Quantifying Sterilizing Membrane Retention Performance

Mark M. Blanchard

In aseptic processing, performance of sterilizing filtration is of critical importance. A filter device must provide high confidence of retaining organisms under production conditions as well as in worst-case validation scenarios. To achieve this performance, it is necessary to understand and quantify retention capabilities of sterilizing grade membranes. Such understanding needs to begin during design of the membrane used in a filtration device. To date, manufacturers have provided this information descriptively or semiquantitatively. The level of confidence or assurance is left to a user's interpretation. Here, I provide a framework for quantifying membrane retention performance when sterile filtrate is required.

In membrane design, the relationship between retention performance and a nondestructive physical measurement has been a key starting point for quantification. This topic dates back at least as far as the work of Johnston and Meltzer (**1**).

PRODUCT FOCUS: PROTEINS, ANTIBODIES, PARENTERAL PRODUCTS

PROCESS FOCUS: ASEPTIC PROCESSING

WHO SHOULD READ: PROCESS DEVELOPMENT AND MANUFACTURING

KEYWORDS: STERILIZING-GRADE MEMBRANES, RETENTION, FILTRATION, BUBBLE POINT, VALIDATION

LEVEL: INTERMEDIATE

They showed that a log reduction value (LRV = logarithm of upstream quantity divided by downstream quantity) of a given bacterial challenge is correlated with membrane maximum pore size as measured by bubble point (BP). These results were used to define a membrane as "absolutely retentive" if it retained a challenge of $>10^7$ colony forming units per square centimeter of membrane area (cfu/cm^2) of *Pseudomonas diminuta* (now *Brevundimonas diminuta*).

This bubble point and LRV relationship has been widely used. For example, ASTM Method F838-05 (**2**) includes LRV as the measure of filter performance. Authors such as Meltzer and Jornitz (**3**) cite the relationship between LRV and BP and suggest that in practice it is approximately linear. Strictly speaking, LRV itself does not directly assure that sterile effluent is achieved in a filtration process. One must infer that a sufficiently high LRV is equivalent to having sterile filtrate.

FDA guidelines state that a sterilizing grade filter "reproducibly removes all microorganisms from the process stream, producing a sterile effluent" (**4**). This moves the objective from a high LRV to sterile filtrate (zero observed organisms), but again it does not quantify the assurance level with which this is achieved. Also left open is the method by which assurance is measured.

PDA Technical Report No. 26 (**5**) requires that filter validation trials include samples from three membrane lots that have BP or other relevant physical measurement at or near the specification limit. The acceptance criterion is that test filters (three or more) show no passage of organisms. (If a test filter shows passage without an assignable cause, a confirmation retest may be performed.) Though not stated in the guidelines, a high probability of success would be needed to demonstrate that the filtration process is adequate.

A technician tests the bubble point of a filter membrane. (WWW.MILLIPORE.COM)

Taken together, the FDA and PDA guidelines establish the need to have a high confidence of obtaining sterile filtrate. From them, one can infer a level of retention assurance. To succeed with the recommended PDA validation study, the confidence that three limit samples retain a high *B. diminuta* challenge needs to be at least in the range of a capable process running in an operational

qualification near its process limits. That implies that validation success probability needs to be 0.99 or higher. That level of success requires the retention probability of an individual sample at the specification limit of a membrane to exceed $0.99^{1/3}$, or about 0.997, because

$$\begin{aligned}\Pr(\text{three retentive independent samples}) &= \\ \Pr(\text{sample A is retentive}) &\times \\ \Pr(\text{sample B is retentive}) &\times \\ \Pr(\text{sample C is retentive}) &\\ > \min(\Pr(A), \Pr(B), \Pr(C))^3&.\end{aligned}$$

The objective, then, is to design a membrane and a measurable specification that will provide such retention assurance. Here, I show how retention confidence of a membrane can be reliably estimated during its design. I select a general model of retention performance and a robust statistical method for making estimates. Both simulations and actual membranes demonstrate method performance.

Method Development

Relationship of Modeling Retention and/or Passage to Pore Size: When selecting a model, it is useful to have a description of membrane retention as a function of some physical property, such as pore size. Meltzer and Jornitz provide a valuable survey of work done on this topic. In it, they cite examples of a correlation between LRV and BP and discuss the effect of BP measurement error on the precision of LRV estimates. Although most of the cited work is empirical rather than theoretical, it does provide a useful starting approximation.

With regard to LRV, the proportion of organisms passing through a filter is defined by Cd/Cu, where Cu is the number of cfu in the upstream feed, Cd is the number of cfu in the downstream filtrate, and LRV is $-\log(Cd/Cu)$. When $\log(Cu)$ is nearly constant, as in the ASTM *B. diminuta* test, the observed linear relationship between LRV and BP also holds between $\log(Cd)$ and bubble point. When the estimated Cd is less than 1.0, a fraction of all tests will have sterile filtrate, and the remaining fraction will have a low Cd. In such cases, the probability of sterile filtrate, or retention confidence, can be calculated from the Poisson distribution:

$$\begin{aligned}\text{retention confidence (\%)} &= \\ \exp(-Cd) \times 100&, \text{ or equivalently,} \\ \exp[-10^{-\log(Cd)}] \times 100&\end{aligned}$$

For example, when $\log(Cd) = -3$, retention confidence is equal to $\exp(-10^{-3}) \times 100 = 99.9\%$. Figure 1 illustrates these concepts.

Some theoretical work suggests that this is a reasonable model. A membrane can be viewed as a sieve with some defining property such as a pore diameter distribution. In such a case, it is straightforward to show that retention (or passage) probability of a given particle size must increase (or decrease) monotonically as the fraction of pores larger than the particle decreases. Also, any retention relationship between pore and particle distributions that has approximately exponential tail behavior will lead to a linear relationship between log(passage probability) and pore size.

Finally, I must note some assumptions in any such analysis. I am evaluating one membrane property, bubble point, as the primary design property affecting retention. In membrane design, it is necessary to confirm whether other properties (such as thickness) have significant effects on retention. These need either to be evaluated at worst-case conditions or kept in a narrow range. If the effects of other properties are to be investigated, the model I develop here can be easily extended to include multiple variables. Further, it is important to establish that size exclusion is the primary mechanism of membrane performance. Additional studies or test conditions can isolate whether size exclusion is the sole or primary mechanism.

Other models and measurements are certainly possible. For example, one can measure the approximate fraction of pores greater than a given size by means of porometry. Or, if using bubble point, one can use the inverse relationship between BP and maximum pore diameter to select a model such as $\log(Cd) = a + b \times (1/BP)$. In practice, these are often not much different from the linear relationship with bubble point, unless the bubble point range approaches very low values.

One limitation of the model $\log(Cd) = a + b \times BP$ is that it implicitly assumes that retention is complete only at an infinitely high bubble point, and that no BP has a zero retention. If there are boundaries

Figure 1: Quantifying retention performance

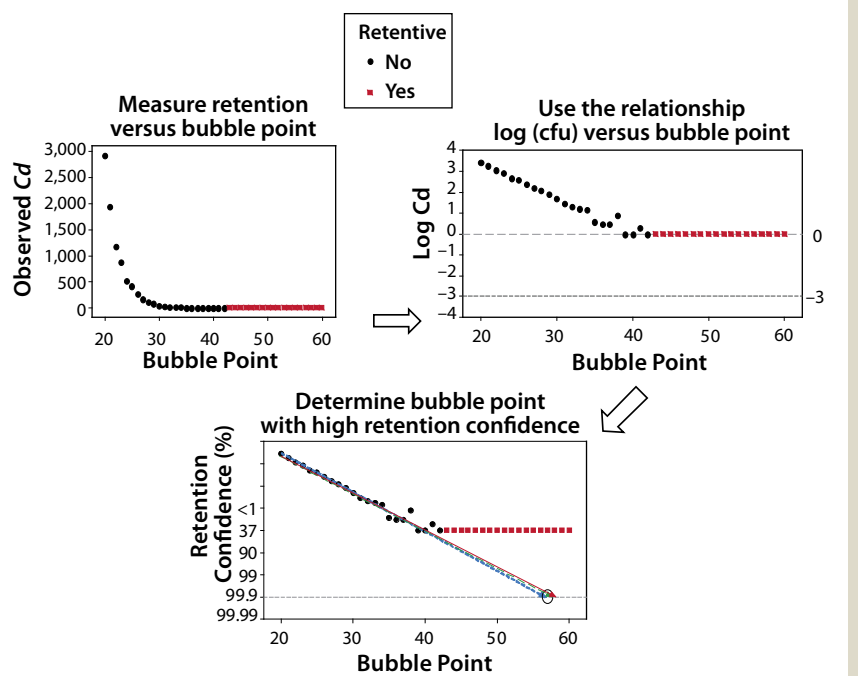


Table 1: Percent standard deviation of statistical methods at 99.9% retention confidence

Simulation	LRC*	LR*	PR*	NBR*
Ideal	2.5%	2.6%	1.9%	2.0%
1.5–3% Measurement Error	3.6%	3.1%	6.2%	2.7%
3–5% Measurement Error	9.0%	18.0%	25.0%	13.7%
3–5% Error + Dilution	1.2%	1.4%	16.7%	9.0%

*LRC = linear regression with censoring; LR = linear regression; PR = Poisson regression; NBR = negative binomial regression

for these two possibilities, one alternative model could take a form such as $\log(Cd) = a + b \times [(BP - \min)/(\max - BP)]$, where

min = a minimum bubble point where retention is zero, and

max = a maximum bubble point beyond which retention is 100%.

Selecting a Method for Statistical Analysis: The next step is to select an analysis method that reliably estimates retention probability. Such a method must be applicable to the models developed above. It must reliably estimate retention in the important region where no downstream colony forming units are observed (filtrate is sterile) under a number of practical conditions, including

- retentive and nonretentive outcomes.
- presence of error in BP measurements
- some Cd values that are determined by diluting a small fraction of the filtrate
- either a linear or monotonic (smoothly increasing or decreasing) underlying relationship between bubble point and $\log(Cd)$

Four statistical methods provide accurate and precise estimates under at least some of these conditions. They are

LR: linear regression of $\log(Cd)$ vs. bubble point, excluding results with zero Cd.

LRC: linear regression with censored data similar to LR, but zero *Cd* is treated in the analysis as a censored outcome at $\log(Cd) = 0$. In this analysis, the interpretation is that the observed *Cd* is less than one, but some proportion of future results could be positive, to be estimated by the model.

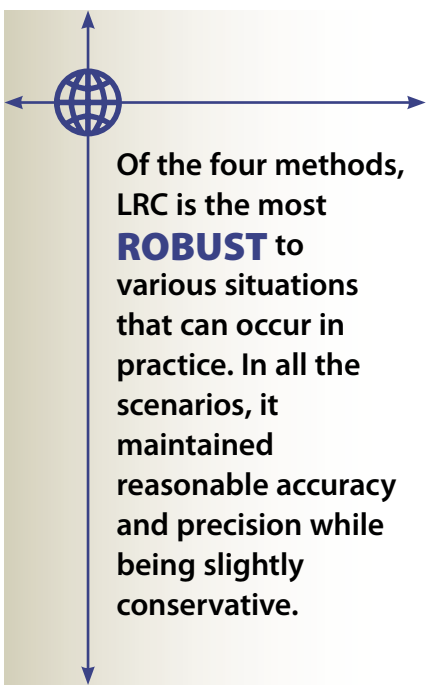
PR: Poisson regression of $Cd = \exp(a + b \times BP)$, where the residual error is assumed to derive from the Poisson variation of biological outcomes. When *Cd* equals zero, the data are included in the analysis.

NBR: negative binomial regression, similar to PR, where the residual error is assumed to be larger than Poisson variation.

Other methodological approaches had drawbacks that prevented their inclusion. For example, a simulation extrapolation method (SIMEX) is described by Carroll, Ruppert, and Stefanski (**6**). But quantitative knowledge of bubble point measurement error must be known to obtain accurate estimates by that technique. In practice, such information is not always available to the degree necessary for the method. Other approaches, such as orthogonal line fitting, have been attempted but are known to have drawbacks. A discussion of these and related methods can be found in Carroll et al.

Evaluating Performance of the Statistical Model and Method: To evaluate performance of the models herein, I conducted a small simulation study, structured in the following way.

- Retention probability increased with BP
- Both retentive and nonretentive samples were included
- A moderately wide range of bubble points were included



- Bubble point measurements had errors
- Some nonretentive results may have required dilution to estimate downstream counts.

These are conditions typically expected in membrane design evaluations.

Specific inputs of the simulation are as follows. The simulated relationship between *Cd* and *BP* is $\log(Cd) = 16 - 0.4 \times BP$ when measured without error. In this model, $\log(Cd) = -3$ at a BP of 47.5 ("psi", for this simulation). This corresponds to a 0.001 probability of a positive count, or a 0.999 probability of a retentive outcome. I will refer to this as the bubble point of 99.9% retention confidence (BP99.9).

The BPs in the simulation data set range from 34 to 70 in increments of 1. This represents a larger quantity of samples than a typical study would include, but the larger simulation data set helps to illustrate the conclusions more clearly. The simulation consists of nine independent trials in which random Poisson counts are generated at each bubble point.

Those runs are referred to as "ideal" runs, because measurement error has not been introduced into the bubble point. Nine additional runs are created by adding to each "true" bubble point a random error with mean of zero and standard deviation of one. This is an example of small to moderate measurement error (about 1.5% to 3% of the true value). Also, another nine runs are created by adding to each true bubble point a random error with mean of zero and standard deviation of two. That is an example of moderate to high measurement error (3 to 6%). Each statistical method is used in each run to estimate BP99.9. The average and percent standard deviation of the nine trials are calculated to obtain accuracy and precision estimates, respectively.

To simulate dilution effects, an additional nine runs were constructed from the high measurement error data set, then supplemented with added data points between 20 and 33. In this range, the expected $\log(Cd)$ exceeds 2.4 (or cfu exceeds about 250, a

typical "too numerous to count" level in bacterial assays). A dilution factor is applied, and a random Poisson count is generated for the "diluted" sample, then multiplied by the dilution factor (10×, 100×, and so on).

Results — Simulations

In the ideal case (statistical model matches actual practice, wide range of data, no bubble point measurement error, no multiple dilutions required), all statistical methods tested provide reasonably accurate and precise results. The estimated BP99.9 with the PR and NBR methods are exactly on the target of 47.5 and just slightly above target at 47.9 for LR and LRC. Percent standard deviations of the estimates (Table 1) are 2.6% or less for all four methods.

However, as measurement error is introduced, method performances start to change. In general, accuracy of BP99.9 estimates become more conservative (higher). Of the four methods, LRC and NBR methods stay closer to target. When BP measurement error is under 3%, all of the methods degrade only slightly, with estimates between 48 and 50. At higher measurement error, the LR and PR methods become over conservative, with BP99.9 estimates of 55 to 58. LRC and NBR continue to remain closer to target, at about 50 and 51, respectively.

Precision also degrades with measurement error. At low measurement error, relative standard deviation of the PR method increases more than the other methods, climbing to about 6%, whereas the other methods are around 3%. At higher measurement error, the LRC method maintains a relative standard deviation under 10%, but the other methods range from 14% to 25%.

In the scenario that includes higher measurement error and a wider range of bubble points that require dilution for evaluation of *Cd* (Figure 2, rightmost case), the performance ranking of the four methods changes. The additional data represent an advantage for the LRC and LR methods, which have estimates within 1% of target and precision of 1–1.5%. However, the change in error structure has a severe impact on the estimates from the PR and NBR methods, which are biased by over 20%, with precision of 10–15%. Although it is theoretically possible to construct a modified NBR method that adapts to the problems posed by sample dilution, it makes the method more complex.

Of the four methods, LRC is the most robust to various situations that can occur in practice. In all the scenarios, it maintained reasonable accuracy and precision while being slightly conservative. Next in performance is NBR, which is best under ideal conditions but suffers larger accuracy and precision penalties under more typical experimental conditions. LR can be equivalent to LRC if there are very few cases with zero cfu in the filtrate. However, it is usually desirable to have such cases to demonstrate membrane retention, and the LR method completely discards them. PR makes more restrictive assumptions than NBR and suffers a greater penalty when they are not met. In all cases, it is important to note that BP measurement error has detrimental effects on any model.

Figure 2: Log(*Cd*) versus bubble point simulation accuracy of simulation methods

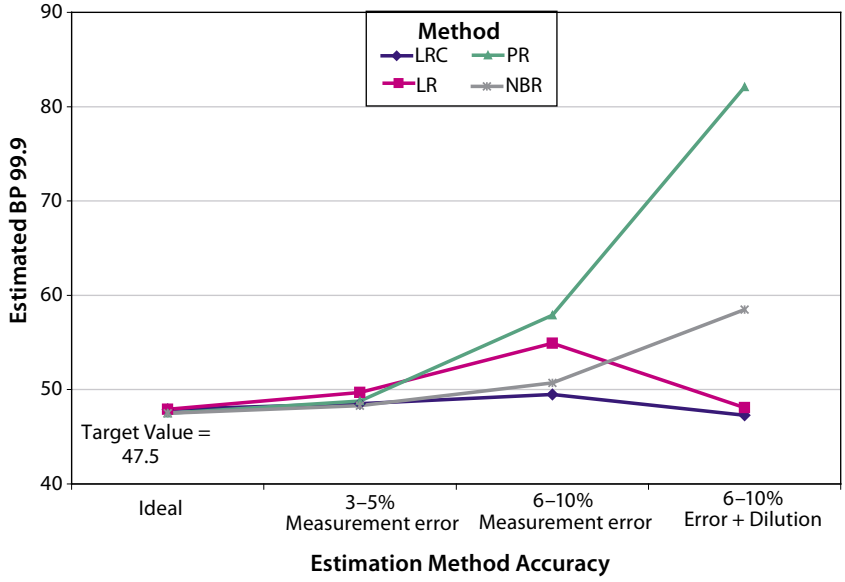


Figure 3: Log(*Cd*) versus bubble point from simulations; log(*Cd*) = *a* + *b* × *BP*

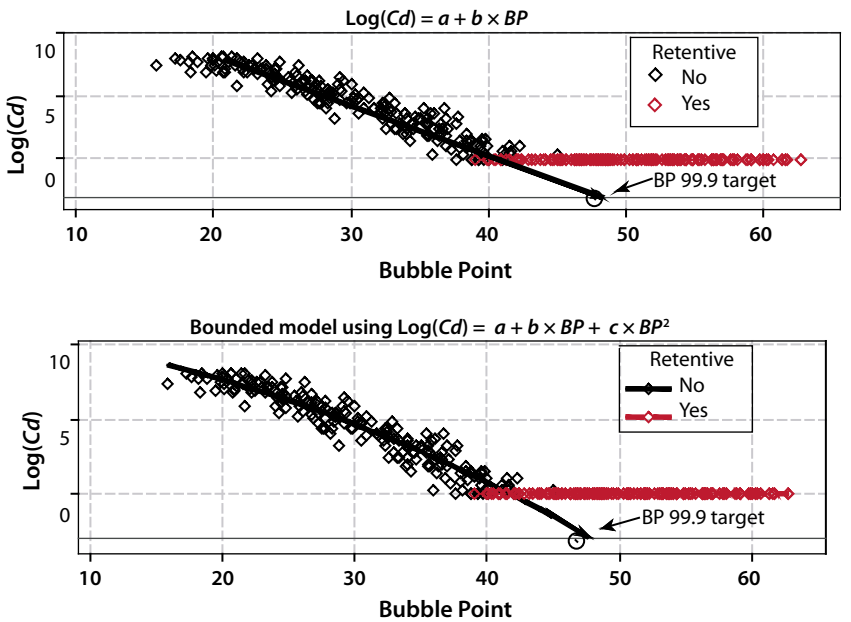
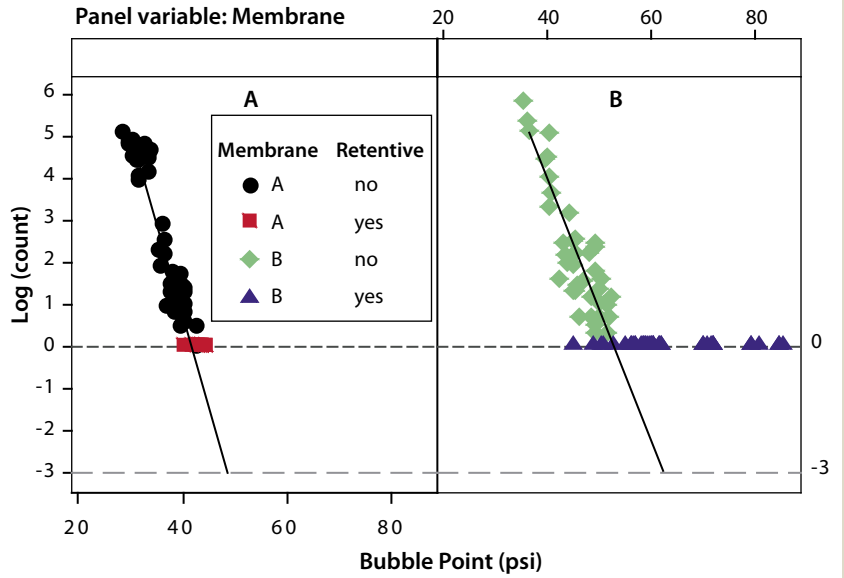


Figure 4: Retention versus bubble point — two membrane type examples



Estimation and control of such variation is always highly recommended as a general principle.

A final simulation exercise extends the potential of the top-performing LRC method to obtain reasonable estimates under an alternative model. In this scenario, a simulation was run using the alternative bounded model described earlier. The relationship between retention and bubble point was defined by $\log(Cd) = 8 - 22.5 \times [(BP - 20)/(100 - BP)]$, with a BP99.9 target of 46.3 Again, nine trials were constructed in the same way as described above, including 3–5% measurement error and sample dilution.

Figure 3 shows all data from this simulation. A line shows the overall fitted relationship between log(*Cd*) and BP. For the bounded model, there is distinct curvature in the fit. When a second order polynomial is used as an approximation for the bounded model (shown in the figure), the estimated BP99.9 is 46.7, which is quite close to the target value of 46.3. The linear model has an estimated BP99.9 of 47.5.

As with any model, there is opportunity for additional development. It is rather straightforward with existing commercial software to include cases where high Cd results are not enumerated and must be treated as censored. We have not attempted to fit the parameters of the bounded model in this body of work, but this would be an area of future interest and would require a combination of nonlinear curve fitting and ability to accommodate censored data.

I have demonstrated the ability of appropriate statistical models to estimate retention confidence with reasonable accuracy and precision under realistic conditions. The final aspect of this work is to apply these results to actual membrane designs.

Results: Membrane Design Examples

Figure 4 shows actual data from two membrane types. The membranes have different materials, formation process, chemistry, and cross-sectional structure. A range of bubble points was manufactured and tested for each membrane.

The graphs show retentive and nonretentive outcomes. Both cases shows a significant linear relationship between log(*Cd*) and BP (slope z-test; –values < 0.001), with no indication of significant curvature (quadratic coefficient z-tests; $p > 0.16$). With membrane A, the estimated high confidence BP is about 49 psi. With membrane B, it is about 62 psi. The slopes of the log(count) and BP lines for both membranes are significantly different (z-test; $p < 0.001$). Both designs have been commercialized and have a successful bacterial retention track record.

Accurate, Precise, and Adaptable

A statistical analysis of filtrate bacterial counts (*Cd*) vs. bubble point (BP) enables estimates of retention confidence. The analysis is a linear regression between log(*Cd*) and bubble point, in which retentive outcomes are handled as censored observations. These estimates are shown to be reasonably accurate and precise under conditions of measurement error and dilution effects. I compared estimates from a simulation with estimates from other statistical analyses that can provide accurate outcomes under a more limited range of conditions. The statistical analysis can be adapted for other models, such as a bounded model where complete retention and passage occur. Two examples of actual membrane results demonstrated that the model and analysis perform as expected.

References

1 Johnston P, Meltzer T. Comments on Organism-Challenge Levels in Sterilizing-Filter Efficiency Testing. *Pharm.Technol.* 3(11) 1979: 66–110.

2 ASTM International. *F838-05 Standard Test Method for Determining Bacterial Retention of Membrane Filters Utilized for Liquid Filtration* (2005); http://webstore.ansi.org/ansidocstore/product.asp?sku=ASTM+F838-05.

3 *Filtration in the Biopharmaceutical Industry.* Meltzer T, Jornitz M, Eds. Marcel Dekker, New York, 1998.

4 US Food and Drug Administration. *Guidance for Industry: Sterile Drug Products Produced by Aseptic Processing — Current Good Manufacturing Practice* (2004). http://www.fda.gov/cber/gdlns/steraseptic.htm.

5 Parenteral Drug Association. *Technical Report No. 26: Sterilizing Filtration of Liquids* (1998); https://store.pda.org/bookstore/technicalreports.aspx.

6 Datta S, Redner S. Gradient and Percolative Clogging in Depth Filtration. *Int. J. Mod. Phys. C.* 9(8) 1998: 1535–1543.

7 Carroll R, Ruppert D, Stefanski L. *Measurement Error in Nonlinear Models.* CRC Press: Boca Raton, FL, 1995.

***Mark M. Blanchard** is a consulting research and development statistician at Millipore Corporation, 80 Ashby Road, Bedford, MA 01730, Mark_Blanchard@Millipore.com.*