By Any Measure…

Thomas R. Adams, P.E., Vaikko P. Allen and Andrea Perley
Vortechnics, Inc.

The introduction of a diverse array of stormwater quality management tools in the last few years has created problems for the growing number of individuals and organizations who would like to compare the performance of these tools. Comparison is complicated by differences in treatment capacities, targeted pollutants, and treatment approaches. Several methods of evaluation have emerged in response to the need for verification of theoretical performance predictions; yet none of these "yardsticks" are appropriate in all situations and results from each are often not readily comparable to results from other measures.

Complicating the matter further is the confusion regarding what is being compared. In some cases, a technology will be compared with another technology. In other cases, the technology is compared to a performance standard. (Analogy: My maple syrup may be better than your maple syrup, but does that make it Grade A maple syrup?)

The confusion stemming from this is greater than meets the eye. For example, many specifying engineers and hydrologists want to meet a performance standard of 80% TSS removal on an annual average basis. They go to guidance manuals and product manufacturers seeking something that will meet the standard. As the selection process develops, they grapple with cost, maintenance, the availability of land needed, and many other issues under the heading of "cost-effectiveness" for their clients. But by the time the selection has been made, it has become more a question of who has the better maple syrup rather than whether or not the selected product meets the standard. That is partly because 80% TSS removal on an average annual basis is virtually an impossible standard to guarantee.

It ***is not the purpose*** of this paper to propose the adoption of one standard over another.

It ***is the purpose*** of this paper to review the merits of various measures of performance and, more importantly, to stress that any monitoring program that attempts to measure performance of a stormwater quality management system should begin with a consideration of how the observations will be reported.

Any monitoring program will consist of:

- Sampling
- Testing
- Reporting

Often, researchers start with the selection of samplers and proceed with questions of deployment and maintenance of the samplers. Only when samples have been collected are the questions of testing, and eventually reporting, given much thought. We propose that the reporting aspect, even though it is "last" chronologically, should dictate how the testing and sampling are done. For example, if the report is considered most informative when its focus is the mass of pollutants removed, as is often the case, there is no need for samplers. To get those results usually requires only the very simple task of periodically measuring accumulations of precipitated sediments or floatable petroleum products (or other pollutants) and doing simple arithmetic calculations to determine the mass removed.

We know of projects that went to the extraordinary effort that it takes to procure, set up, operate, and maintain automatic samplers to obtain influent and effluent concentrations, only to "back into" an estimate of the mass of pollutants removed by a convoluted set of "volume-times-concentration" calculations. This is a classic case of doing something the hard way, not to mention the very expensive way. So the first rule of thumb for any monitoring project is to ***decide first how to report the results***. Then design the sampling and testing around the information that is to be reported.

In this paper, we will present the following four "measures" of performance from which to choose in making those decisions. We consider these measures to be the current "status quo of the art."

1. Mass of pollutants removed
2. Event mean concentration (EMC)
3. Lines of Comparative Performance (Minton, et. al.)
4. Plotting efficiency versus operating rate

First, we would like to discuss in some detail some of the broader issues involved in monitoring stormwater treatment systems and measuring and reporting on their performance.

Setting

The setting is nearly always in the field or in a laboratory. We feel that technical professionals and non-professionals alike generally underestimate the many benefits of testing in a laboratory. What we refer to as the setting tends to be pre-determined well before a study gets underway. Since the setting tends to influence the important decision of what performance measure is best, this cursory overview is provided for perspective.

Field testing

The drawback with field testing is that it cannot be replicated very well. Every site is different. Every storm is different. There are "wet years" and "dry years." There are seasonal variations that can produce easily treated heavy sediment loads in winter and spring; hard-to-treat loadings, such as pollen and grass clippings, in summer; and moderately treatable loadings (leaves etc.) in the fall. It is poor science to compare the results of a field test of any treatment system to the results of a test of the same system at a different time and place.

Still, field testing has tremendous appeal because the stormwater and the sediments are "real."

- Field testing of individual facilities is usually adopted to evaluate the facility's performance in comparison to a performance standard (such as 80% TSS removal) to see if it "measures up." This setting is the simplest and most common, and is adaptable to any of the measures. The treatment system can be set up to treat all runoff or it can bypass flows that exceed the treatment capacity.

- "Side-by-side" field testing of several different facilities is increasingly popular, at least in concept. But, to our knowledge, this approach has not yet been successfully implemented. A key element is the design of the "flow-splitter" that takes all of the runoff and "splits" it into an equivalent discharge to each treatment system. The easy part of the design of the flow-splitter is achieving equal flow rates of ***water*** to each of the facilities (and even this "easy part" is not always all that easy). The hard part is getting equal discharge of ***pollutants*** to each of the facilities.

Whether testing an individual system or multiple systems, researchers have the potential to learn something very basic and very important from field testing that has, to our knowledge, eluded researchers to date. That is the determination of an appropriate threshold for bypassing peak flows. Consider a site that is estimated by conventional runoff modeling to discharge stormwater runoff at a rate of 3 CFS in a 10-year storm. Consider further two proposals for treatment. One claims efficiency of 90% and a treatment capacity of 1 CFS. The other claims 80% and a treatment capacity of 3 CFS. Which treatment option should be selected? If the prevalent standard of 80% is in place, the "safe" choice would seem to be treating all runoff from a 10-year event with 80% efficiency. But what if price is a factor? Maybe the system with 1 CFS capacity costs less. Even at the same cost, would 90% efficient be preferable to 80%? And with that consideration comes the $64,000 question. Will the system that claims 90% efficiency with a capacity of just 1 CFS even meet the standard of 80% overall? How much of the pollutant load will bypass the system altogether?

Some proponents of small-flow/high-efficiency technologies have stated that 90% of all storms are less than 1 inch of total rainfall and, therefore, treating 90% of all rain at 90% efficiency will yield a net annual removal of 81%. This argument is fundamentally flawed. It assumes that the 90% of rain from small storms carries 90% of the pollutants. This is simply not the case. The rate of mobilization of virtually all pollutants depends on rainfall intensity, not depth of rain. Therefore, it is important to treat high-intensity flows resulting from the ***infrequent event***, which tend to carry a disproportionately high pollutant load.

If 1 inch of rain falls in 24 hours, virtually any system that is reasonably proportioned, designed and, of course, maintained for the treatment of stormwater will do a good job. Efficiencies of TSS removal should be in the 90% range if the runoff is fairly dirty with silty-to-fine sandy sediments.

It is questionable, however, as to whether or not all of the runoff would be dirty if the rain that produced the runoff totaled 1 inch and fell over a 24-hour period. Intuitively, the "last flush" of such a storm would be very clean. But even the first flush may be very clean in comparison to what it would be if 1 inch of rain fell in 1 hour. This highly variable "dirtiness" gives rise to another interesting question when trying to measure efficiency. That is the question of how to account for the inevitable reduction in treatment efficiency when the water to be treated is clean in the first place. No treatment system can remove what is not there. So it has been argued that some accounting should be made for the fact that there is some lower limit to the physical treatment that can be provided. Minton's "Lines Of Comparative Performance" (see figure 2.) take this important consideration into account and are discussed later.

The "Double Whammy" of the "2-Month Storm"

Infrequent, high-intensity storms are important to the effective treatment of stormwater for two reasons:

1. Over time, the higher intensity of less-frequent rainfalls, and the resulting higher stormwater runoff velocity, is what transports most of the sediment off of streets.
2. The treatment facility is overloaded by the high flow of water that is transporting the sediment at the same time that most of the sediment is being transported to it.

Schueler and Shepp (1993) performed monthly observations documenting a random pattern of accumulation and loss of sediment in a study of 17 different oil/grit separators in Maryland.

Overall, the losses of sediment "outnumbered" the accumulations. In other words, the observed systems lost previously accumulated sediments ***once every two months.*** We have inferred from their work that the "2-month storm" is a reasonable benchmark for stormwater treatment. To be "measurably" better than the poorly reputed conventional oil/grit separator, a system or a facility should be able to demonstrate, at a minimum, that it can continue to function in the 2-month storm. If a system is found to lose sediment in a 2-month storm, it should not be considered any better than conventional technology. Similarly, if a system needs a bypass to protect it from washing out in 2-month storms, it should be considered only marginally better than conventional oil/grit separators. Bearing in mind that high flows transport much of the total sediment, treatment systems should be able to handle more than ***the 2-month storm without bypass.*** Otherwise, much of the total sediment load may be discharged to the receiving waters that the system is supposed to protect.

Clearly the statements of the preceding paragraph are more of a hypothesis than a statement of fact. One way to validate or invalidate the hypothesis is described in the following section on side-by-side testing in the field.

Side-By-Side Testing

Testing stormwater facilities "side-by-side" has recently become a very popular idea. The premise is that a well-run comparison of systems treating "the same stormwater and the same pollutants at the same rates of flow" will go a long way to reduce the tremendous "scatter" in the data that has been obtained to-date by testing individual systems at different sites. If two systems are evaluated at different sites, even if the same researchers using the same protocol carry out the study, the results will probably not be comparable. Every site is different, and from the point of view of stormwater treatment, differences that appear slight can actually be significant. We have observed a dozen systems installed on a single site (a large shopping mall parking lot) which were specified by the same engineer and installed by the same contractor at more or less the same time and, of course, subjected to the same weather. They exhibited decidedly different results when we measured sediment accumulations in the systems. The sediment depths ranged from a light dusting to accumulations of over two feet in less than a year.

So it is important in "side-by-side" testing that there be just one flow stream to the two (or more) systems being tested and that the flows be split, so that each system gets exactly the same rate of flow and the same pollutant concentration at all times.

The main benefit of "side-by-side" testing is that it can provide an answer to the question, posed earlier, of whether it is preferable to have, the arrangement should be as shown in Figure 1.

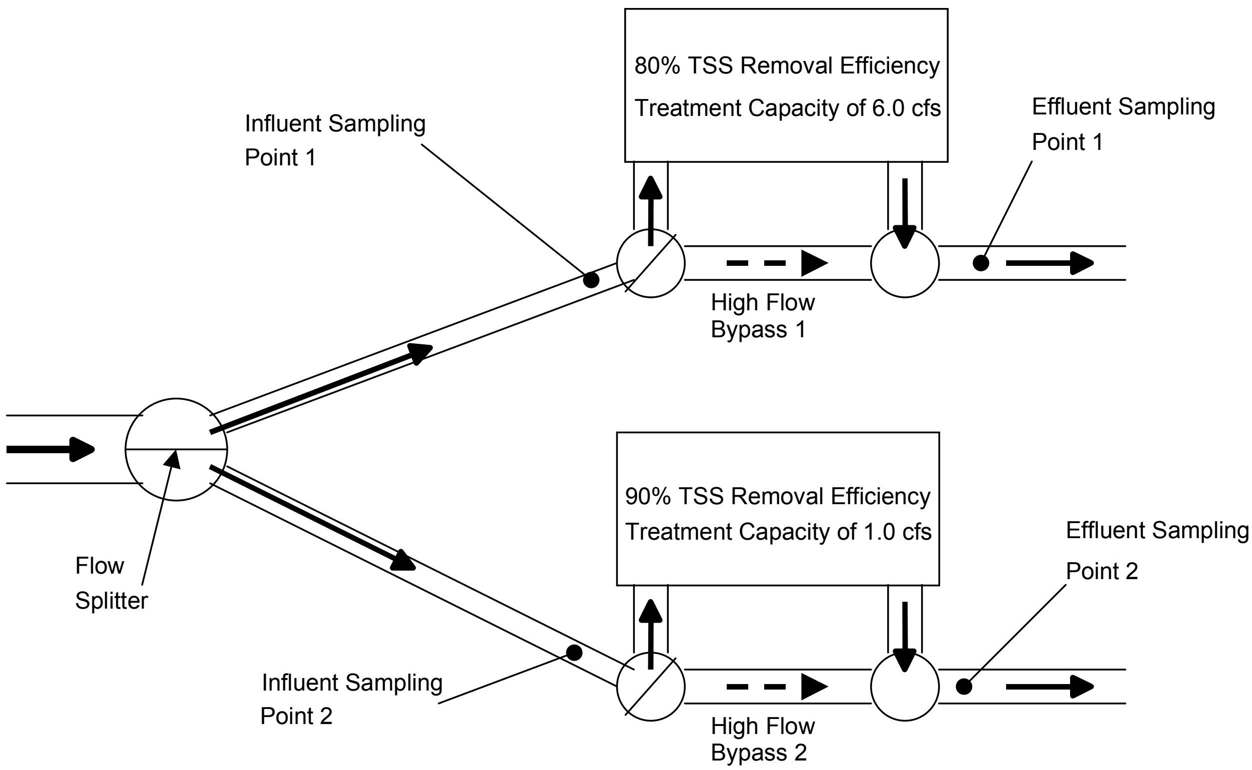

Figure 1. Recommended Arrangement For Side-by-Side Field Testing.

By sampling at points 1 and 2, the overall efficiency of the treatment system and bypass can be assessed objectively. Also, the question of "Which is the better system?" is answered. There are two shortcomings:

1. Lack of repeatability. If one system gets 80% efficiency overall and the other gets 70% overall during one year of testing, there is no assurance whatsoever that either number will be repeated the next year. The test results should be regarded as indicative of performance. They are certainly not an assurance of performance over time. Such is the inconsistency, or "noisiness," of stormwater treatment data. A study like this should be conducted over a period of no less than two years. If the second year's results are reasonably close (in terms of statistical correlation) to the first, it can be considered complete. If not, it would be tempting to average all results. We feel that it would be poor science to do so, however. With the seasonal variability of stormwater pollutant loadings, one year's results will produce a single data point. A second year's results will produce a second data point. Many people seem to regard each storm result as a single data point, but as long as standards continue to be based on "average annual removal efficiencies," that is simply not the case.

2. Variability with other sites. We have already mentioned the differences from one site to the next. The basic premise of side-by-side testing is to determine relative performance of two or more systems (i.e., which is best). As long as such a study is limited to this premise, the variability from one site to the next will not be a problem. But we know from

experience that any "study," even the most cursory, tends to be overly generalized. We can only caution against doing so.

Laboratory Testing

Testing stormwater treatment systems in a laboratory setting offers some very significant advantages over field testing.

1. It is repeatable and demonstrable.
2. It is more productive in the sense that decades of rainfall can be simulated in a matter of days.
3. It is more economical in terms of labor, sampling equipment, and flow-metering equipment costs.

Laboratory testing achieves these benefits by controlling operating rates, particle sizes, and pollutant loading. When influent concentrations are very low, removal efficiency will be low; but for concentrations that are generally recognized as representative for stormwater, all concentrations tend to produce comparable removal efficiencies.

In the lab, a set of tests can be run using one particle size (at representative concentrations) at operating rates from zero to the system's capacity. At the conclusion of these tests, a curve can be drawn plotting efficiency versus operating rate on the y-axis and x-axis, respectively. Such a curve typically slopes downward to the right, reflecting reduced efficiency and higher operating rates. Any point along a constructed curve should be reasonably reproducible when using the same influent sediment load.

Subsequently, a whole family of other curves can be constructed using different particles. Also, to more closely simulate "typical" sediment, a graded sediment sample can be developed and tested in the same way.

Laboratory testing should not be considered the "last word" in documentation of a system's performance, but can be considered a "benchmark" which is very useful in comparing systems operating at flow rates up to their capacity. Some field testing, where it is feasible, should supplement the work in the lab and, as previously discussed, side-by-side field testing is the only way to determine the impacts of bypasses on different systems.

The Four Most Common Measures of Performance

1) Mass of Pollutants Removed

This is easily the simplest approach to stormwater treatment measurements in the field. By measuring the depths of sediment accumulations in the facility on a periodic basis, it becomes a simple arithmetic exercise to calculate the volume and mass of sediments removed by the system.

Additional information is made available by this measurement. It may be recalled, from our earlier discussion of the 2-month storm, that Schueler and Shepp used measurements of sediment accumulations to document the poor performance of conventional oil/grit separators.

Researchers should consider using the same approach for the newer technologies that have come along since their important work was published. The approach can be made even more informative by correlating observations to such things as activities in the drainage area (e.g., winter sanding, sweeping, a spill, etc.) or meteorological events such as observed rainfall intensities or precipitation depths.

2) Event Mean Concentration (EMC)

These are sometimes referred to as "flow-weighted" or "flow-based composite" samples. They are nearly always obtained using automatic samplers, a flow-meter and a flow totalizer that arithmetically converts the flow rate measured by the flow-meter to flow volume over time and keeps track of the volume.

The sampler receives a signal that causes it to take a sub-sample when a programmed volume of flow is measured. For example, one sub-sample might be taken every 200 cubic feet of flow through the system. Over the course of the storm, all sub-samples would be combined into one large sample container from which the concentration will be obtained that represents the flow-weighted average for the entire storm.

Without a flow meter, the samplers could be set up to take a "time-based composite" sample; i.e., to sample every 30 minutes. Flow-weighted samples are much more representative, as a simple example will show. Consider a volume of 1,000 gallons with a uniform concentration of 300 mg/l flowing at a uniform rate past the sampling point in 15 minutes, followed by half as much volume (500 gallons) with a 100 mg/l concentration flowing by in the next 15 minutes. The correct representation of the concentration would be calculated as:

$$\frac{(1{,}000x300)+(500x100)}{1{,}000+500}=233$$

Flow-weighted sampling will more accurately reflect this. For example, if the sampler were programmed to pull a sample every 500 gallons, then 2 samples at the higher concentration would be taken and just one at the lower concentration. The average concentration would be calculated as:

$$\frac{300+300+100}{3}=233$$

Time-based sampling would, if the programmed time interval were 15 minutes, take one sample with a concentration of 300 and another with a concentration of 100, and the average would be calculated as:

$$\frac{300+100}{2}=200$$

Automatic samplers that can take flow-based composites have become a very valuable tool for sanitary engineers measuring concentrations of pollutants in wastewater. We believe that they have been too quickly applied to stormwater monitoring without regard for some of the inherent differences. Waste streams have "highs and lows" of both flow rate and concentration, but they are not nearly as wide as the variability of stormwater, which can change from flow rates of zero to a deluge in a matter of minutes and concentrations that can also exhibit a minimum of zero. These "spikes" can cause very brief periods of negative efficiency if a system is prone to washing out (as stormwater systems were shown by Schueler and Shepp to do regularly). If a wash-out occurs, it is an important phenomenon to note, but the briefly elevated concentration in the effluent will be "composited" with the rest of the (presumably lower) effluent samples. This will reduce the "event-mean-concentration," but will not reveal that a washout has occurred. Noting washouts, and the flow rate that caused them, is a very important aspect of a stormwater monitoring program; but they are not likely to ever be revealed by EMC data.

The second drawback of EMC data is that when influent concentrations drop to very low levels that cannot be further reduced by physical treatment, the efficiency, as measured by EMC's, will be reduced. This tends to obscure the fact that higher efficiencies can be achieved when they need to be achieved; i.e., when influent concentrations are higher.

3) Minton's "Lines of Comparative Performance"

It is widely acknowledged that there is a lower limit to the capabilities of physical treatment systems for stormwater. This means that it is very unlikely that effluent concentrations would ever be zero. It also means that very low concentrations would not be significantly reduced.

Minton et. al. has proposed the following mathematical expression to describe this lower limit:

$$\frac{Influent - LowerLimit}{Influent}$$

If the lower limit is 20 mg/l, then a plot of this expression is that shown in figure 2.

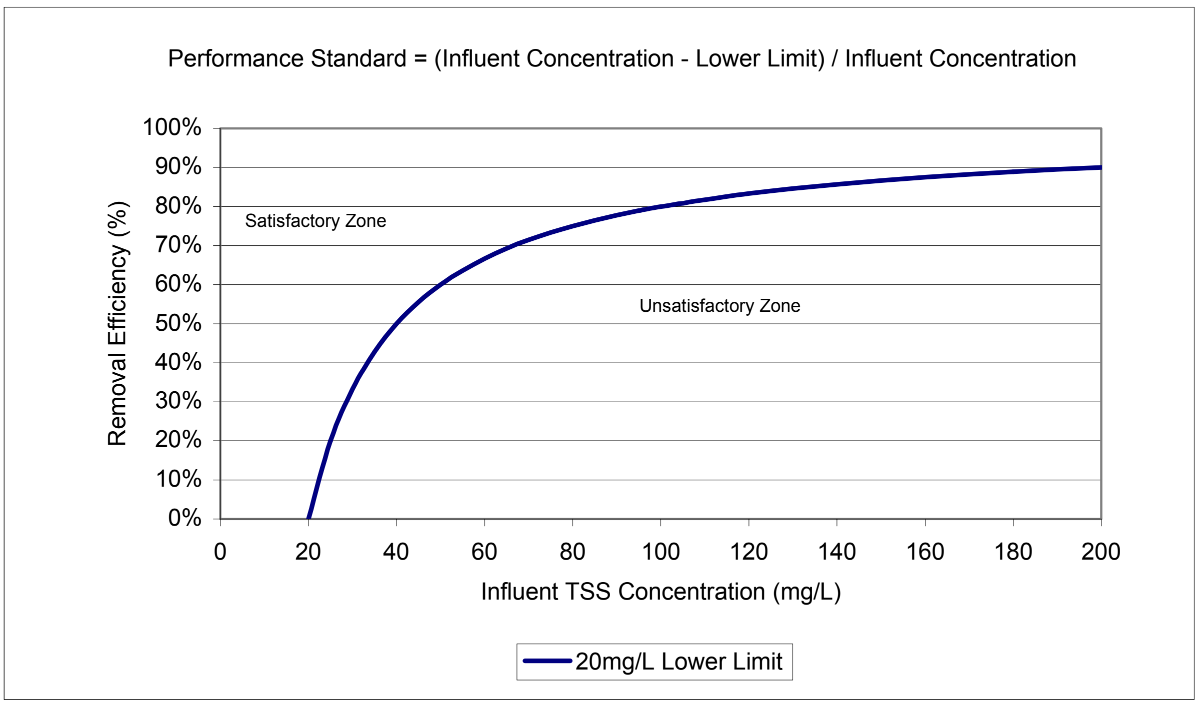

Figure 2. Line of Comparative Performance.

Plotting individual data points on such a graph can be very informative. Observed efficiencies above the line (designated a "line of comparative performance" by Minton) are considered satisfactory, while those below the line are unsatisfactory. Best of all, EMC's can be used without unfairly representing the efficiency. While the efficiency may be reduced by low influent concentrations, it may still be shown to be "satisfactory." More research will be needed to determine what is an appropriate lower-limit value to use in this approach.

4) Plotting efficiency vs. operating rate.

This approach was essentially described earlier under the heading of Laboratory Testing. By using discrete sediment particles in the laboratory, a family of curves such as those in Figure 3 can be developed.

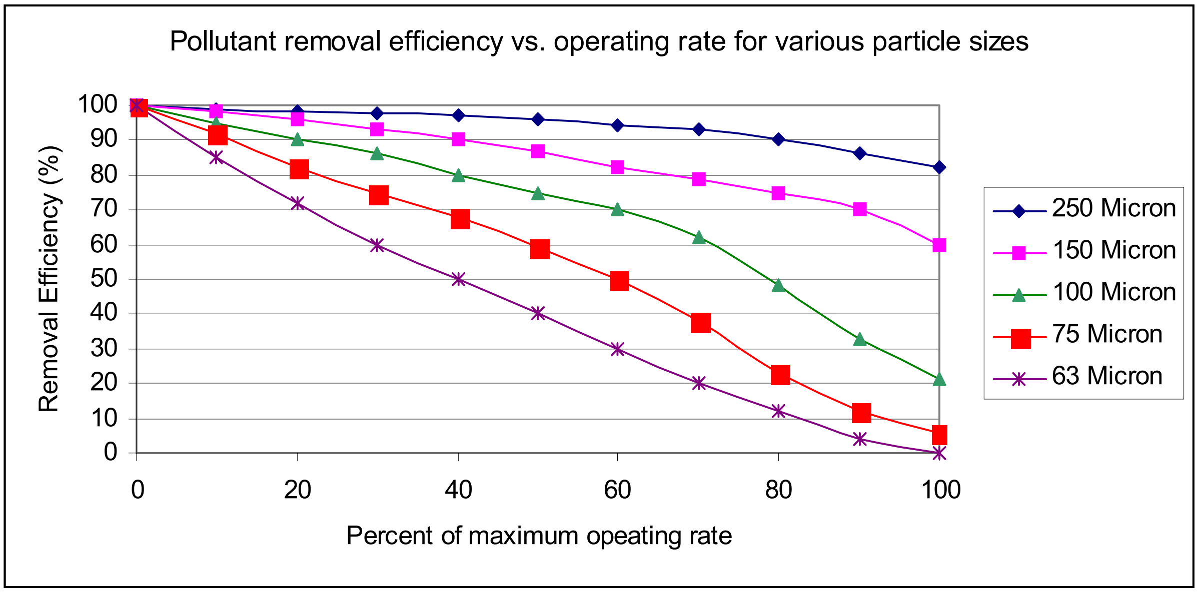

Figure 3. Removal Efficiency versus Operating Rate.

Field data is less likely to fit the relatively tight curves that can be generated in the lab. At the same operating rate, you may have vastly different influent concentrations, particle gradations, organic content, etc., depending on such factors as the time of the storm, antecedent dry period, and time of year. Removal efficiency is a function of all of these factors combined.

We feel that this performance measurement technique and presentation is the most informative. Its repeatability under controlled conditions makes it ideal for comparing one system to another. Certainly, if one system's performance curve on 100-micron particles, for example, is higher at all flow rates than another, it could reasonably be judged to be the higher efficiency system. If the curves are similar at low rates of operation, but either system drops down to zero efficiency at some higher flow rate, that flow rate should, of course, be considered the peak capacity for that system. This approach cannot show compliance with any standard for a stated percentage of TSS removal on an annual average basis.

Conclusion

To our knowledge, these four measures represent all of the techniques that have been used to measure the effectiveness of various stormwater treatment systems.

Measuring sediment accumulations in the field provides a good deal of useful information on mass removals and the ability to retain (or fail to retain) previously captured pollutants during periods of high flow. This approach costs very little to implement.

Event-mean-concentrations are the most widely accepted measure, but may not report all efficiencies and will almost certainly allow any failures to go undetected. This approach requires the use of automatic samplers at considerable cost, in terms of both time and money.

Minton's Lines of Comparative Efficiency are fairer to the treatment system because they account for the inability of any system to remove pollutants that are not present (or present in very low concentrations). If EMC data is collected to plot against the lines, then there are the same drawbacks of cost and automatic samplers allowing failures to go undetected. Both of those drawbacks can be overcome, but only with a very dedicated effort to take samples manually. Taking manual samples throughout the duration of all storms is very time-consuming and unpleasant work. For that reason, it is almost never done.

Plotting efficiency versus operating rate, whether in the field or in the laboratory, is arguably the most informative approach. In the field, automatic samplers are used (with individual samples in individual bottles and not composited), so there are those costs to consider. In the lab, samplers are needed but the construction of a model treatment facility, and the pumps and tanks to handle the required flow rates and volumes of water, will more than offset that cost saving.

Since none of these measures provides an ironclad confirmation that the widely prevalent standard of 80% TSS removal is being met, we submit that a different standard should be adopted by stormwater management jurisdictions.

References

1) Minton, Gary, M. Blosser, P. Bucich, B. Leif, J Lenhart, J Simmler and S. True, in press, Protocol for the Acceptance of New Stormwater Treatment Technologies in the Puget Sound Watershed, APWA Washington Chapter.

2) Schueler, Tom and David Shepp, The Quality of Trapped Sediments and Pool Water Within Oil Grit Separators in Suburban Maryland, Metropolitan Council of Governments, 1993 (revised).