BMP Performance Expectation Functions – A Simple Method for Evaluating Stormwater Treatment BMP Performance Data

James H. Lenhart

James H. Lenhart, P.E., D.WRE, is the chief technology officer for CONTECH Stormwater Solutions Inc. in Portland, Oregon. This article was presented at the ninth Bienniel Conference on Stormwater Research and Watershed Management, hosted in May 2007 by the UCF Stormwater Management Academy at the University of Central Florida.

Many regulatory agencies are struggling with how to set simple yet realistic goals for Best Management Practice (BMP) performance. Many regulations provide for simple removal rates of pollutants, such as an 80 percent TSS removal or a total annual TSS load reduction of 80 percent. Some agencies use other parameters, such as a 65 percent total phosphorus removal requirement. What is problematic is that these simple requirements do not reflect the reality of how BMPs actually perform in the field.

In efforts to get away from percent removal requirements, other attempts at setting flat effluent standards for BMPs (e.g. 20 milligrams per liter [mg/l] TSS effluent) are also problematic because the level of treatment required to constantly meet these standards is very high. To some degree, by definition, the performance of a BMP is probabilistic and presumptive; therefore, it has not been deemed practicable to constantly expect performance levels that meet effluent standards. There are also concerns that the cost of this level of treatment and associated maintenance are too high and that setting a fixed effluent standard may introduce complexities in terms of monitoring and compliance.

Clearly both approaches are problematic, yet both have beneficial aspects that perhaps could be combined to form a simple, realistic, and achievable performance standard for BMPs–a standard that can add a level of confidence that the BMP is going to meet the standards through analysis of field data. BMP performance claims should be based and verified with the confidence that a percent removal or effluent concentration or load reduction will occur given a range of influent concentrations and/or particle size distributions.

The Trouble with Percent Removal

From a purely analytical perspective the simplistic 80 percent removal requirement has some serious flaws. First, let's assume that influent concentrations are extremely low, say 20 mg/l. For an 80 percent reduction, the effluent would need to be 4 mg/l, which is often below the probable quantitative limits (PQL) set by commercial laboratories. In other words, with a PQL of 5 mg/l, the best any technology could ever achieve is 75 percent removal.

Another issue, which is even more significant, is the notion that there are irreducible concentrations (Schueler, 1996). This is predicated on the notion that given the operation of BMPs, that there is no expectation that the effluent will be below some amount. Many stormwater professionals accept that the irreducible concentration is at 20 mg/l (or greater) for TSS. In fact, advanced wastewater treatment regulations typically set effluent guidelines at 20 mg/l of TSS. Why would we expect a relatively simple stormwater BMP to outperform a plant with primary treatment, secondary treatment, automation, intensive maintenance and operators?

The irreducible concentration could also be viewed as a baseline effluent concentration. As an extreme example, say that water with zero mg/l of TSS enters a wetland. More than likely the effluent will not be zero and could easily be 20 mg/l. Though there is a net export of mass, at these concentrations this type of BMP behavior should not be a surprise.

Using the example above, given an influent concentration of 20 mg/l and a 20 mg/l irreducible concentration, the expectation for percent removal is zero. Clearly this is far from the 80 percent rule, yet given the practical reality of BMP performance, it is acceptable.

Data analysis using percent removal is typically not an accepted practice. The arithmetic averaging of percent removal, though sometimes used, generally is not accepted because it can be deceptive. For example, a series of small storms with small runoff volumes may yield higher removals due to long-term settling of displaced water. Less-frequent, higher-magnitude storms yield low removal rates but have much greater volumes of water being discharged. Simple arithmetic averaging could yield a result that the BMP worked well, when in fact, in terms of mass load, the BMP did not work well at all.

On the other hand, if influent concentrations are continuously low, the average percent removal is low and the BMP is judged not to work, when in fact, given irreducible concentrations, all that could really be concluded is that the site has a low pollutant load and the function of the BMP is indeterminate.

Another issue which has been discussed is plotting percent removal versus influent concentration. Typically when plotted, a characteristic curve is the result. The nature of the curve shows removal efficiency increasing with increasing influent concentrations. It has been shown that error plays a part in the characteristic (de Ridder et.al., 2002). The error is most pronounced at low concentrations due to analytical resolution, but are there other influences on the curve characteristic that may have a direct bearing on BMP performance?

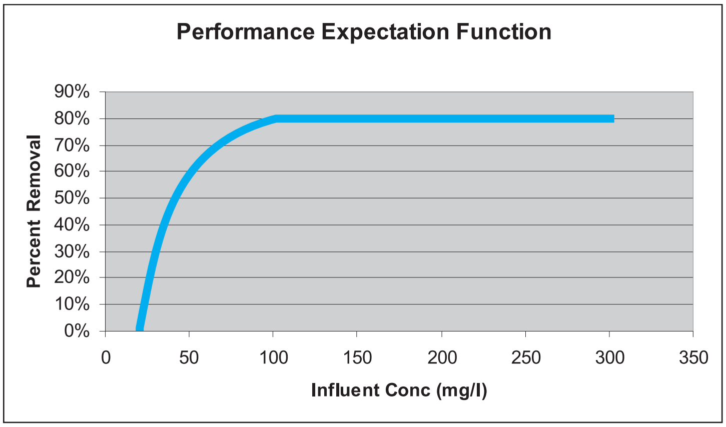
One major influence can be particle size distribution. Lehman and de Ridder (2005) showed a direct correlation between intensity and TSS concentration. In general, as storm intensity increased, influent TSS concentration increased as well. This finding is consistent with physically based models in which increased intensity results in more detachment energy, higher peak flows, and transport energy. Though not applicable in all cases, it leads to the hypothesis that higher removal efficiencies at higher concentrations are the product of transporting larger particles, which are easy to remove.

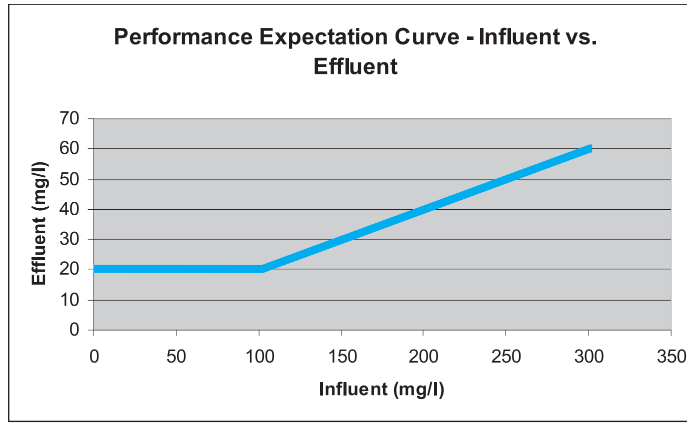
So it appears that there are both advantages and disadvantages in using concentration alone to evaluate BMP performance, but clearly, given these issues, simple percent removal as a standalone measure of performance should not be done.

Load Versus Concentration

Many will argue that the sum of loads or

Continued on page 30





Figures 1 and 2: Sample PEF Functions expressed as influent vs. percent removal and influent vs. effluent

Continued from page 28

mass load calculations are the only way to evaluate BMP performance. Others will argue that concentration is most important.

Mass load reduction is also a simple concept. Basically, mass load reduction is done by calculating the event mean concentration (EMC) of a storm times the runoff volume to yield the total mass of the influent and effluent. The percent reduction of the mass load is calculated from there.

While this method seems straightforward, there are issues with it as well (Strecker et. al., 2004). Say, for example, that a BMP gets a series of small storms with EMCs of about 100 mg/l. The EMCs of the effluent are at about 70 mg/l, which yields a 30 percent removal of TSS; however, a large storm transports a huge amount of mass (possibly consisting of large volumes of sand) at a concentration of 1,000 mg/l with an effluent of 100 mg/l for a percent removal of 90 percent. When the sum of loads is conducted, the amount of mass and high removal of the one storm outweighs the others and leads to the conclusion that the BMP achieved an 80 percent reduction of mass load; therefore, it was working.

What is problematic is that even though an 80 percent mass load reduction was achieved, the effluent concentrations were high and still exhibit significant water quality impacts, so in this case, one might accept a BMP that really does not meet water quality needs.

On the other hand, let's say that a BMP has influent EMCs of 50 mg/l and effluent EMCs of 20 mg/l for five storms, and one storm at 120 mg/l in and 24 mg/l out. The sums of loads removal is calculated to be about a 66 percent removal. This result may lead to the conclusion that the BMP does not meet the 80 percent goal and is therefore rejected, even though given the concentrations, the BMP actually performed very well. Clearly, more data with higher concentrations may be needed to be conclusive, but these data are not sufficient to reject the BMP for low performance.

These situations lead to the conclusion that in order to understand the operation of the BMP, one must look at both load and concentration for making decisions on performance.

Performance Expectation Functions

From these discussions, it should be evident that simplistic percent removals on either a concentration or mass load basis do not allow for proper evaluation of BMP performance and effluent guidelines are not practicable; however, one may consider combining the two into a method that considers both concentration and mass load while being simple, flexible, measurable—and most important— achievable by many BMPs.

A Performance Expectation Function (PEF) can achieve these goals. The basis of the PEF is that the regulatory agency defines the PEF based on the agency's specific water quality goals. The agency defines the irreducible or baseline concentration (typically 20 mg/l for TSS) that constitutes an effluent guideline for concentration below a threshold amount. Then, for influent concentrations above the threshold, percent removal (typically 80 percent) is used.

For example, with a baseline concentration of 20 mg/l, an agency would set an effluent guideline of 20mg/l for influent concentration of 100 mg/l or less. For concentrations greater than 100 mg/l, the performance expectation is 80 percent. Put simply, the PEF would be "for concentrations less than or equal to 100 mg/l, the expected effluent is 20 mg/l and for influent concentrations greater than 100 mg/l, the expected effluent is 80 percent of the influent."

Figures 1 and 2 show how the PEF can be illustrated in two ways. The first is a plot of influent versus percent removal and the second is of influent versus effluent.

This curve now defines the performance expectation of the BMP. Since the BMP performance is probabilistic, one would expect that some of the data points will be above the line and some will be below the line.

It is important to realize that the PEF can be used for other pollutants such a phosphorus and metals or can be more complex. For example, the city of Portland wants the concentration percent removal to rise to 90% at concentrations exceeding 280 mg/l

Using the PEF to Evaluate BMP Performance

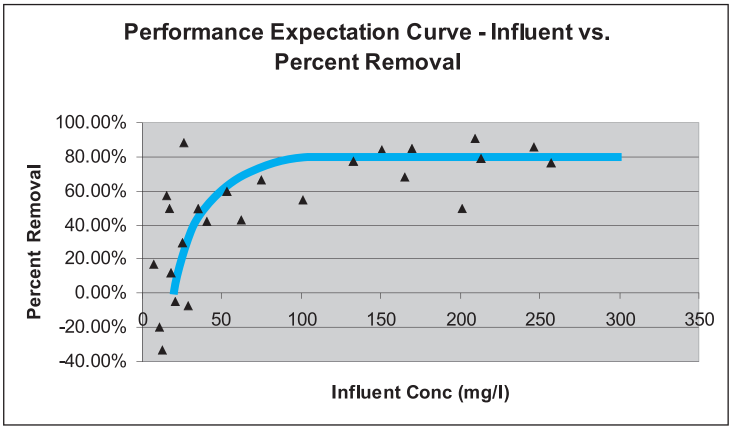
Once a PEF is defined by the regulatory agency, observed performance data from a qualified BMP monitoring project can be used to compare how the observed performance meet the expected performance as defined by the PEF.

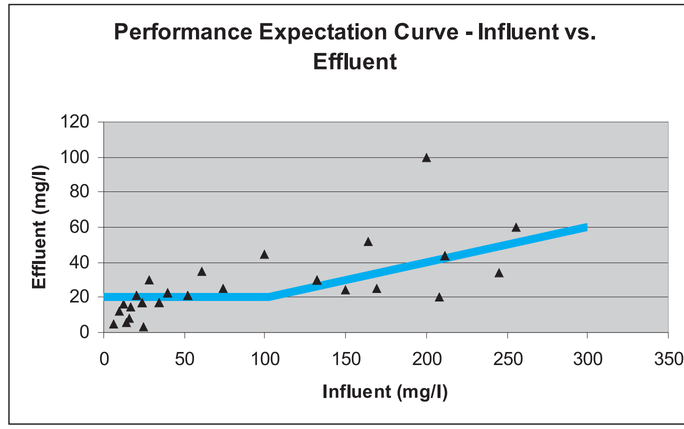
For the sake of illustration, a hypothetical data set was constructed and is shown in Table 1. The sample population is 25, which for a field monitoring population would be considered substantial.

These data can then be plotted against the PEF to gain a visual perspective on performance. Figures 3 and 4 present the data in a graphical format with the PEF.

Once the data are plotted against the PEF, one can begin with a numerical and visual analysis of the data. Though both graphs are presenting the same data, the influent versus percent removal seems to convey more information. Looking at the influent versus the effluent, it seems that in viewing the points, the question of what fraction (percent) is removed is always asked. Some additional visual aspects are:

1. **Spread of the data points**. Do the data points have a tendency to group or scatter? Data points that form tighter groups should represent a more robust and predictable technology. Scattered points indicate a lot of variance in the performance





Figures 3 and 4: Performance Expectation Functions vs. Observed Data

characteristics.

2. **Position of the points about the line.** For percent removal, points above the line are exceeding expectations, whereas points below the line are not meeting expectations. If the majority of the points are tightly clustered and above the line, this is a good indicator that the technology is meeting or exceeding expectations. Clusters below the line indicate the technology is not meeting expectations. Finally, clusters about the line may be visually indeterminate.
3. **Outliers.** Note that in the example there are two points which may represent outliers. For the analysis, one may decide to include or exclude the points.

Data Analysis— Observed Versus Expected

It is important to understand that the PEF is defined by the "user" and the observed data points are plotted about the line; therefore, the PEF is not the outcome from a regression analysis of the points but is a defined performance standard from which one can compare observed versus expected.

One method of comparison is the sign test. This is a simple nonparametric statistical test to estimate if the scatter of the points about the line represent the same population or a population which rests above or below the line. For example, if the BMP performance characteristic did follow the PEF, it would be reasonable to expect that 50 percent of the points would rest above the line and 50 percent below. If higher frequencies of occurrence lay either above or below the line, then this may indicate that the BMP is either outperforming or underperforming expectations.

Influent (mg/l)	Expected Effluent (mg/l)	Expected Percent Removal	Observed Effluent (mg/l)	Observed Percent Removal
6	20	0.00%	5	16.67%
10	20	0.00%	12	-20.00%
12	20	0.00%	16	-33.33%
14	20	0.00%	6	57.14%
16	20	0.00%	8	50.00%
17	20	0.00%	15	11.76%
20	20	0.00%	21	-5.00%
24	20	16.67%	17	29.17%
25	20	20.00%	3	88.00%
28	20	28.57%	30	-7.14%
34	20	41.18%	17	50.00%
40	20	50.00%	23	42.50%
52	20	61.54%	21	59.62%
61	20	67.21%	35	42.62%
74	20	72.97%	25	66.22%
100	20	80.00%	45	55.00%
132	26.4	80.00%	30	77.27%
150	30	80.00%	24	84.00%
164	32.8	80.00%	52	68.29%
169	33.8	80.00%	25	85.21%
200	40	80.00%	100	50.00%
208	41.6	80.00%	20	90.38%
212	42.4	80.00%	44	79.25%
245	49	80.00%	34	86.12%
256	51.2	80.00%	60	76.56%

Table 1: Hypothetical Data Set for Example Analysis

Sign Test

The Sign Test is a nonparametric test that may be of use when it is necessary to know only if observed differences between two conditions are significant. That is to say, with appropriate use of the sign test, it would be possible to determine if X is really "more" than Y, however the conditions are arranged. The sign test is structured so that plus (+) and minus (-) "signs" are used to denote change in magnitude, as opposed to a quantitative measurement.

In a sign test, the concentration differences are calculated by subtracting the

Continued on page 33

Continued from page 31
observed from the expected. Positive numbers are then assigned a plus sign and negative numbers are assigned a negative sign. Differences of zero (i.e. Observed = Expected) are omitted.

The outcome of the number of points above and below the line is compared to a population when it is expected that half the points are above the line and half are below. Using a binomial distribution, the probability that the number of occurrences above (or below) the line, as explained by chance, is calculated. The probability is then evaluated to decide if the samples do or do not represent the PEF. There are three outcomes from this test.

1. The probability is high that the observed data match the expected.
2. The probability is high that the observed data do not match the expected and are greater (+).
3. The probability is high that the observed data do not match the expected and are lesser (-).

With Outcomes 1 and 2, the hypothesis that the BMP meets or exceeds expectations would be accepted, at least on a concentration basis. Outcome 3 indicates the BMP is below expectations and should be rejected.

$$P(X) = \frac{n!}{(n-X)!\,X!} \cdot p^X \cdot q^{n-X}$$

Where:
P(S) ....The symbol for the probability of success (+)
P(F) ....The symbol for the probability of failure (-)
p ....The numerical probability of a success (use 0.5)
q ....The numerical probability of a failure (use 0.5) (P(S) = p and P(F) = 1- p = q)
n ....The number of trials
X ....The number of successes (positives)

In the example there are a total of 25 samples. Of the 25 samples, 13 are above the line(+) and 12 are below (-). This indicates a 50 percent probability of occurrence, which clearly indicates this BMP is meeting expectations. As an example, however, let's assume that of the 25 pairs, there were 17 below the line and eight above. Then there is about a 5 percent chance of this occurring, which would lead to the conclusion that the BMP was not meeting performance expectations (http://home.clara.net/sisa/pairwise.htm).

Mass Load Balance Calculations

As mentioned previously, simply looking at the influent versus percent removal or influent versus effluent does not tell the whole story. These graphs convey no information on load reduction.

Load reduction evaluation is a quantitative method based on calculating both the expected load removal (expected concentration times the actual runoff volume) and the observed load removal. The difference between these two values represents a residual that can then be further analyzed. Table 2 shows these calculations.

Continued on page 34

Influent mg/l	Expected Effluent mg/l	Expected Percent Removal	Observed Effluent mg/l	Observed Percent Removal	Volume (liters)	Mass IN (g)	Effluent Mass Observed (g)	Effluent Mass Expected (g)	Mass Removed Observed – Expected (g)
6	20	0.0%	5	17%	2000	12.0	10.0	40.0	-30.0
10	20	0.0%	12	-20%	500	5.0	6.0	10.0	-4.0
12	20	0.0%	16	-33%	300	3.6	4.8	6.0	-1.2
14	20	0.0%	6	57%	500	7.0	3.0	10.0	-7.0
16	20	0.0%	8	50%	1500	24.0	12.0	30.0	-18.0
17	20	0.0%	15	12%	150	2.55	2.25	3.0	-0.75
20	20	0.0%	21	-5%	2000	40.0	42.0	40.0	2.0
24	20	16.7%	17	29%	800	19.2	13.6	16.0	-2.4
25	20	20.0%	3	88%	1900	47.5	5.7	38.0	-32.3
28	20	28.6%	30	-7%	350	9.8	10.5	7.0	3.5
34	20	41.2%	17	50%	800	27.2	13.6	16.0	-2.4
40	20	50.0%	23	43%	1100	4.0	25.3	22.0	3.3
52	20	61.5%	21	60%	5000	260.0	10.5	100.0	5.0
61	20	67.2%	35	43%	2000	122.0	70.0	40.0	30.0
74	20	73.0%	25	66%	5000	370.0	125.0	100.0	25.0
100	20	80.0%	45	55%	2000	200.0	90.0	40.0	50.0
132	26.4	80.0%	30	77%	1600	211.0	48.0	42.2	5.76
150	30	80.0%	24	84%	9000	1350.0	216.0	270.0	-54.0
164	32.8	80.0%	52	68%	3000	492.0	156.0	98.4	57.6
169	33.8	80.0%	25	85%	1800	304.0	45.0	60.8	-15.8
200	40	80.0%	100	50%	800	160.0	80.0	32.0	48.0
208	41.6	80.0%	20	90%	5000	104.0	100.0	208.0	-108.0
212	42.4	80.0%	44	79%	30000	6360.0	1320.0	1270.0	48.0
245	49	80.0%	34	86%	9000	2210.0	306.0	441.0	-135.0
256	51.2	80.0%	60	77%	2000	512.0	120.0	102.0	17.6

Table 2 – Mass Load Balance Calculations

Total Mass In (KG)	Total Mass Out (KG)	Total Mass Out Expected (KG)	Observed – Expected (KG)
13.83	2.93	3.04	-.011

Table 3 – Summary of Table 2

Continued from page 33

Note in this case that a negative number reflects a positive result. In other words, less mass left the BMP than expected, so one could conclude from a mass basis that the BMP met expectations as well.

Note that on a mass basis, the expected percent removal calculates to be 78 percent and not 80 percent. Clearly if the water was much cleaner with lower EMCs, the mass removal could be, say, 50 percent and still meet performance expectations; however, one may ask the question of how well the BMP would operate at higher concentrations, which would warrant additional samples at higher concentrations.

The load reduction assessment can be further refined if there is an infiltration component. If a fraction of the entire runoff volume is reduced through infiltration or evaporative processes, then the expected mass load would be a product of the (influent volume)x(Expected infiltration component)x(expected percent removal) and the actual mass load would be the (Effluent volume)x(effluent concentration).

This allows an assessment of how well the infiltration component is working, rather than assigning a simple percent which perpetuates the issue. One should use caution however because the infiltration capacity is most likely not constant and reduces over time with progressive loading.

Parameter	Influent (mg/l)	Effluent (mg/l)	% Removal	Expected Effluent mg/l
TSS	50	20	60	20
Total P	0.3	0.21	30	0.22
Ortho P	0.10	0.12	-20%	0.10

Table 4 – Summary of Example Total-P Performance Expectation Function

Comparison to the Expected Rainfall Distributions

Another issue about the use of storms is how they are distributed. Another way to misinterpret data is to not evaluate how the unit was sized as compared to the magnitude of the storms or storm flows that occurred. In most areas, one can use local rainfall data to construct a cumulative rainfall depth frequency curve or a cumulative flow duration curve. These curves can be used to adjust flow data (or runoff volume data) to normalize what actually happen during the monitoring period versus what would be expected to happen over a much longer period of time.

In most (if not all) cases, one would find that BMPs tend to work better during small storms (especially BMPs that rely on volume storage and settling), and one would also find that the highest frequency of storm occurrence is smaller storms. It therefore stands to reason that an additional weight should be added to the data set to provide an adjustment which weights the data to be more representative of what will statistically occur over a period of time, versus what just happened during the sampling period.

Analysis of Outliers

Analysis of the outliers can be done for both the concentration and load. One method is to analyze the residuals (observed minus expected) to determine if they are normally distributed about the mean, which in this case would be zero. Box and whisker plots can then be used to identify the points outside the second or third standard deviations.

Continued on page 39

Continued from page 34

Performance Expectation Functions for other Pollutants

A PEF can be constructed for other pollutants as well. In some cases the PEF may be more complex due to the more complex nature of the pollutant. Total phosphorus, for example, has a soluble component to it. Most BMPs do not address Ortho-P and in many cases can generate Ortho-P from the decomposition of organic matter. Typically the reduction of Total-P is associated with the organic and mineral phase of Total-P associated with the TSS (Wigginton et al, 2000). The soluble component adds a layer of complexity in that the higher the fraction of Ortho-P to the Total P, the BMP relative performance will significantly drop.

So, in the case of a PEF for Total-P, there could be two base lines. The first is the Ortho-P baseline and the second is the fraction of the particulate Total-P associated with the baseline TSS concentration. The PEF for the ortho fraction could be set to zero, and the particulate fraction could be then tied to the TSS removal or some function of the TSS removal.

For example, if an influent sample had 0.3 mg/l of Total P, of which 0.10 mg/l was Ortho-P, then the remainder could be associated with the TSS. If the influent TSS is at 50 and the expected percent removal is 60 percent (conservatively assuming a linear relation between TSS and TP), the removal expectation for the TSS fraction of the TP is 60 percent of 0.2 mg/l, which is 0.12 mg/l. This gives an expected effluent of (0.10 mg/l + 0.12 mg/l) = 0.22 mg/l. Thus, the expected percent removal is only 27 percent. In this case the observation was 0.21 mg/l; therefore, the BMP was exceeding expectations for TP, even though the Ortho-P fraction was elevated on the effluent side.

Conclusion

This method of analysis is relatively simple and does not use "heavy statistics"; however, it does provide a reasonable balance between the need to simply define expected BMP performance while taking into consideration much of the practical reality of how BMP's actually perform. This method takes into account both concentration and load and allows for a realistic comparison to expected performance that is characteristic of most accepted BMPs.

The use of the PEF also allows the regulatory agency to stipulate the expected BMP performance. This allows for a connection between the BMP performance and water quality needed to meet the water quality requirements for the receiving waters.

References

- de Ridder, Scott, Sean Darcy, James Lenhart (2002) Influence of Analytical Method, Data Summarization Method, and Particle Size on Total Suspended Solids (TSS) Removal Efficiency. 9th International Conference on Urban Drainage. Report No. PE-C063. Portland, OR.
- Lehman, Jeremiah and Scott de Ridder (2005) Predicting Solids Concentrations from Storm Event Variables, StormCon, Orlando, 2005.
- Schueler, Thomas (1996) Irreducible Pollutant Concentrations Discharged from Urban BMP's, Watershed Protection Techniques, Vol 2, No 2, Spring 1996, page 75.
- Strecker, Eric, Marcus Quigley, Ben Urbonas and Jonathan Jones (2004) Stormwater Management, State-of-the-Art in Comprehensive Approaches to Stormwater, The Water Report, Issue #6, August 15, 2004.
- Wigginton, Byran O., James Lenhart (2000). Using Iron-Infused Media and StormFilter Technology for the Removal of Dissolved Phosphorus from Stormwater Discharges. Water Environment Federation – 73rd Annual Conference and Exposition. Anaheim, CA.