Construction General Permit – Stormwater Deadline: 5/4/07 5pm

WSPA

Western States Petroleum Association
Credible Solutions • Responsive Service • Since 1907

Kevin Buchan
Senior Coordinator, Bay Area and State Water Issues

RECEIVED
MAY - 4 2007
SWRCB EXECUTIVE

May 4, 2007

Chair Doduc and Members of the Board
State Water Resources Control Board
P.O. Box 100
Sacramento, CA 95812-0100

Re: WSPA Comments on the Preliminary Draft General NPDES Permit for Construction Activities

Ms. Doduc, and Members of the Board:

The Western States Petroleum Association (WSPA) is a trade association that represents the companies and other entities that conduct most of the petroleum-related operations in the western United States. These operations include production, transportation, refining and marketing of petroleum and petroleum-based products. WSPA member facilities operate under National Pollutant Discharge Elimination System ("NPDES") permits for stormwater discharges, especially the General Industrial Permit and the General Construction Permit. WSPA members have a keen interest in State Water Board's (SWRCB's) draft construction permit, and its interpretation and application of permit requirements. We appreciate the opportunity to submit the following comments on the preliminary draft general permit.

We would like to thank the Board for its effort to allow stakeholder participation in the permit's development by asking for stakeholder input on the "preliminary" draft. The draft permit contains many new and significant elements that we believe require careful consideration. We applaud your efforts to solicit early feedback and comment on those issues.

Statewide Stormwater Policy

We are disappointed that staff seems to have abandoned development of a comprehensive statewide stormwater policy in favor of continuing to set policy on a permit-by-permit basis (Fact Sheet, page 21, Section III.A.3). We vehemently disagree that a permit-by-permit approach will lead to a "statewide stormwater policy at a lower cost and in less time." WSPA believes that a comprehensive statewide stormwater policy is important and needed. General permits that address specific types of discharges cannot address the myriad of issues associated with stormwater runoff and its potential regulation under other state programs, such as the Ocean Plan, the NPS program and the ASBS program, to name just a few. A statewide policy is needed to cover stormwater regulation under these and other programs. It is premature for the State Board to disregard the benefits of a statewide stormwater policy and piecemeal policy through the general permits. Important issues such as the use of numeric effluent limits within stormwater permits need to be addressed in the broader context of a statewide policy.

Numeric Effluent Limits (NELs)

The preliminary draft permit proposes NELs for TPH, pH and TSS for discharges from Active Treatment Systems (ATS). Even though it proposes relatively few technology-based NELs, the proposal to use NELs is inconsistent with the recommendations provided by the Blue Ribbon Panel in the report entitled “The Feasibility of Numeric Effluent Limits Applicable to Discharges of Storm Water Associated with Municipal, Industrial and Construction Activities” dated June 19, 2006 (Blue Ribbon Panel Report). The draft permit does not address the reservation and concerns expressed in that Report related to the development and use of NELs. In particular, the use of design storm criteria (to determine when NELs apply and when they do not) is not addressed at all in the draft permit. WSPA strongly recommends that the ATS requirements and the associated numeric limits be removed from the preliminary draft permit and that this permit-term be used develop the appropriate protocols and data to support technology-based NELs in the next round of permits, should they be deemed necessary.
Attached are the September 1, 2006 letters from Flow Science and Professor Gary Lorden to the SWRCB on the issues and difficulties related to the development and implementation of scientifically-based Technology Based Effluent Limits. The two letters recommend that the SWRCB undertake a multi-year, well designed data collection effort to develop the appropriate data needed for workable NELs. We urge you to delete the ATS and NEL requirements in the preliminary draft permit and consider implementing the recommendations in those two attached letters.

Toxicity Testing

WSPA does not believe toxicity testing, especially chronic toxicity is appropriate in the stormwater program. Currently toxicity testing may take 7-10 days to complete, way beyond any storm duration. WSPA urges that toxicity testing be deleted from the permit and suggests that a research/pilot program be initiated for toxicity testing to evaluate its potential use in the next permit.

Action Levels

The Blue Ribbon Report suggested that the SWRCB use Action Levels as a way to allow the SWRCB to learn and gather data to support NELs for the future. WSPA endorses this approach. WSPA believes that if Action Levels were developed, set and used consistent with the recommendations of the Report, Acton Levels in the permits could be effective in improving stormwater quality. But the use of Action Levels in the draft permit must be consistent with the Report’s recommendations. In particular, they must be used to identify “bad actors” and an appropriate monitoring program must be developed and implemented so that Action Levels can be used for their intended purpose. The September 1, 2006 Flow Science letter includes a number of recommendations for the development and use of ALs that we urge the board consider for the next permit.

Regional Board Approvals

WSPA is very concerned that the preliminary draft permit lays out an elaborate regional board "approval" process that is unworkable. WSPA urges the SWRCB to clearly state in the permit, that the permit is effective once all the required documentation is submitted. The heart of the permit is the Storm Water Pollution Prevention Plan (SWPPP) -- a living document that must be constantly upgraded and modified to reflect construction progress and stages. Requiring formal board approval of the SWPPP will make the permit unworkable. For the system to work, the board must rely on the competence and integrity of stormwater professionals that certify the SWPPP (in fact the preliminary draft permit enhances the qualifications for these stormwater professions), of course verified by board inspections as required by law.

WSPA also directs your attention to the CASQA letter on this preliminary draft permit, The CASQA letter addresses a number of other issues important to WSPA, WSPA urges the board to work with CASQA, WSPA and other stakeholders as the board moves to the next draft of the construction permit.

Thank you for considering WSPA's comments. Please contact me at 916-498-7755 if you have any questions or wish to discuss our comments. Thank you.

s/Kevin Buchan
(sent via email)

Enclosures: Flow Science, letter dated September 1, 2006
Lorden, letter dated September 1, 2006

Flow Science Incorporated
723 E. Green St., Pasadena, CA 91101
(626) 304-1134 • FAX (626) 304-9427

September 1, 2006

Ms. Song Her, Clerk to the Board
State Water Resources Control Board
P.O. Box 100
Sacramento CA 95812-0100

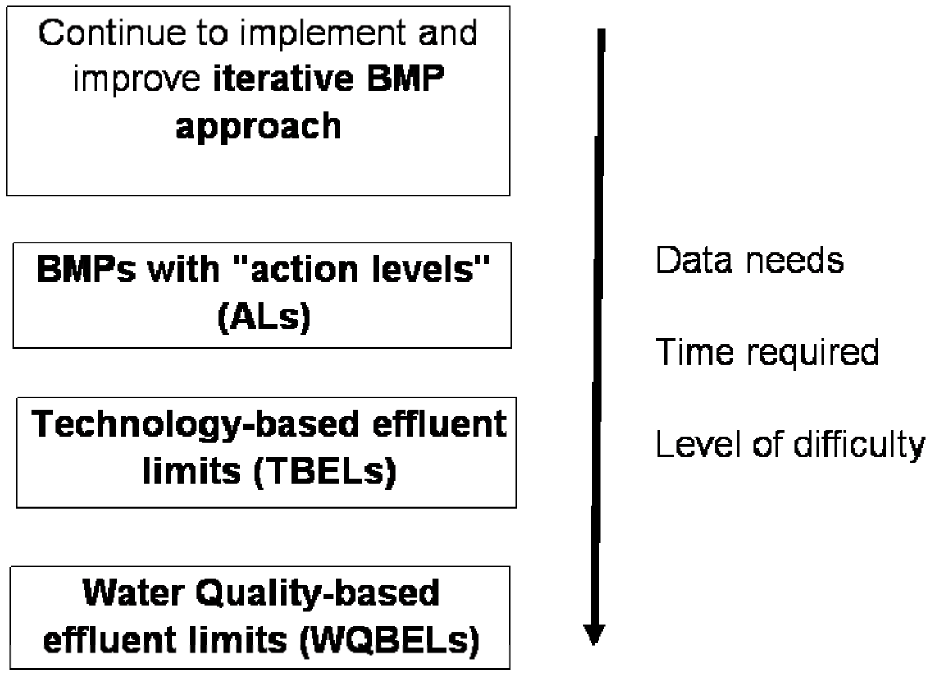
Re: Comments on the Storm Water Panel of Experts Report entitled *The Feasibility of Numeric Effluent Limits Applicable to Storm Water Discharges*
FSI 044018.2

Dear Ms. Her:

Flow Science was retained by the Western States Petroleum Association (WSPA) to comment upon the State Board Expert Panel ("Expert Panel") report entitled *The Feasibility of Numeric Effluent Limits Applicable to Storm Water Discharges*. I also provided testimony to the State Board on July 21 and July 28, 2006, and incorporate those presentations by reference. We appreciate the opportunity to comment on this report and look forward to working with the State Board as it addresses questions related to improving storm flow water quality. Many of our comments are general, and apply broadly to the municipal, industrial, and construction sectors, while others are more tailored to storm flow discharges from industrial facilities.

As detailed in Figure 1, there are four major options for the regulation of storm water. The types of data and the amount of time required for use in the State's storm water regulatory program will depend upon the type of limit to be developed, the methodology used to establish numeric limits, and the monitoring and compliance strategies to be used both to establish datasets upon which limits can be based and to evaluate compliance and improvements in water quality as a result of program implementation. The comments below first discuss major issues common to all options: storm flow characteristics, the utility of the existing dataset, and program design considerations. Finally, each type of limit is discussed in turn, including the amount and type of data and estimated time frames that would be required to develop each type of limit.

Figure 1. Options for storm water regulation.



Storm flow characteristics. Storm flows are quite different from many other types of discharges, particularly in the arid west. Most notably, storm flows exhibit highly variable flow rates, flow volumes, and constituent concentrations. Storm flow water quality is a complex function of watershed size, slope, soils, vegetation types, rainfall (storm size and intensity), antecedent conditions (a function of the time since last rainfall), land use, and climate.

Available data demonstrate that storm flow constituent concentrations can vary by an order of magnitude or more on timescales of an hour or less (see Flow Science, 2005). Constituent concentrations can also vary just as widely between storm events, and at any given time between relatively closely located sites. Analysis of existing data demonstrate quite clearly that storm flow constituent concentrations do not follow a neat, "log-normal" statistical model (see separate analysis by Dr. Gary Lorden). This is important because the procedures the State and Regional Boards currently employ to develop numeric limits for non-storm flow discharges rely upon the assumption that data are log-normally distributed. For storm flow data, this assumption is incorrect, so that new methodologies will be needed to develop numeric limits (especially WQBELs, as discussed below). Importantly, these methodologies will need to develop means to account for extreme events (e.g., high rainfall intensities, changed site conditions, etc.)

that can result in measured storm water concentrations that fall outside of the "normal" range of observations.

Constituents enter storm flows from a variety of sources, including both natural sources (site soils, airborne dust, wildfire ash, combustion products, etc.) and manmade sources (atmospheric deposition of anthropogenic origin, such as automobile exhaust and road dust, building materials, site activities and practices, application of pesticides, etc.). Available data suggest that, for many constituents, the bulk of the storm flow concentration and loading may arrive from atmospheric deposition (see, e.g., Sabin et al. (2004), Sabin et al. (2005)). These considerations are important because it may prove easier and more cost-effective to control pollutants at the source rather than to treat and remove pollutants from storm flows.

Strategies available to improve storm water quality range from best management practices (BMPs) to storage and treatment approaches. All approaches are challenged by the high volumes and flow rates of storm flows, which necessitate hydrologic design criteria, such as a "design storm" or other hydrologic specifications. As noted by the Expert Panel, exceedances of limits can be expected to occur several or more times per year, based largely on hydrologic considerations alone.

Existing data. Most of the available data on storm flow quality, both from individual sites and in receiving waters, are in the form of a single grab sample per storm event, and generally for a relatively limited number of constituents. Thus, it has not been possible to date to develop relationships between parameters that affect storm flow quality (rainfall amount and intensity, antecedent conditions, site conditions, etc.) or to predict or explain the full range of variability observed in storm flows.

Data have been collected as required by the State's General Industrial Permit for four constituents: total suspended solids (TSS), conductivity, oil and grease or organic carbon, and pH. These data are in the form of grab samples collected once per storm. There is little information in the database on storm size or intensity, site conditions, BMP and treatment measures in place, and other factors that affect storm water constituent concentrations. To date, there has not been a broad, controlled program of data collection that would allow us to compare water quality concentrations between facility types, regions, or in response to hydrologic influences.

Data are available for additional constituents from a small sampling of individual facilities, but are generally in the form of grab samples. Very few data are available to describe variations in concentrations during a storm or in the form of event mean concentrations (EMCs, or composite samples), and, to our knowledge, the few data that are available do not represent discharge water quality but rather were collected interior to a site. To implement, for example, CTR criteria in the form of numeric limits applicable to storm flows would require data on a similar temporal scale to the objectives (e.g., acute criteria are expressed as one-hour averages, requiring data on a one-hour or shorter timescale; see also comments by Dr. Gary Lorden).

The degree to which the variability of storm flows must be characterized will depend upon the type of limit to be adopted, and the actions triggered by observed exceedances of those limits. For this reason, fewer data would be required to establish ALs than for TBELs, and fewer data would be required for TBELs than for WQBELs.

Program design considerations. As described below, existing data may be sufficient to establish Action Levels for a handful of constituents, including those for which data are available as a result of data collection undertaken pursuant to the State's General Industrial Permit (TSS, conductivity, oil and grease, organic carbon, pH). Additional study will be required to determine if these data are sufficient for this purpose, and the answer may depend upon the way in which Action Levels are to be established and how they are to be used.

In any case, future data collection would be required to establish numeric limits or other quantitative measures of compliance. The type and quantity of data to be collected are very much dependent upon the type of limit to be developed, the methodology to be used to compute limits, and the monitoring and compliance strategies to be used after limits are established. For example, if compliance with numeric limits is to be determined using grab samples, then the data collection effort necessary to develop limits would likely be more data-intensive, so that those grab samples can be related to variations in concentration within a storm or to EMCs (flow-weighted composite concentrations).

Options for storm water regulation.

Continue to implement and improve the iterative BMP approach. As shown in Figure 1, the first option is to continue to implement and improve the existing approaches to managing storm flows using an iterative BMP process. As noted by the Expert Panel, improvements can be made in this process, including utilizing BMP performance data and knowledge about the impairments or constituents of concern in a receiving water to select better and more efficient BMPs. With this option, compliance and enforcement would be based upon selection of appropriate BMPs, then continued implementation and maintenance of the selected option. Examples of additional data and information that could be collected to improve the iterative BMP approach include:

- Development of a list of BMP options
- Data collection and research into BMP unit design and efficiency
- BMP design criteria (a "design storm" or other hydrologic design criteria)
- Information on gross receiving water quality (identification of constituents of concern and flow characteristics, etc.)
- Detailed analysis of maintenance and enforcement options

This program could begin immediately. In effect, the program itself could be iterative, with improvements made pursuant to a coordinated, well-designed program of data collection and subsequent development of program guidance, likely at the direction of the State Board.

BMPs with "Action Levels" (ALs). As envisioned by the Expert Panel, ALs would serve to identify "bad actors" (those discharges or sites with a propensity, based on monitoring data, to contribute disproportionately to high concentrations of constituents in receiving waters) and to trigger an iterative management approach. The Expert Panel presents three options for the development of ALs. To proceed with implementation of ALs, Flow Science would recommend that the State Board would first identify both the methodology to be used to establish ALs and to clarify how measurements would be compared to ALs and the actions that would be triggered. As discussed above, existing data in the database compiled pursuant to the General Industrial Permit may be sufficient to establish ALs for those limited constituents, again depending upon the methodology to be used. Examples of additional data that would be required to develop and implement ALs are:

- Development of a list of BMP options
- Data collection and research into BMP unit design and efficiency
- BMP design criteria (a "design storm" or other hydrologic design criteria)
- Information on gross receiving water quality (identification of constituents of concern and flow characteristics, etc.)
- Process and procedures for establishing ALs
- Actions required when ALs are exceeded at a certain frequency
- Data on effluent constituent concentrations for those constituents that will have ALs

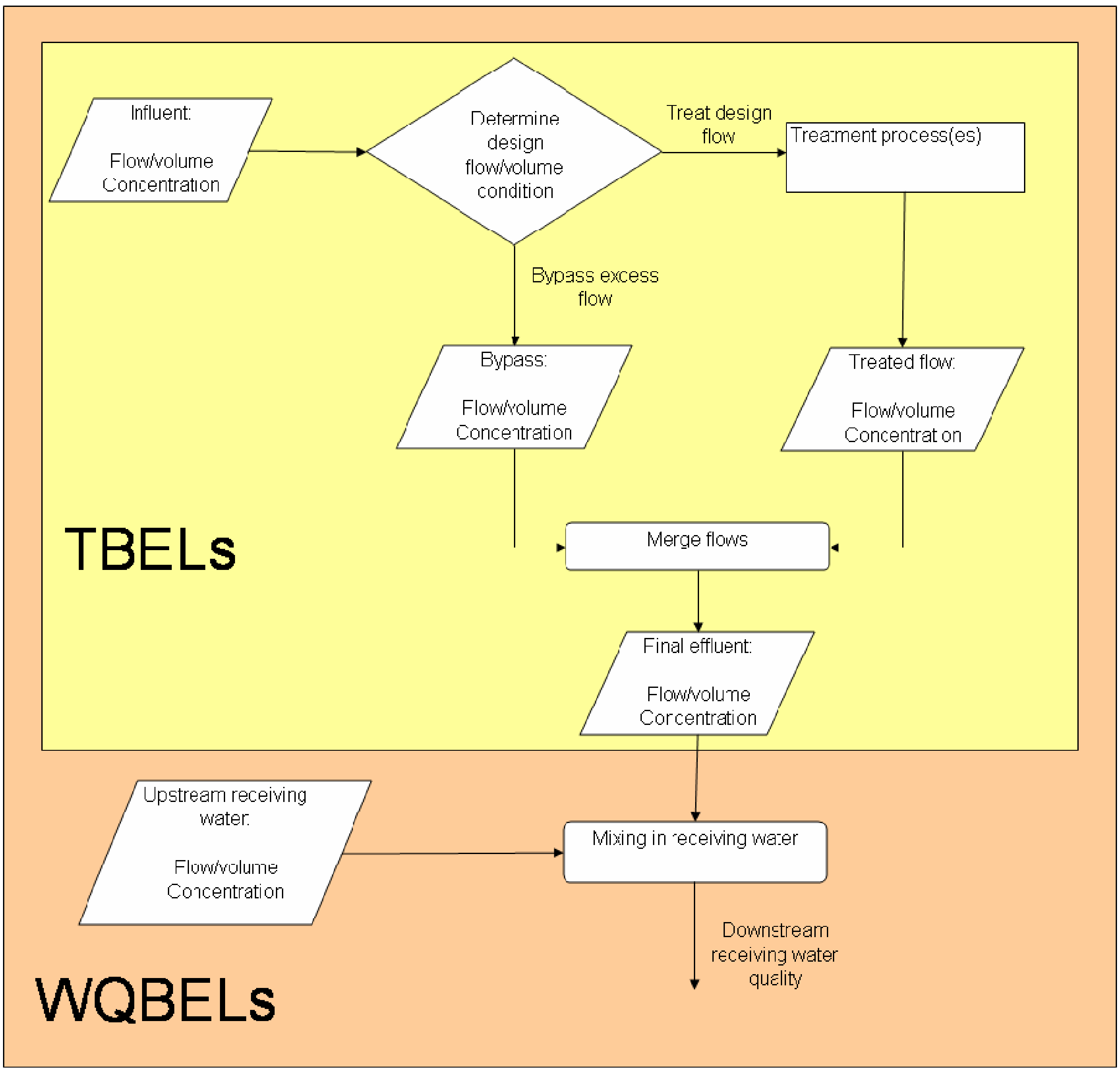
AL development could be staged, so that – if data are sufficient – ALs would be developed in the near-term for industrial discharges for some subset of the four constituents (TSS, conductivity, oil & grease or organic carbon, and pH) for which grab sample data are available. A second phase could involve collection of data for additional constituents and development of ALs for those constituents. We estimate that development and implementation of ALs could take from 0-3 years for constituents with readily available data, and 3 or more years for constituents for which additional data collection will be required.

Technology-Based Effluent Limits (TBELs). TBELs are numeric limits based upon available technology and the treatment efficiency of those technologies. For storm flows, TBELs would need to be developed in consideration of the volume or flow rate to be treated, the efficiency of the treatment process, and the quality of storm flow influent to the treatment process. As shown in Figure 2, the final effluent stream will be a mixture of treated effluent and untreated effluent (i.e., effluent beyond the hydraulic capacity of the treatment system). Data requirements for TBELs may include:

- Detailed characterization of influent (raw) water quality

- BMP and treatment system performance data, which would be required for a range of influent concentrations and under field, not laboratory, conditions
- Process for setting TBELs that would recognize the variability of storm water flow rates/volumes and constituent concentrations
- Monitoring and compliance options (e.g., grab v. composite samples, sampling frequency, etc.)

Figure 2. Considerations in development of TBELs and WQBELs.



An important consideration with TBELs is that an initial dataset used to establish limits may not capture the full range of conditions that may occur during the life of a project. Any TBELs, or the compliance and enforcement program associated with TBELs, would need to define both an allowable frequency of exceedance and a process for handling water quality excursions due to extreme events. Based on available information, we estimate that a total of at least 4 to 6 years would be required to establish appropriate TBELs. This estimate is derived as follows: we estimate that 1-2 years would be required to design the data collection program and to develop the methodology for calculating TBELs; a minimum of 2-3 years would be required for data collection; and at least one additional year would be required to calculate limits and place them into permits. Design and construction of controls and treatment systems may require additional time. Note that these timelines would depend upon the process used for limit development and could further be influenced by the availability of funding for monitoring, development of work groups, advisory committees, and peer review and notice/workshop/hearing processes, among other factors.

Water Quality-Based Effluent Limits (WQBELs). WQBELs must consider both effluent and receiving water quality in the limit development process. To date, the methods available for developing WQBELs are based on relatively simple, idealized data distributions – e.g., normal or lognormal data distributions. However, as discussed above, storm flow data do not follow these idealized distributions, but rather are "heavy-tailed" or "extreme value" distributions. For these reasons, existing WQBEL methodologies are inappropriate for storm flows, and new methodologies must be developed. As shown in Figure 2, variability in effluent quality and flows and variability in receiving water quality and flows all contribute to final receiving water quality, and these variables are a function of both time and space. Either dynamic modeling or statistical approaches could be considered to incorporate these considerations into limit calculation procedures, as described briefly in EPA's Technical Support Document (U.S. EPA, 1991). Data that would be required to calculate WQBELs may include:

- Detailed (hourly or sub-hourly) effluent quality and flow data
- Detailed receiving water quality and flow data
- Information on the means of compliance to be employed
- Methodology for determining reasonable potential ("RPA") and for calculating effluent limits
- Development of monitoring strategies and enforcement options
- Means to relate TMDLs to WQBELs

Note that it is usually envisioned that TMDLs would be implemented in permits as WQBELs. However, all the same considerations would apply to WQBELs calculated from TMDLs as to WQBELs calculated in water bodies or for constituents where TMDLs have not been established. In other words, concentration-based TMDL allocations should not be inserted directly into permits as numeric limits; rather, the same calculation methodology used to establish WQBELs should apply to development of WQBELs based on TMDLs.

We estimate that at least 7-10 years would be required for WQBEL development, which was derived as follows: ≥ 2-3 years to design the program and develop the methodology for calculating limits; ≥ 3-5 years for data collection; ≥ 2 years to calculate and implement limits. Again, design and construction of controls could require additional time.

Note that all of the above timelines are our best estimates, and they would depend upon the process used for limit development and could further be influenced by the availability of funding for monitoring, development of work groups, advisory committees, and peer review, and notice/workshop/hearing processes, among other factors.

We appreciate the opportunity to comment on this issue, and we look forward to working with the State Board in the future. Please contact us if you have any questions.

Sincerely,

Susan C. Paulsen, Ph.D., P.E.
Vice President and Senior Scientist

References

Flow Science, 2005. "Storm Water and Best Management Practices Analysis."

Sabin, Lisa D., Schiff, Kenneth C., Lim, Jeong Hee, Stolzenbach, Keith D., 2004. "Atmospheric dry deposition of trace metals in the Los Angeles coastal region." Southern California Coastal Water Research Project Biennial Report 2003-2004.

Sabin, Lisa D., Lim, Jeong Hee, Stolzenbach, Keith D., Schiff, Kenneth C. 2005. "Contribution of trace metals from atmospheric deposition to storm water runoff in a small impervious urban catchment." Water Research, Volume 39 (2005), pp. 3929-3937.

U.S. EPA, 1991. Technical Support Document for Water Quality-based Toxics Control. PB91-127415, March.

September 1, 2006

Ms. Song Her, Clerk to the Board
State Water Resources Control Board
P.O. Box 100
Sacramento CA 95812-0100

Re: Comments on statistical aspects of Panel of Experts report "**The Feasibility of Numeric Effluent Limits Applicable to Storm Water Discharges**"

Dear Ms. Her,

I was retained by Flow Science in connection with their work for the Western States Petroleum Association and was asked to evaluate and comment upon statistical issues associated with the formulation of numerical limits for constituent concentrations in stormwater effluents.

My qualifications.

My professional background as a statistician began with my research specialization in the field for my Ph.D. in mathematics from Cornell in 1966. Since that time, I have been continuously engaged in research and teaching of statistics at Northwestern University, UC Berkeley, and Caltech, where I have been Professor of Mathematics since 1977 and was department chair from 2003 to 2006. I am a fellow of the Institute of Mathematical Statistics, and have been active for the last thirty-five years as a statistical consultant for Caltech colleagues and for various governmental agencies and private companies. I have also served as a statistical expert witness in a variety of legal and regulatory matters, including statistical issues of water quality.

Variability of pollutant concentrations in stormwater flows.

It is clear from stormwater datasets I have examined in my statistical consulting that the pollutant concentrations associated with stormwater flows are highly variable, even over short time scales. Within a given hour the measured concentrations of pollutants in grab samples can be expected to vary substantially in relation to the mean for that hour. Therefore the probability that a single-grab-sample-per-storm monitoring system will accurately reflect the true impact of effluents on receiving waters is low. Moreover, the application of numerical effluent limits to grab-sample data is inherently less effective from a statistical point of view than the use of composite samples, for example. Because CTR criteria are specified as one-hour (or longer) average concentrations, it is essential to consider the additional variability of effluent concentrations that occur on a sub-hour basis. In particular, it is important to recognize the fact that any numerical limit applied to grab samples inherently imposes a smaller numerical limit on hourly averages.

Sources of Data Variability.

It is critical that numerical limits be derived with adequate consideration given to *all significant sources of variability of constituent concentrations in stormwater discharges.*

The sources of variability can be described in three categories:

1) Input Variables. These include influent characteristics, storm characteristics (e.g. rainfall intensity, and the rate and volume of flow), site-specific hydrologic features, and receiving water characteristics, such as those affecting dilution.

2) Treatment Characteristics. As discussed in the panel's report, the treatment capabilities of different facilities under different inputs vary widely, as do limits on their capacity for handling different flow rates and volumes.

3) Output Variables. Constituents in discharges vary in concentration in different parts of the effluent flow, the results of laboratory analysis are subject to measurement errors, and sampling techniques like one-grab-sample-per-storm introduce additional variability compared with hourly averages or storm composite concentrations.

Because of the systematic and widespread differences in these characteristics from facility to facility, storm to storm, and sample to sample, it is necessary to carry out an extensive and well-designed data collection effort at a representative set of facilities over a period of years. One or two years of data cannot represent the range of variability in number and severity of storms from year to year.

Unusual events and exceedance probabilities.

The report of the panel of experts contains repeatedly acknowledges the need to consider that storm water flows are dramatically affected by "unusual events". Here are a few examples:

- "...there is wide variation in stormwater quality from place to place, facility to facility, and storm to storm." (p.6)
- "Since the storm-to-storm variation at any outfall can be high, it may be unreasonable to expect all events to be below a numeric value." (p.6)
- "...several to more times each year, the runoff volume or flow rate from a storm will exceed the design volume or rate capacity of the BMP. Stormwater agencies should not be held accountable for pollutant removal from storms beyond the size for which a BMP is designed." (p.10)

- "The Panel recommends that Numeric Limits and Action Levels not apply to storms of unusual event size and/or pattern (e.g. flood events)." (p.18)

"Unusual events" aside, the statistical approach used in the State Implementation Policy (SIP) relies heavily on two features:

- the assumption that pollutant measurements in stormwater flows follow a lognormal distribution, and

- the idea of setting numeric limits for a facility by considering "exceedance frequencies" based upon calculations using the lognormal distribution.

The latter idea is revealed clearly on page 10 in Step 5, which discusses "a factor (multiplier) that adjusts for the averaging periods and exceedance frequencies of the criteria/objectives ...".

Clearly the use of "never to be exceeded" limits for stormwater effluents in permits needs to be eliminated. In light of the Panel's discussion and the statistical rationale used in setting limits, provision should be made for two kinds of exceedances—

- exceedances caused by carefully-defined "unusual events"—for example, storms whose severity and/or flow volumes exceed a "design storm", and
- "random" exceedances—resulting from the unavoidable fact that even ideal data, such as data from the assumed "standard", lognormal distributions, have some frequency of exceedance of any specified numerical limit.

Is the assumption of Lognormal Distributions valid? Analysis of datsets.

The most frequently used statistical model for datasets of stormwater constituent measurements in the lognormal distribution. In particular, the SIP relies heavily on the assumption that lognormal distributions adequately describe such data. It is therefore important to evaluate the validity of that assumption— i.e. to test it on actual data.

Flow Science provided me with three datasets containing grab sample measurements of copper at three outfalls of a California facility.I analyzed the datasets A, B, and C to determine whether the *maximum value* in each sample is too large to be reasonably explained by the lognormal distribution best fitting the data as a whole. In my experience with stormwater datasets, the largest value is "too large", indicating that the so-called *right-hand tail* of the actual data distribution is *heavier* than it would be if the data distribution were lognormal. In other words, the largest values that are found in datasets are not explained by the "general shape" of the lognormal distribution.

The following table gives the results of my analysis.

Summary Statistics for Copper datasets at 3 Outfalls
(water quality objective= 14 mg/L)

	Outfall A	Outfall B	Outfall C
Sample Size	23	32	20
Sample Median	2.8	2.8	3.2
Sample Maximum	55	12	39
p-level*	.003	.012	.110

*defined below

The p-level is a quantity used by statisticians to measure the *statistical significance* of a finding obtained from data analysis—in this instance, that the maximum value in each of the three datasets is large in relation to the sample as a whole. It is calculated as the probability that such an extreme value would occur "due to chance" *assuming that the data do come from a lognormal distribution.*

For Outfall A, the p-level is .003, indicating that a value of the sample maximum as extreme as the one observed (in relation to the overall dataset) would occur only about 3 times in 1000 samples. Similarly, for Outfall B, the p-level of .012 indicates that a maximum value as high as 12 (in relation to the other values) would occur only about 12 times in 1000 samples. These two results are what statisticians call "highly statistically significant", meaning that they provide strong evidence that the hypothesis being tested (in this case, that the true data distribution is lognormal) is not true. The p-level for Outfall C is .110, indicating that a maximum as large as 39 (in relation to the other measurements from Outfall C) would occur in about 1 sample out of every 9. This is what statisticians consider "marginally significant".

Taken as a whole, these three analyses indicate strongly that the lognormal distribution does not describe the behavior of copper measurements at these outfalls, in that the largest values in samples of size 20 to 30 are "too large to be explained by the lognormal distribution". Since the setting of numerical limits for constituent concentrations in stormwater effluents, as well as the monitoring of compliance with them, is mainly concerned with the behavior of the largest values in datasets, this apparent failure of the lognormal distribution to describe that behavior seriously undermines the statistical methods used in the State Implementation Policy (and elsewhere) to establish numerical limits.

How much data is needed to set numeric limits? An example.

As the report of the panel of experts and the State Implementation Policy (SIP) statistical methods make clear, setting numerical limits is critically related to controlling frequencies of exceedance—i.e. satisfying numerical criteria or objectives.

It is very important to recognize that the amount of data needed to determine frequencies of exceedance reasonably accurately and with high confidence is large—larger than would be expected on "common sense" grounds.

For example, suppose one takes a sample of n measurements designed to "demonstrate with 95% confidence that the value L is exceeded at most 5% of the time". (Here L is arbitrary and the statement is equivalent to saying that "the 95^{th} percentile of the true distribution of the data is L or smaller".) If n=153 (say), then it turns out that that confidence statement will be true *provided that at most 3 exceedances occur.*

Since 3 is about 2% of 153, we note that in order to *demonstrate* that the true exceedance percentage is at most 5%, we need to have exceedances of 2% or less in the sample. This is because we want to have *95% confidence* in the conclusion that the required 5% rate is truly satisfied.

Moreover, if we ask "How low does the true exceedance percentage have to be so that we can have 90% confidence that there will be at most 3 exceedances in the sample?", then the answer is "about 1.1%". If the *true* exceedance rate were 2%, we would get about 3 exceedances *on the average* in a sample of 153, but nearly half the time we would "flunk the demonstration" by getting more than 3 exceedances. Thus, one needs an "extra margin of safety"— in this example, a 1.1% true exceedance rate—to have a 90% probability of a successful *demonstration* that L is exceeded at most 5% of the time.

This disparity between *having* a low exceedance rate and *demonstrating* a low exceedance rate I call the "Caesar's wife effect". To avoid the *suspicion* of "having more than 5% exceedances" of a numerical limit, L, one must actually have a *true exceedance rate* much lower than 5%.

The sample size n=153 was used for this example. If a smaller sample size were used, the Caesar's wife effect would be even more pronounced—that is, either one would have to have an *extremely low exceedance rate* (a fraction of 1%) or else would have a probability larger than 10% of getting too many exceedances in the sample to succeed in demonstrating that the 5% objective is attained.

To perform more difficult analyses, such as testing the "fit" of distributions better suited than the lognormal to describe the data, requires even larger sample sizes. Investigation of a large body of data can shed light on the question of whether the lognormal distribution can be replaced by some other shape of distribution that better represents actual data. My expectation is that no *simple* family of distributions can represent adequately the range of behaviors of datasets. Accordingly I expect that the most useful statistical calculations will turn out to be based upon so-called *nonparametric* or *semiparametric* methods, relying more upon estimating from data the actual frequencies of high concentrations rather than upon estimating parameters such as the coefficient of variation.

Recommendations for developing a sound statistical basis for numerical limits.

To establish a statistically sound basis for setting numerical limits for storm water effluents, I believe it is necessary to carry out a well-designed data collection effort at a representative set of facilities over a period of years sufficient to incorporate the substantial year-to-year variability in the number and severity of storms.

Investigation of a large body of data can shed light on the question of whether the lognormal distribution can be replaced by some other shape of distribution that better represents actual data. My expectation is that no *simple* family of distributions can represent adequately the range of behaviors of datasets. Accordingly I expect that the most useful statistical calculations will turn out to be based upon so-called *nonparametric* or *semiparametric* methods, relying more on estimating actual frequencies from data than on estimating parameters like the coefficient of variation. It is an inescapable "statistical fact of life" that substantial sample sizes are required to determine appropriate numerical limits and to monitor compliance with them.

Sincerely,

Gary Lorden

Gary Lorden