The slowly decaying part of the trace is due to heat dissipation from the nanoparticles into the environment. The lines in Figure 5 are results from finite element simulations of the heat dissipation process, which require the heat capacity (C_p), density (ρ), and thermal conductivity (κ) of the different materials, and the thermal interface conductance (G) as input

parameters. For Au and H_2O C_p , ρ , and κ are well-known; however, G is not and determining a value for G is the object of the experiment. The finite element simulations were performed with different values of G , and the results were compared to the experimental data. The best fit to the data is obtained for $G = 150 \text{ MW/(m}^2\text{ K)}$, and the data are bounded on the low and high sides by $G = 100 \text{ MW/(m}^2\text{ K)}$ and $G = 220 \text{ MW/(m}^2\text{ K)}$, respectively (shown as the shaded area on the graph). Thus, the best estimate of G , and the range of G values that are consistent with the data, is $G = 150 + 70/-50 \text{ MW/(m}^2\text{ K)}$. Note that it is hard to be precise about the type of error here, so it is probably best to be conservative in the estimate of the error range. Also, Figure 5 is a fairly simple example because there is only one parameter being varied. In cases where there is more than one parameter in the analysis, all the parameters have to be varied to determine a reasonable range of values that describe the experiments. Unfortunately, the parameters in this type of analysis are often correlated, so determining a unique set of parameters with errors can be challenging.

Best Practices. The error that should be reported depends on the situation. For example, in electron microscopy analysis of nanoparticles the average dimensions of the materials should be reported with standard deviations (as this characterizes the range of sizes, which is what is important). For solar or electrochemical cells (or other types of devices) the average values of the cell parameters should be reported with 95% confidence limits. The average values and confidence limits should be determined by measuring the properties of a number of nominally identical cells, and the number of cells examined should be given. For quantities obtained from fitting experimental data, use the errors generated by the fitting program (if your program does not give error estimates, use a different program). You should always state the type of error being reported (standard error, 95% confidence limit, standard deviation, etc.). For example, for Figure 2 the result could be reported as “ $\tau = 2.9 \pm 0.1$ ns (error equals 95% confidence limit).” If you choose not to use 95% confidence limits, then it is very important to give the number of measurements, so that the reader can work out the 95% confidence limits.

Tables of experimental data should always have errors. If all the errors are the same (e.g., they are all 5%), it is acceptable to state the errors in the table caption, rather than listing them in the table. Likewise, plots of experimental data should have error bars. Note that error bars are not typically needed for primary data, such as spectra, transient absorption traces, current versus voltage, or magnetization versus field measurements. Because the focus in a graph is usually on the dependent quantity, it is essential to have y-axis error bars. However, error bars may also be needed for the x-axis. An example is in nanomaterials studies, when properties are plotted against sample dimensions. When the error bar is smaller than the size of the symbol in the graph, this should be stated in the figure caption.

Error Pitfalls. There are a number of common mistakes that should be avoided in reporting errors. First, avoid reporting too many significant figures. In Figures 2–4 above, the insets that contain the analysis results are fine for the Supporting Information but should not be included in the main text. In the main text just give the key result with the error and the correct number of significant figures. Second, R^2 and χ^2 values are not errors. These quantities give information about the suitability of the model used to fit the data, and

whether the number of parameters used in the model is justified. However, they do not directly give the uncertainty in the parameters being measured.

Sometimes least-squares fitting routines give unreasonably small errors (for example, time constants in transient absorption traces with errors of a few femtoseconds). If the errors look too small, then you should perform a sensitivity analysis to determine the uncertainty (vary the parameter and see what range of values gives a reasonable fit to the data). Likewise if the errors are very large, then the data are not very sensitive to that parameter, and you may be able to fit the data without the parameter. In time-resolved experiments you should not report time constants smaller than the instrument response function, unless you know the instrument response function very well and have deconvoluted it from the signal. For multiexponential decays data should be recorded over several lifetimes to accurately determine the time constants. And even though errors can be obtained from the fitting programs from these experiments, the measurements should be always be repeated!

The examples discussed above are very basic. There are a number of textbooks on error analysis that provide a more sophisticated description of how to statistically analyze data—for example, how to check for correlations and test for significant differences between measurements. At the end of the analysis, it is important to perform a qualitative check on the error being reported. If the error appears to be too large or too small, then there may be a mistake or inaccuracy, and the assumptions behind the fitting or model may need to be questioned. Some reporting tips are listed in Table 2.

Table 2. Some Tips for Reporting Errors in Physical Chemistry Experiments

best practices for reporting errors
The number of significant figures should match the error.
State the type of error being reported.
Report the standard deviations for the dimensions of nanomaterials.
Include all sources of uncertainty in error propagation.
Know the limitations of your instrument.
Give the sample size when reporting mean values and errors.

Further reading:

Lyons, L. *A Practical Guide to Data Analysis for Physical Science Students*; Cambridge University Press: Cambridge, 1991.

Taylor J. R. *An Introduction to Error Analysis*; University Science Books: Sausalito, CA, 1997.

Gregory V. Hartland, Senior Editor  orcid.org/0000-0002-8650-6891

AUTHOR INFORMATION

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpca.0c01403>

Notes

The author declares no competing financial interest.

Views expressed in this Viewpoint are those of the author and not necessarily the views of the ACS.

This Viewpoint is jointly published in *The Journal of Physical Chemistry A*, *B*, and *C*.