Global Environmental Operations, Inc.

Better Statistics mean better permits

By

David L Russell, PE

When we have an effluent do we really know what the discharge is, or are we simply making a best guess, or do we know how our permit limits are set? In the US, the general structure of an discharge permit is that it has a few critical numbers: Daily Maximum, Monthly Average,

Under the Clean Water Act's latest revisions, the penalties for non-compliance can be substantial. Under 33USC1319, when the Administrator of the EPA finds violations of Section 1311, 1312, 1316-1318, 1328, or 1345 relating to violation of permit conditions the first step is an Administrative Order or Notice of Violation. The AO or NOV takes effect only after the RA has had the opportunity to confer with the individual. Then the potential fines, can be $2,500 to $25,000 per day of violation or a year in jail or both, with each day being a separate offense in civil or criminal actions.

Statistical Basis for Permitting

It is important to determine your permit limits at the time the permit is issued. The type of statistical distribution applied in setting the permit limits should be identified and agreed upon.

The maximum amount of contaminants you can legally discharge are generally set by water quality standards. The minimum discharge may not be zero even when the facility is shut down, due to environmental and non-process related factors[i].

Municipal and industrial treatment systems can be upset by conditions beyond the Operator's control, and these minor and major upsets can cause permit violations. The maximum amount

4797 Dean Lane, SW. Lilburn, GA, 30047 (770) 923-4408; cel: 404-667-2427

E-mail: dlr@mindspring.com

of contaminants discharged is limited only by the size of the event causing it, the type of contaminant, the effect on the wastewater treatment system, and environmental factors (rainfall, snowmelt, firewater discharges, etc.) and the Operator's ability to intercept and stop the event.

Some wastewater discharge permits, are based upon a lognormal distribution[ii] for toxics, and a Student's T or Normal distribution for "conventional" contaminants. Generally, the daily average discharge is a factor of 2 times the monthly average. With a standard normal distribution[iii] for conventional pollutants (BOD_5, TSS, NO_3, etc), the daily maximum is set at two times the standard deviation on either side of the mean (monthly average) value. The 2 times standard deviation is supposed to represent 95.5% of the possible values of the population of possible discharges. But is this realistic?

We constructed a sample population of 51 random numbers between 5 and 60 in Excel® as a sample of an waste treatment plant discharge. The ranked and ordered data are shown in Figure 1. The average of the discharge is 30.73 mg/l. The standard deviation of the population of discharges is 15.85 mg/l. That means that 95.5% of all possible discharges from a facility would be between -0.975mg/l and 62.44mg/l. The upper and lower 95.5% confidence intervals are outside the range of the population. So, what's happening?

The idea is that over time an equal number of data points will fall above and below the mean, but they won't. This distribution violates rationality, because no facility can have a negative discharge value and discharge values significantly below the mean are rare, and some operators and engineers would say highly improbable.

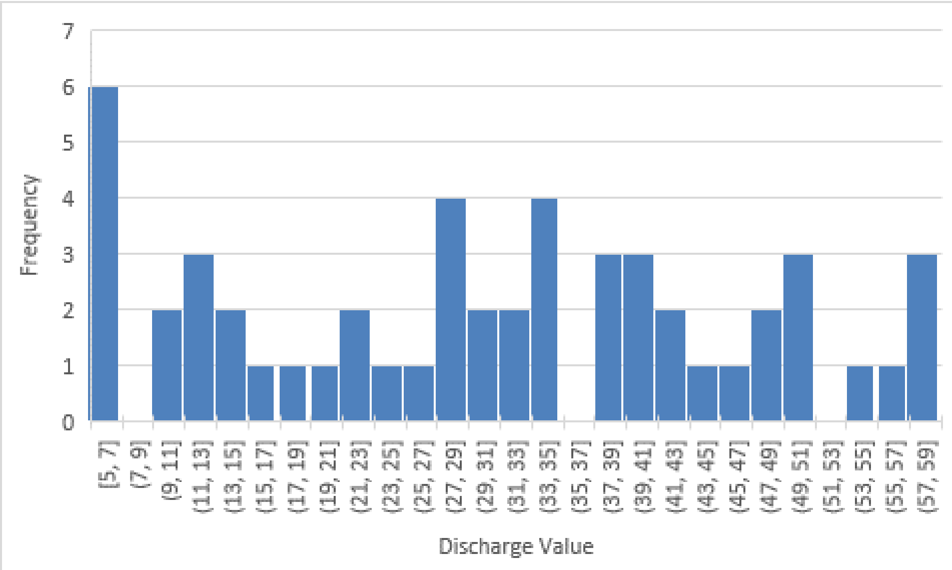

Figure 1. Frequency of 51 random numbers between 5 and 60.

A wastewater treatment plant lower discharge value may be close to zero but the maximum discharge value is most often beyond the operator's control and depend on outside influences such as spills, rainfall, industrial chemicals in the influent, etc.

A lognormal analysis of the effluent data shows us a slightly better situation. The lognormal distribution is prepared using normal distribution procedures, with the exception that the Log10(x) or ln(x) is used rather than x in the population construction. The lognormal distribution has the advantage that it has no negative values.

A lognormal distribution also has an fat upside tail, meaning a more forgiving and larger daily maximum discharge value. This is also true of a Weibull distribution. The probability function for a lognormal distribution is shown in eq. 1

$eq.1 \quad X = e^{(\mu+\sigma Z)}$ Lognormal Distribution equation

Where Z is a standard normal distribution variable, and σ and μ are the standard distribution and the mean of the population of the variable's natural logarithm. The parameters of the lognormal distribution can be evaluated once σ and μ are known. These are often referred to as the scale and location parameters. We ran a lognormal analysis our sample population, and found that the scale parameter was 0.6992 (±0.2812), and the location parameter was 3.2335 (±0.3893) from that, one can construct the Probability Density Function and Cumulative Density Function for the sample. The distribution has a correlation coefficient (r^2) which varies between 0.96 at lower values to 0.64 at higher values.

Weibull Distributions

The Weibull distribution is shown in Equation 2, and while it looks formidable, it is very easy to compute:

Eq.2: $f(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-\left[\frac{x}{\alpha}\right]^{\beta})$

When the shape parameter β is above 1 the curve begins to look like a standard distribution. At β = 3.44 the distribution looks like a lognormal distribution. Values above 4 start to indicate a specific type of material or equipment failure. The shape parameter of the distribution can provide information about failure modes, and it is widely used in manufacturing for evaluation of failures and preventive maintenance on process equipment, bearings, to pumps and compressors. One could argue that treatment plant systems depend upon the reliability of their mechanical equipment and thus a Weibull analysis is relevant.

A Weibull distribution is easy to compute, and has a longer right tail as well, and fits the data. It could be an ideal way to establish permit limits when a facility is in startup or when it is running. A Weibull analysis will simulate normal, beta, lognormal, and Weibull distributions based on your data. Inexpensive statistical programs.[iv][v] and spreadsheets can be used to calculate a Weibull distribution. There are also a number of websites and papers which discuss the application of the Weibull distribution.[vi]

Weibull Distribution family shows various Values of Shape factor β
Source: https://applicationsresearch.com/WeibullEase.htm

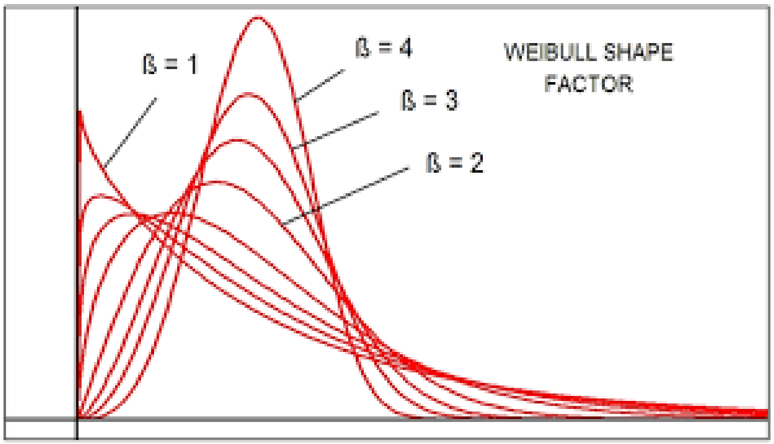

Figure 2 Illustrates the different type of curve shapes obtained through Weibull analysis.

Comparing the Distributions:

For our example the Weibull distribution determined the following coefficients:

β = shape = 2.010, α = scale = 34.61

Correlation Coefficient for the distribution R^2 = 0.9531 varies from 0.89 at lower values to 0.97 at higher values, which is better than the lognormal distribution.

The goodness of fit of the Weibull and lognormal distributions is illustrated by the linearity of the data points in the Quartile-Quartile (QQ) plots shown below. A straight line has been drawn through each of the plots for illustration.

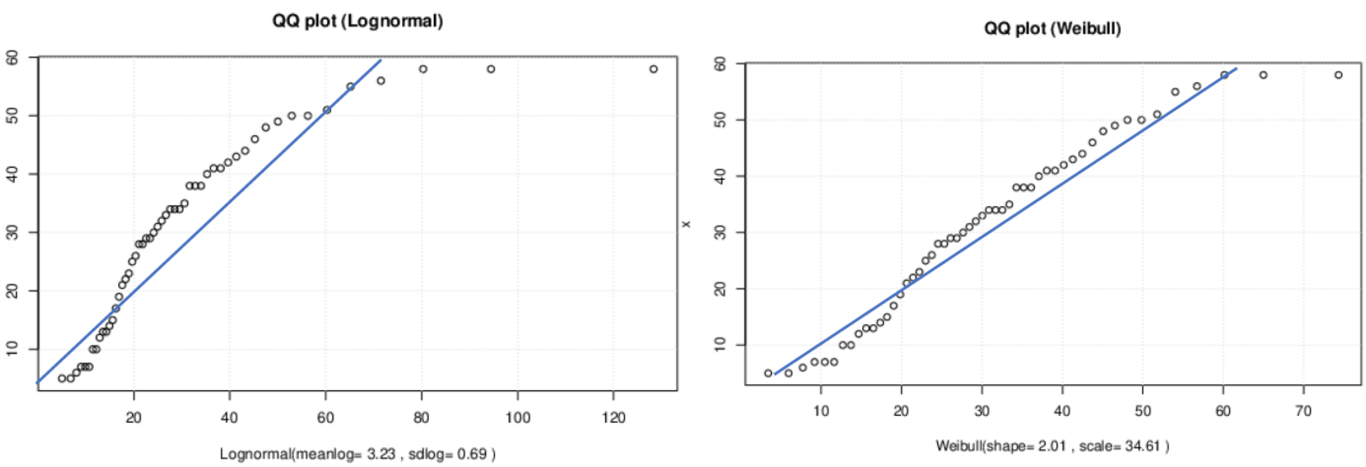

Figure 3: Comparison of Weibull and Lognormal distributions

Creating a Weibull Distribution

There are two easy procedures for creating a Weibull Distribution on your data using a spreadsheet. The first involves using the spreadsheet Weibull distribution function on the spreadsheet itself. Enter the data in column form, arrange the data from smallest to largest values and specify the range on the built in software and let the software do the work.

A second procedure requires slightly more work but you have control and can select datapoints to remove zeros.

Start by sequentially numbering the number of data points you are to put in Call this Column A. If you have 25 data points, n goes from 1-25

Next enter your data in a column we call Z

Next rank the data (Z) from smallest to largest. Note that you may have to copy the data using the Paste Values command

Create a Column F which will calculate the rank of Z by computing individual numbers from A into Column F by the following formula . Note that "a" is a rank number in column A.

F =(a-0.5)/(n)

This gives you values for F -- [the -0.5 prevents you from topping out the scale, and can be omitted for very large data sets].

Label a column X then take natual log of values of Z placing them into Column X

Create a column Y which contains natural log of the natural log of (1/(1-F) formula is

Y =nl(nl(1/(1-F)) for all values

Plot Y vs X, and adjust the graph for clarity. You want values of the natural log of Z as the X axis. You will come up with a series of points on an graph which will have some negative values on the y axis.

Use the software to plot a straight line through your data and give you the equation for the line and the correlation coefficient. It will be in the form of y= mx +b. (b will probably be negative but that's all right)

m is the Weibull shape parameter called β. Now compute η which is obtained by solving η= exp(-b/m) where exp is base e to get the other parameter. Congratulations, you have created a Weibull plot which is mathematically rigorous[vii].

For the sake of convenience I've shown the calculation of a Weibull Distribution using a spreadsheet in Figure 4.

Weibull Plot Example in Excel

Rank	Data	Ordered			
A	Z		F	X	Y
1	5	5	0.0714	1.609	-2.602
2	20	20	0.2143	2.996	-1.422
3	35	28	0.3571	3.332	-0.817
4	28	35	0.5000	3.555	-0.367
5	45	40	0.6429	3.689	0.0292
6	58	45	0.7857	3.807	0.4321
7	40	58	0.9286	4.06	0.9704
	Shape = 1.4122		Scale	39.44	

Figure 4. Calculation of a Weibull Distribution

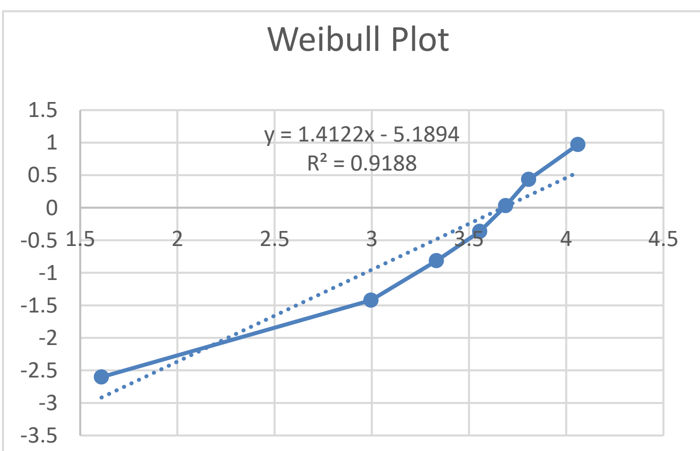

Figure 5 The Correlation Plot in Excel® used to create the β and α values for the distribution.

Note that this was done with an extremely limited data set, and a larger data set could provide different and more accurate values. There are other sources of Weibull distribution calculations on the World Wide Web, including one which calculates the beta and eta values for you: http://www.wessa.net/rwasp_fitdistrweibull.wasp.

Probable Errors and Permits

Another challenge is to make accurate measurements regarding our discharges—to find out what we are actually discharging, and to be as accurate as reasonably practical with respect to the discharge.

The discharge is mass based, and the formula is:

Discharge= Concentration * Flow

If we have any uncertainties in the measurement of any of these parameters, we have a permit uncertainty or probable error. Sometimes, the uncertainty can be as large the measured value. The probable error in a measurement is composed of individual and independent functions, f1…..fn, and the measurement M is composed of the functions as follows:

M = f1(x) + f2(x) + f3(x)… fn(x)

And the total probable error in M is given by the following formula:

$$e^2 = (e_1 \mathrm{d}x/\mathrm{d}f_1)^2 + (e_2 \mathrm{d}x/\mathrm{d}f_2)^2 + \cdots$$

Where e_n is the error of measurement in the individual parameter.

Most flowmeters including magmeters, and doppler flowmeters have an accuracy to within ±2% depending upon the meter and design. Weirs and open channel flowmeter can have an probable error of up to 10% depending upon installation conditions.

For a specific case, we can estimate the accuracy of the permit measurement. Our suspended solids concentration is 10 mg/l. The flow is 11880 $M^3 day^{-1}$ measured by our flowmeter. The published accuracy for Suspended Solids test according to Standard Methods[viii] is 15%. The measured discharge 118.8 kg day^{-1}

If the flowmeter is accurate to within about 2% and the laboratory is accurate to within 15% on the particular test, then the relative error is

e = square root of ($0.15^2 + 0.02^2$) or e= 0.15132 or 15.13% At the specified conditions, the error could put your discharge anywhere between 100.1 and 135.85 kg d^{-1}. Similarly if you are close to your maximum permit value, you could be over or under, but you don't know.

The value of the error will vary with the accuracy of each of the individual parameters, and that error can be reduced by multiple sampling or by detailed analysis using the partial differentials above. Multiple analysis of the TSS will have a great impact because the probable error is reduced by the square root of the number of analyses.

SAMPLING ERRORS

There are a number of other errors in sampling, and most of them depend upon the sampler. We assume that the samples we collect are representative but they may not be. The EPA attempted to address the issue of sampling devices and accuracy. Their efforts indicated that the ratio of the composite sample concentration to actual concentrations could vary between a low of 68% to a high of 135% depending upon the sampler and flow conditions[ix] Most of the samplers averaged between 90 and 99 accuracy as shown below:

RATIO OF COMPOSITE SAMPLE CONCENTRATION TO ACTUAL CONCENTRATION

q FLOW \ CONC k	$1-t$	$1-\frac{t}{2}$	$\cos\frac{\pi t}{2}$	e^{-t}	$\sin\pi t$
c	0.90	0.97	0.92	0.95	0.99
	0.90	0.97	0.92	0.95	0.99
	0.90	0.97	0.92	0.95	0.99
	0.90	0.97	0.92	0.95	0.99
t	1.35	1.09	1.26	1.14	0.99
	0.90	0.97	0.90	0.97	0.90
	0.86	0.96	0.87	0.95	0.89
	0.87	0.96	0.89	0.95	0.97
1-t	0.68	0.87	0.72	0.82	0.99
	0.95	0.98	0.98	0.96	1.12
	0.92	0.97	0.95	0.95	1.09
	0.92	0.97	0.93	0.95	0.97
sinπt	0.90	0.97	0.88	0.97	0.80
	1.01	1.00	1.00	1.00	1.01
	0.90	0.97	0.92	0.95	0.98
	0.90	0.97	0.92	0.95	0.97

Line 1. T_cV_c – Simple composite

Line 2. T_cV_v – Volume proportional to flow rate (q)

Line 3. T_cV_v – Volume proportional to flow (Q) since last sample

Line 4. T_vV_c – Time varied to give constant ΔQ

REPEAT MEASUREMENTS:

The US Bureau of Reclamation's publication: Water Measurement Manual, is available on-line[x] and is an excellent reference for flow measurement. It should be a part of every environmental engineer's technical library.

Accurate discharge measurement is a science and needs to be approached with accuracy and caution.

More information and a paper on sampling is available in the downloads section of my website: http://www.globalenvironmental.biz

Dave Russell

PS:

I have a new book coming out in April 2019 which addresses some of the same materials covered here, and also details what you need to know about equipment and your designs.

Practical Wastewater Treatment 2nd Edition is published by John Wiley & Sons. And it is 400 pages of all that you really wanted to know about the practical considerations of wastewater, groundwater, and industrial process water treatment. Also available on Amazon.com

---

[i] Some of those factors might include groundwater, or contaminated stormwater flowing through the plant or into the plant sewer system. Note that this is not strictly industrial, but may include municipal systems as well.

[ii] https://www.epa.gov/sites/production/files/2015-09/documents/pwm_2010.pdf

and you need to see the Technical support Document of Water Quality Based Toxics Control https://www3.epa.gov/npdes/pubs/owm0264.pdf and www.cormix.info The latter of which is the EPA's expensive model for calculating water quality in mixing zones.

[iii] There is a difference between the Student's T and the Normal distributions, but because the differences are minor, we have chosen to equate them as interchangable.

[iv] In construction of this article, the author used the following programs:

On Youtube.com, under the heading of Weibull Analysis there are a number of helpful websites.

One of many Statistical Programs : Trial Version of MaxStat lite with a free limited duration trial. https://maxstat.de/en/home-en/. The program cost is about $100. A free website for analyses is: https://wessa.net = look for the number of distributions available.

[v] See very good, well written articles by P Barringer: at http://www.barringer1.com/pdf/Chpt1-5th-edition.pdf and https://reliabilityweb.com/articles/entry/a_guide_for_using_the_weibull_distribution_in_failure_mode_analysis/

[vi] http://www.weibull.com is run by Reliasoft which is dedicated to reliability analysis, on the Web, one can look at https://Ncalculators.com/math-worksheets/Weibull-distribution-example.htm and a number of other websites.

[vii] There are a number of mathematical justifications of the procedure, including some papers on Weibull analysis from https://www.slideshare.net/melvincarter/using-microsoft-excel-for-weibull-analysis, several helpful procedure videos on www.youtube.com under the heading Weibull Analysis, and the Barringer paper cited above.

[viii] See Standard Methods for the Examination of Water and Wastewater under Solids. Published by APHA, ASCE, and WEF.

[ix] EPA-600/2-75-065, December 1975, An assessment of Automatic Sewer Flow Samplers and EPA-600/4-82-029 September Handbook for Sampling and Sample Preservation of Water and Wastewater.

[x] https://www.usbr.gov/tsc/techreferences/mands/wmm/index.htm