Helpful Practices for Addressing Point Sources and Implementing TMDLs in NPDES Permits

Prepared by EPA Region 9 June 2015

Forward

Based on "lessons learned" over the past 25 years in developing and implementing Total Maximum Daily Loads (TMDLs), this document describes approaches and practices for developing and interpreting TMDLs in ways that facilitate incorporation of appropriate requirements in National Pollutant Discharge Elimination System (NPDES) permits. We hope the suggestions in this document help practitioners avoid unintended mistakes that are often repeated and can be easily avoided. TMDL writers should confer with NPDES permit writers early and often in the TMDL development process to ensure they can be properly implemented through permits. Also, permit writers should confer early and often with TMDL developers when interpreting existing TMDLs to better understand wasteload allocations and underlying assumptions. Better communication will help reduce misunderstandings, disconnected information transfer, and most importantly, improve the effectiveness with which TMDLs guide implementation of permits. Clear expression of point source control requirements will help restore impaired waters.

This is not a U.S. Environmental Protection Agency (EPA) guidance document; nor does it represent official EPA policy. This document discusses practices and ideas that, in the experience of staff in EPA Pacific Southwest Region (Region), can prove effective in integrating TMDLs and permits. The Clean Water Act (CWA) and associated implementing regulations contain the legally binding requirements associated with the development of TMDLs and NPDES permits. This document does not substitute for the CWA or associated implementing regulations. Approaches and practices identified in this document are not binding; the TMDL and permitting authorities may consider other approaches consistent with the CWA and associated implementing regulations. When EPA makes a TMDL or permitting decision, it will make each decision on a case-by-case basis and will be guided by the applicable requirements of the CWA and associated implementing regulations and information applicable to individual cases. This document is intended to be consistent with but does not modify existing EPA policy or guidance.

Table of Contents

INTRODUCTION

The Total Maximum Daily Load (TMDL) process provides a critical opportunity to assess the relative importance of pollutant sources and apportion pollution control burdens among various sources. NPDES permits are the main vehicles through which point source pollution control responsibilities identified in permits are implemented and enforced. As the TMDL and NPDES programs developed, they often operated independent of each other. Operating guidance and customary practices in each program have not always been well coordinated. As a result, we have observed that many TMDLs are written in ways that make them difficult to implement through permits and, conversely, many NPDES permits contain effluent limitations that do not accurately implement the intent of associated TMDLs.

This document describes approaches for improving effectiveness in developing TMDL wasteload allocations (WLAs) and implementing WLAs in NPDES permits. Based on "lessons learned" through our experience in Region 9 developing and reviewing TMDLs and NPDES permits, we identified practices that TMDL and NPDES permit writers may want to use and common mistakes to avoid. In addition, while the focus of this document is on point sources and the translation of TMDLs into NPDES permits, many "helpful practices" are also relevant to the development and implementation of TMDLs focusing on nonpoint sources that receive load allocations (LA). We've included appropriate TMDL and permit examples that illustrate these practices.

Our key "lesson learned" is that WLAs can be difficult to translate accurately and effectively into numeric water quality based effluent limitations (WQBELs) in NPDES permits unless the WLAs are accompanied by interpretive information that specifies where, when, and how the WLAs should be incorporated into NPDES permits. This document is roughly divided into two sections. Section A discusses practices for effectively addressing point sources when developing TMDLs and associated implementation plans. Section B discusses issues that arise in interpreting existing TMDLs when developing NPDES permits for discharges to waters for which there is an approved, applicable TMDL.

In 2013, EPA issued "A Long-Term Vision for Assessment, Restoration, and Protection under the Clean Water Act Section 303(d) Program." The new Program Vision details enhancements made to the CWA 303(d) Program informed by the experience gained over the past two decades in assessing and reporting on water quality and in developing approximately 65,000 TMDLs. It is designed to enhance overall efficiency of the CWA 303(d) Program, and in particular encourages focusing attention on priority waters and provides States flexibility in using other available clean water program tools beyond TMDLs to attain water quality restoration and protection. EPA recognizes it may be appropriate in some cases to use alternative restoration approaches to evaluate pollution problems in watersheds and devise solutions to those problems. These alternative restoration approaches are not "in lieu" of a TMDL. However, EPA recognizes that under certain circumstances there are alternative restoration approaches that may be more immediately beneficial or practicable to achieve water quality standards than pursuing the development of TMDLs in the near future. While these alternative restoration approaches may not incorporate all elements of formal TMDLs, they will often have similar characteristics. We believe the suggestions in this document will help ensure that such alternative restoration approaches effectively account for point source discharges and that the provisions of such alternative approaches are properly incorporated in NPDES permits. In considering whether alternative restoration approaches are appropriate to address situations in which point source discharges contribute to water quality threats or impairments, we recommend that practitioners carefully consider whether these alternatives will yield the legal and technical rigor necessary to properly inform NPDES permitting. If alternative approaches are used that do not specifically articulate point source control approaches necessary to protect and restore water quality, it will be necessary during the permits process to fully evaluate whether each NPDES permitted

source has reasonable potential to contribute to water quality standards exceedences, and how permit requirements should be structured to ensure that applicable standards are met.

Here are several important terminology reminders to help avoid future problems:

a. We refer to TMDL "attainment" or "achievement" and NPDES permit "compliance" since permits, unlike TMDLs, are directly enforceable. TMDLs guide implementation actions but are not directly enforceable.

b. TMDLs and associated implementation provisions cannot directly incorporate compliance schedules. While TMDLs can identify appropriate implementation timeframes, "compliance schedules" can be implemented only through NPDES permitting decisions. The permitting decision record must demonstrate that compliance schedule regulatory provisions are met for that permit, regardless of what the TMDL might say about implementation timeframes. Before compliance schedules may be implemented in NPDES permits, their use must be authorized pursuant to CWA section 303(c) and approved by US EPA. Moreover, compliance schedules cannot be used to provide time to develop or revise a TMDL, a water quality standard, or a mixing zone analysis.

A. NPDES PERMIT-FRIENDLY TMDL DEVELOPMENT

Over the past 25 years, both TMDL development and NPDES permitting have evolved and become more complex. TMDLs are addressing increasingly complex watershed settings and pollutant or stressor problems, and the challenge of addressing nonpoint sources through TMDLs has become a high priority. The range of discharges requiring NPDES permits has expanded beyond traditional wastewater facilities, and some discharges are increasingly being addressed through general permitting approaches. Both TMDL developers and NPDES permit writers have learned valuable lessons on how to meet each others' needs while improving the use of TMDLs to strengthen control of discharges from point sources. Section A discusses lessons learned and recommends practices to help ensure that TMDLs provide the critical information that is needed in order to support the issuance of effective permits.

1. Address All Point Sources in the Watershed

Some TMDLs do not include wasteload allocations (WLAs) for all NPDES permitted discharges in the area covered by the TMDL. In some cases, the TMDLs mention a facility or permit but do not provide a WLA. It is very difficult for a permit writer to address a facility discharging to the TMDL water if the TMDL does not include a clear and source-specific WLA. Moreover, in situations where a facility does not discharge the pollutant identified in the TMDL or discharges an insignificant amount of the pollutant, many TMDLs are unclear about how the NPDES permit should account for this situation. Here are some suggestions for details to include in the TMDL to address these situations.

a. Name All Permitted Discharges.

The TMDL should name all NPDES permitted discharges in the TMDL watershed. This includes all major and minor NPDES discharges, including discharges covered by individual and general NPDES wastewater, stormwater, construction, industrial, and other permits. For example, we have found many TMDLs ignore state and federal highways, which are often regulated under municipal separate storm sewer system (MS4) NPDES permits. Watershed scale TMDLs should generally account for all point sources that discharge to tributaries or upstream from the water segments targeted by the TMDL. If the

TMDL does not specifically address all point sources that discharge to tributaries or upstream of the TMDL segment or analysis area, the TMDL can incorporate an allocation or assumption regarding contributions of pollutant loads from upstream sources and discuss how their NPDES permits should address these pollutants.

b. Include NPDES permit number and the facility name.

The TMDL WLA should specifically include the NPDES permit number and facility name as they appear in the permit itself to ensure accurate incorporation of the TMDL's WLA when the associated permit is reissued or revised. If feasible, the TMDL can discuss specific WLAs for specific outfalls if the facility has more than one outfall, or suggest an approach to applying an aggregate WLA to a facility with multiple outfalls.

c. Include a specific WLA for each permitted facility.

In the case of facilities permitted by a general NPDES permit, express the WLAs such that they can be effectively implemented on a facility-by-facility basis. See Section A.2 below regarding disaggregated WLAs. For example, concentration-based WLAs are probably easiest to implement in situations where multiple facilities are covered by the same WLA and it is difficult to disaggregate WLAs by discharger. Another example would be to clearly explain how the load-based WLA was calculated (e.g., based upon XX mg/L TSS and YY discharge flow rates).

d. Account for permitted facilities that do not discharge the pollutant of concern or that discharge insignificant amounts of this pollutant in the TMDL.

A TMDL writer has several options for addressing a facility that does not discharge or is an insignificant discharger of a pollutant in comparison with other sources. The TMDL source analysis should account for all known point source dischargers, noting that some may not discharge the pollutant of concern or discharge insignificant amounts that would not need to be limited with a specific wasteload allocation in order to achieve applicable standards. Care should be taken in evaluating insignificant or "de minimus" discharges to ensure that they are really unimportant at all geographic scales and need not be limited. In cases where individual permitted facilities do not discharge or discharge insignificant amounts of the pollutants of concern, it greatly assists permit development if the TMDL specifies how the NPDES permits should account for these discharges. Potential options for addressing this situation include:

1. The TMDL can specify that a particular point source need not be addressed by a WLA or permit limitation (likely including monitoring requirements to help ensure the facility does not discharge the pollutant at significant levels in the future). In this case the TMDL would explain why no allocation is necessary for this facility.
2. The TMDL can specify that the permit for a facility should incorporate performance-based limitations to ensure its loading of the pollutant of concern does not increase in the future.
3. The TMDL can incorporate an explicit margin of safety (MOS) to account for all insignificant sources along with discussion of how this MOS may be available for use in calculating future permit limits (e.g. performance based limits).
4. The TMDL can incorporate a WLA of zero for facilities that do not discharge the pollutant of concern, in which case the associated permit would generally prohibit discharge of the pollutant.

Example TMDLs:

- Glen Annie/Los Carneros (CA) Nitrate TMDL (2014) – The TMDL source analysis explains that non-point sources are responsible for almost all nitrate loading and that point sources contribute insignificant loads. The TMDL narrative identifies these very small/insignificant point sources and provides a rationale for not including numeric WLAs, while acknowledging these point sources can continue to discharge nitrate at existing very low levels.

- Spokane River (WA) DO TMDL (2011) – This TMDL includes a WLA appendix to explain how the WLAs were developed and are intended to apply to permits, especially upstream waste water treatment plants (WWTPs) in Idaho. Also, seasonal WLAs were clarified on a permit-specific basis.

- Bear Creek (MO) Nutrients TMDL (2011) – The TMDL determined that one of four WWTPs in the watershed was the most significant (3.12 MGD) source of TN and TP and a WLA with loading reductions was assigned to this facility. The other WWTPs received WLAs equivalent to their existing permit limits as additional load reductions from these facilities were not needed to achieve the allowable TMDL.

- Shenandoah River (VA) Mercury TMDL (2011) – Minor municipal facilities and facilities discharging under general permits were considered insignificant sources of mercury and were not assigned wasteload allocations. WLAs were assigned to industrial and major municipal facilities. The TMDL states that the Virginia DEQ will reevaluate those NPDES permits with assigned WLAs to ensure compliance, and that NPDES permits should include the following provisions:

    - Additional monitoring requirements using low-level detection techniques; and
    - If such monitoring results show exceedences of the applicable standard, the permittee is required to submit for review and approval a Pollutant Minimization Plan (PMP).

- Kiskiminetas-Conemaugh River (PA) Metals TMDLs (2010) – The TMDL provides gross WLAs for "negligible" point source discharges that were assumed to be discharging below the applicable water quality standard. The gross WLAs were calculated based on application of water quality critieria and available information about facility flows. The TMDL lists specific permits covered under the aggregate WLA for negligible sources.

- Dumps Creek (VA) Sediment (2004; 2010) – The TMDL and approval was first issued in 2004. Several years later Virginia DEQ realized it had omitted 3 point sources and so it re-opened the TMDL to make a narrow revision to include these three small sources and their WLAs. The re-approval of TMDL occurred in 2010, just prior to re-issuance of the permits for those 3 point sources. The LAs were decreased to accommodate for the increased loading within the WLAs.

- Upper Moncacy (MD) Sediment TMDL (2009) – The TMDL addresses each of the 34 permitted facilities, including individual industrial facilities, POTWs, mining operations permitted under the state's industrial stormwater general NPDES permit and MS4s. Permits were grouped into process wastewater and stormwater sources of sediment. Based on an analysis of the permit information, the TMDL writer determined that the total permitted load from these process wastewater sources equaled 0.2% of the total watershed load. Because these sources were so small relative to other sediment loads in the watershed, the TMDL established WLAs for this group of permitted facilities based on existing loads.

- Guyandotte River (WV) Metals, Fecal Coliform and pH TMDLs (2004) – Based on the types of activities and the nature of their discharges, permitted non-mining sources as shown in Table 3-3 of the TMDL are believed to be negligible. In this TMDL, these minor facility discharges are assigned WLAs that allow them to operate under their current permit limits.

e. Account for NPDES discharges to tributaries.

Preferably, TMDLs will clarify how they apply to permitted discharges from facilities located in tributaries upstream from the TMDL study area. The Region recommends setting WLAs for all permitted facilities discharging to tributaries upstream from the study area unless available information indicates those facilities do not discharge the pollutant of concern at levels that could reach the downstream waters for which TMDLs are being established.

Example TMDLs:

- Chesapeake Bay TMDL (2011) – EPA Region 3, in collaboration with the Chesapeake Bay jurisdictions, established this federal TMDL which allocates wasteloads and loads to sources in the states and the District of Columbia, many of which are upstream from the Bay.

- Columbia River Basin (OR) Dioxin TMDL (1991) – The TMDL established WLAs for the chlorine bleaching pulp mills in the basin. This TMDL addressed pulp mills discharging to Columbia River as well as those on its tributaries; e.g., on the Willamette and Snake Rivers and those further north within Washington. Recent state-developed TMDLs have defined pollutant load allocations at the upstream stateline. While this newer approach reduces potential concern about a downstream state setting WLAs for facilities located in an upstream state, it reduces clarity about how upstream sources need to be specifically controlled to ensure the TMDL is met.

f. Clarify how to calculate a performance-based limit if a WLA is set at current performance.

Many TMDLs include numeric WLAs for less significant sources based on their "current performance." As there are many possible approaches to calculating "current performance," it is preferable for TMDL documents to specify how WLAs based on this concept should be translated into permit limitations.

Many permits interpret "current performance" in terms of an extreme upper range statistical value estimate of that performance(e.g., the 95th or 99th percentile performance level) to help ensure the limit is set at a level that is highly unlikely to be exceeded. While this approach reduces the likelihood that a facility will violate the "performance-based" limit, it may not be consistent with the method used in the TMDL to determine that current performance is an adequate limitation for the facility given its pollutant(s) contribution within the watershed in the first place.

TMDLs should provide clear direction on how to translate performance-based WLAs into permit limits in a manner consistent with how the TMDL load assessment was conducted. It may be clearest for the TMDL to set a specific mass and/or concentration-based value or set of values for a specific facility that reflects the specific load analysis conducted for the TMDL, with instruction to the permit writer to set the WQBELs to ensure these current performance levels are not exceeded in the future. Alternatively, the TMDL could set a particular percentile representation of current performance to serve as the basis for the permit limit. The percentile value should be set at a reasonable upper value (e.g. 90th percentile) but not at a value so extreme that it results in a limitation far less stringent than contemplated in the TMDL.

g. Consider setting WLAs at current performance if current loadings are lower than available loading capacity.

While it may be legally permissible under federal regulations to set WLAs and associated permit WQBELs at a level that is less stringent than current performance, it may be preferable for TMDL developers to maintain current performance levels if they are more stringent than might otherwise be established under a TMDL. This approach helps to ensure that antidegradation and antibacksliding requirements are met, and is consistent with overall CWA goals to reduce point source discharges. This may not apply in situations where a facility is experiencing growth in effluent flows or pollutant loads, in which case the TMDL may provide the appropriate mechanism for planning for discharge growth.

Example TMDLs:

- Rio Hondo (NM) Sediment/Nutrients (2005) – The TMDL cited the state's antidegredation policy and maintained existing loads in setting the WLA for the Village of Taos Ski Valley even though loading capacity calculation would have allowed an increase in phosphorus loads from the facility. It appears the extra loading capacity was put into an explicit margin of safety.

- San Diego Creek/Newport Bay (CA) Toxics TMDLs (2002) – Hydrodynamic and water quality models were used to estimate the existing loads vs. the calculated allowable loads using numeric targets. WLAs were set at the lower or more protective of the two values. See TMDL Appx. G.

2. Disaggregate WLAs As Much As Possible

a. Provide separate WLAs for each permitted discharge.

WLAs that apply to multiple dischargers can be difficult to implement unless they are scalable to individual discharge situations. Separate WLAs should be set for each permited discharger if at all possible. In the case of a general permit which covers a large number of facilities or sites that discharge to the impaired waterbody, the WLAs should be designed to clearly delineate responsibilities for different owners or operators and in different areas of watersheds addressed by TMDLs. Options for disaggregating WLAs at the individual discharge level include (in rough order of preference):

- By discharge facility, which can be outfall by outfall, or using representative outfalls;
- By jurisdiction;
- By sub-watershed;
- By land use type.

It may be efficient to express disaggregated WLAs based on an analysis of discharges from representative outfalls as surrogates for discharges from the entire jurisdiction, sub-watershed, or land use type represented by the representative outfall. For a concentration-based WLA, pollutant analysis results from the representative outfall would be used to set the WLAs applicable to a larger jurisdiction, subwatershed, or land use type represented by that outfall. For mass-based WLAs, pollutant loading analysis for a representative outfall can be used in a similar manner after weighting the loading results for the representative outfall based on the proportion of the jurisdiction, subwatershed, or land use type drained by that outfall. The benefit of the representative outfall approach to the discharger is that fewer outfalls would need to be monitored for compliance. The benefit to the TMDL and permitting authorities is that more specific WLAs would be simpler to calculate, individual permit limits easier to establish and permit compliance easier to determine.

b. Provide guidance on how to divide up grouped WLAs in permits.

If the TMDL writer cannot disaggregate the WLAs on a geographical or jurisdictional basis, it is still important to explain how the grouped WLA was calculated so it can be implemented accurately in individual permits. If the WLAs are expressed on a concentration basis, it may be more straightforward to implement them as numeric limits for each permitted discharger covered by the WLA. If WLAs are expressed on a mass basis, the TMDL document should explain how the available loading capacity in the WLA was calculated and divided among different discharging point sources. For example, the Ballona Creek, CA metals TMDL and San Gabriel River, CA metals and selenium TMDLs divided available mass loading capacity in a grouped urban point source WLA based on the percentage of land area in the TMDL watershed occupied by the permitted facility. In this scenario, for example, if an industrial facility regulated under an industrial general permit occupied 1% of the TMDL watershed area, it could be assigned 1% of the available load in the grouped WLA.

c. Provide for future point source discharges.

As it can be difficult to revise an approved TMDL, it may be advisable to include a growth WLA when developing the TMDL to account for new and/or expanded permitted source(s) that arise after the TMDL is approved. Fully allocating all available loading capacity to existing sources leaves little or no flexibility to address new sources. Some existing TMDLs have taken the following approaches to providing for future sources. First, a part of the available loading capacity is reserved for potential future allocation to point or nonpoint sources as an explicit margin of safety, assumption, or unspecified allocation. Second, some TMDLs establish a "future growth WLA" to be used specifically for future permitted sources. In either case, the TMDL writers should explain how this reserved loading capacity can be used when permitting point sources that do not have WLAs under the TMDL For example, Minnesota requires a "reserve capacity" to be included in its TMDLs and describes that all or a portion of the reserve capacity would be available through future permitting action(s).

Example TMDLs:

- Chest Creek (PA) Sediments (2011) – Pennsylvania DEP identified one point source discharge within Chest Creek watershed; however, an additional allocation of 1% of the TMDL was incorporated into the WLA as a bulk reserve to take into account future permit activity. It also contained a 10% explicit MOS and the remaining load was allocated to nonpoint sources.

- Cedar River (IA) *E. coli* (2010) – EPA Region 7 wrote this TMDL and at the request of Iowa DNR, they included "reserve WLA" for unsewered communities that were likely to become connected to sewage treatment plants in the future. This was applied to certain river segments and not to others.

- Lake Houston (TX) Watershed Bacteria (2009) – TMDL developers analyzed Houston-Galveston area future population projections to adjust TMDL allocations based on projected loads per day. Growth estimates per sub-watershed ranged from 79% to 182% between 2008 and 2035. WLAs for wastewater treatment plants were adjusted based on these projections.

- Rio Hondo (NM) Nutrients (2005) – This TMDL set aside 2% of the loading capacity for a growth allocation to account for unknown or future nitrogen discharge sources, whether point source or non-point source.

- Long Island Sound (CT) Nitrogen (2001) – Connecticut established a total watershed POTW WLA based on current plant flows and a 5.5 mg/L TN numeric target. The TMDL establishes a a trading program that facilitates trading of loads among sources. Future increases in TN discharges require improved treatment or purchase of excess nitrogen credits from POTWs to ensure overall attainment of the TMDL WLA by the POTWs as a group.

d. Allocations for discharges that may become subject to NPDES.

In situations where it is unclear whether a discharge source is subject to NPDES permitting requirements, the TMDL can include language indicating that if a source receiving a load allocation is later found to be subject to NPDES requirements, its LA is to be interpreted as a WLA for NPDES permitting purposes. For example, areas near urban centers may not now be subject to requirements of municipal stormwater permits but may later become subject to NPDES requirements following adjustment of urban boundaries resulting from population census updates. A TMDL could include language such as: "If any sources currently assigned load allocations are later determined to be point sources requiring NPDES permits, the portion of the load allocations applied to those sources are to be treated as wasteload allocations for purposes of determining appropriate water quality based effluent limitations pursuant to 40 CFR 122.44(d)(1)." This interpretation of a TMDL would not require resubmittal or reapproval of a TMDL.

3. Clarify Where WLAs Apply

TMDLs are sometimes unclear about where WLAs apply because they only identify a facility but do not clearly specify the actual point of compliance where a WQBEL based on the WLA should apply. While the location for the point of compliance is obvious for some traditional facility discharges, in more complex discharge situations, such as large industrial facilities with multiple stormwater outfalls, it can be difficult to define the correct point of compliance when applying a WLA during permit preparation. WLAs associated with MS4 stormwater permits may be notably challenging since there can be many outfalls and often several jurisdictions whose discharges are authorized by the same MS4 permit that are assigned one numeric WLA value.

a. Specify Location where the WLA was calculated.

In most situations, the WQBEL based on a WLA would be applied as an end of pipe limit. If the WLA is mass-based and there are multiple outfalls, the available load may need to be divided among several facility outfalls. Concentration-based WLAs generally can be applied to all outfalls. Some states (e.g., California) sometimes express WLAs for receiving water locations that are downstream of outfalls to receiving waters. There are advantages and disadvantages to this approach. Expressing WLAs and associated WQBELs as receiving water limitations is usually simpler to calculate as it would not require outfall-by-outfall analysis, may result in reduced monitoring costs, and can be protective of water quality if it is reasonably clear how responsibility for any violations will be apportioned among discharges upstream of the receving water point of compliance. On the other hand, it may be difficult to apportion responsibility for violations if multiple entities discharge upstream from a receiving water point of compliance, and similarly, it may be difficult to detect and address causes of violations. In situations where receiving water-based WLAs are developed, the TMDL should include a specific monitoring plan and specify a mix of receiving water and outfall monitoring that is sufficient to: (i) detect violations under different flow and discharge scenarios and (ii) support apportionment of responsibility among different responsible permittees discharging upstream from the receiving water point of compliance. It may be necessary to incorporate a monitoring design that requires monitoring at points upstream and downstream

of changes in jurisdictional status to help distinguish pollutant loads associated with different jurisdictions. In watersheds for which individual WLAs cover multiple dischargers or jurisdictions, the monitoring plan should be designed to to accurately "bracket" different responsible entities and facilitate distinguishing among the multiple contributions to overall pollutant loadings.

b. Special considerations for MS4 stormwater WLAs.

In the case of MS4 permits, which may regulate dozens or even hundreds of separate outfall locations, specifying the point of attainment requires more detailed analysis and description. The TMDL should account for multiple outfall situations by describing how WLAs apply. It may be possible to develop WLAs in multiple outfall situations that focus on a subset of representative outfalls. We have seen three approaches to stratifying WLAs within larger stormwater-permitted settings. For example, the land uses within a stormwater-permitted jurisdiction can be stratified consistent with the TMDL's pollutant source analysis to distinguish among the different levels of pollutant discharges associated with different land uses. Subwatersheds within a larger stormwater-permitted jurisdiction can also be similarly stratified to distinguish how WLAs apply to smaller management areas. Finally, for MS4 permit situations in which there are multiple co-permittees, it is possible to stratify the WLAs based on jurisdiction to distinguish the requirements applicable to individual co-permittees. If WLA stratification methods are used, it may be appropriate to identify representative outfalls for each stratified land use, sub-watershed, or sub-jurisdiction to ensure requirements are clear and monitoring can be conducted that will enable detection and clear apportionment of responsibility for violations of WQBELs based on WLAs.

Example TMDLs:

- Rock Creek (MD) Sediment TMDL (2011) – The TMDL included a five-page technical memorandum regarding WLAs for regulated point sources within the watershed. Permits were placed into two groups: process water and stormwater. Maryland Dept. of Planning applied land use classifications to establish individual and aggregate WLAs for two Phase I MS4 permits, one Phase II MS4 permit and "other NPDES regulated stormwater" sources. While the TMDL technical memo does not supply an exhaustive analysis, it describes information the TMDL writer relied on and cites existing land use methodology previously described in another MDP document to develop WLAs.

- Lower St. Johns River (FL) Nutrients TMDL (2009) – Some facilities addressed by this TMDL asked the State to combine their WLAs into an aggregate WLA to allow flexibility so that reductions from one facility can be shifted to another as long as the net reduction reaches the aggregate WLA. For these aggregate allocations, Florida DEP plans to issue watershed permits that will require compliance with the aggregate WLA.

- Black River (MI) Bacteria TMDL (2010) – Both individual and aggregate WLAs were provided. The Port Huron WWTP received an individual WLA that applies only during high flow conditions and equals 0.1% of the total annual discharge from this facility. Twelve municipal and industrial stormwater permits were given an aggregate WLA with acknowledgement that land use analysis could be used to break this down further at the time of permit development.

c. Account for Mixing Zones.

If the existing permit allows for a mixing zone, it is helpful for the TMDL WLA to specify whether the WLA is to be met before discharge to the mixing zone or at the edge of the mixing zone. Both the TMDL and permit writer should carefully consider whether a past mixing zone analysis remains appropriate in

light of the pollutant loading and effects analysis, and associated allocation decisions, contained in a TMDL

Example TMDL:

> Indian River (DE) Temperature (2004) –The WLA is established as an end-of-pipe allocation and must be met at the Indian River Generating Station (IRGS) discharge outfall on Island Creek. While applicable water quality standards allow for a mixing zone for thermal pollutants, Island Creek does not provide enough dilution capacity to allow a mixing zone.

4. Clarify When WLAs Apply

Many, but not all water quality standards clearly specify the duration and frequency of allowable excursions in addition to their magnitude. The duration and frequency elements vary substantially depending upon the kinds of effects different pollutants have on human or ecological health. TMDL writers should be particularly attentive to the duration and frequency elements of standards when calculating TMDLs, WLAs, and LAs.

Generaly, EPA permit guidance provides that permit writers should calculate limits differently when implementing standards that focus on aquatic life protection as opposed to those that focus on human health protection. Where both acute and chronic aquatic life criteria are applicable, permit writers generally calculate both short term (i.e., daily or weekly) and long term (i.e., monthly limits). Where criteria focusing on human health protection are used, permit writers generally calculate a long term (i.e., annual or monthly) limit from which a short term (i.e., daily) limit is derived. In cases where pollutants are of concern both for aquatic and human health protection, the most limiting set of effluent limitations are generally included as WQBELs in the permit. TMDL allocations delineated in daily time increments can pose problems for permit writers when translating a daily load into the weekly, monthly or annual limits under the permit regulations. Although TMDLs generally establish allowable daily loads, this does not preclude establishing supplemental wasteload information expressed on different timescales (e.g., weekly, monthly, seasonal, or annual loads or concentrations, or daily allocations that vary by month or season). These additional calculations should be linked back to the daily loads to assist the permit writer in demonstrating the additional loads are consistent with the WLAs. The challenge for TMDL writers is to be keenly aware of these required permit elements and ensure the TMDL document and specific WLAs provide sufficient guidance on how they should be implemented in permits consistent with these long-standing permitting principles. In these cases, close and frequent coordination and communication between the permit writer and TMDL writer is critical.

a. Set clear averaging periods for WLAs.

For point sources, NPDES permitting regulations at 40 CFR 122.45(d) specify how WQBELs should address the duration component of standards that apply for different types of discharges. EPA's Technical Support Document (TSD) guidance provides detailed procedures for appropriately converting standards and associated WLAs into appropriate permit WQBELs. Since TMDL writers generally express allocations as daily loads and it is often useful to include WLAs expressed in other timeframes and durations, we encourage early discussions between TMDL and permit writers to determine the most appropriate time components that are consistent with the applicable regulations and guidance. Here are some general approaches used in developing permit limits that should be considered in developing WLAs for consideration:

- For continuous POTW discharges, EPA regulations provide that effluent limitations should be set as average weekly and average monthly WQBELs unless it is impracticable to do so. In our experience, we have observed that permitting authorities have generally found that weekly average limits for POTWs are impracticable and that daily maximum limits and average montly limits are appropriate discharges.
- For continuous non-POTW discharges, WQBELs are generally set as maximum daily and average monthly WQBELs; and
- For non-continuous discharges, WQBELs should specify the frequency of allowed discharges, total mass allowed per discharge event, maximum discharge rate, and concentration-based or other limitations as necessary.

TMDL writers should incorporate such duration components in WLAs, or at least indicate the underlying temporal assumptions of each WLA. TMDLs should distinguish the time periods of concern for different pollutants addressed by TMDLs. For pollutants that are of principal concern due to short term environmental exposures (e.g., ammonia and other acute toxicants), WLAs should include a shorter averaging period based on the applicable standard. For pollutants associated principally with longer term environmental effects (e.g., bioaccumulative toxins), WLAs should include a longer averaging period based on the applicable standard. As many pollutants are problematic both in the short term and the longer term (e.g., nutrients that can cause short term dissolved oxygen swings and long term problems with nuisance plant growth), it may be appropriate to set both short and longer term averaging periods and associated WLAs to be consistent with how the underlying water quality standard is expressed. The TMDL document should specify the translation methods to convert the WLAs into WQBELs. The TSD provides sound methods for translating WLAs based on acute standards into short term permit limits, such as daily or weekly, and WLAs based on chronic standards into long term limits, usually monthly.

Some TMDLs focus on the effect period of principal concern (e.g., chronic effects are of much greater concern than acute effects) and do not include information on averaging periods based on other effect periods. We often see mass-based TMDLs with very long averaging periods (monthly, seasonal, or annual). We also see concentration-based TMDLs that are intended principally to address short term loading and effect periods. WLAs set for these TMDLs should clearly discuss: (i) how permit limits should be expressed both for the effect period of principal concern and (ii) whether and how to set permit limits for other effect periods of lesser concern. If TMDLs do not explicitly describe the most sensitive exposure durations, then permit writers are likely to be uncertain how to make valid assumptions that are consistent with the applicable WLAs.

b. Document any allowable exceedance assumptions in WLAs.

Some water quality standards are written such that they can be exceeded a certain amount of the time without resulting in adverse impacts to protected designated uses (e.g. 1 excursion in 3 years for many toxic pollutant criteria). Where possible, WLAs should reflect allowable exceedence frequencies, consistent with applicable water quality standards, in order to avoid a need to set permit limits more stringent than necessary. Following either State NPDES implementation procedures or EPA's TSD and WLA guidance methods, it is possible to calculate WLAs that account for magnitude, duration, and frequency components of water quality standards and can be converted easily into permit WQBELs. Load duration curves can also be very useful in identifying the allocable pollutant load that corresponds to an allowable exceedance frequency within the water quality standard.

For some pollutants, applicable water quality standards are set as values never to be exceeded (e.g. some bacterial indicator standards). In these cases, associated TMDLs and permit WQBELs generally do not incorporate an allowable exceedance frequency. We have found that WLAs and associated permit WQBELs expressed as values never to be exceeded can be difficult to implement and meet in practice,

especially for sporadic, flashy discharges such as many stormwater discharges. Note that EPA's current criteria for bacterial indicators are no longer expressed as values never to be exceeded.

In cases where the applicable standards do not specify an allowable exceedance frequency, it may be possible to evaluate the standard during the TMDL process to identify an appropriate exceedance frequency. For example, TMDLs for bacterial indicators for discharges to Southern California coastal waters often are based on the concept of exceedance-days. Relatively undeveloped reference watersheds were monitored to determine the frequency with which applicable standards are exceeded (i.e., average days per year). In this case the State formally revised the water quality standard to authorize this approach concurrent with the TMDL adoption action. The WLAs were then expressed as the number of days per year or percent of days that exceed the frequency of exceedance in the reference watersheds. WLAs developed through this type of approach may be more realistic to attain than WLAs expressed as never to exceed values. Again it is critical that the allowable exceedances within WLA be consistent with those explicitly expressed in the applicable water quality standard.

c. Describe the seasonal, flow-based, or wet weather/dry weather WLAs.

As pollutant loadings and effects often vary substantially based on seasonal or rainfall-runoff conditions, TMDLs often set different requirements for different seasons, flow conditions, or rainfall-runoff conditions. The TMDL writer usually has the best knowledge with regards to the applicable flow and duration components in the TMDL. It is therefore very helpful for permit writers when TMDLs can feasibly discuss how permits should be written to reflect seasonal, flow-based, or wet weather/dry weather-based WLAs.

d. Clarify how WLAs work under critical and non-critical conditions.

As many TMDLs are set based on critical low flow or high flow conditions, it is vital for the TMDL document to explain in detail the types of permit limits that need to be developed in order to address the full range of flow and discharge conditions likely to occur. Traditionally, low-flow based TMDLs are designed to support a single set of acute and chronic NPDES permit WQBELs that, if implemented, ensure the water quality standard will be met throughout the year. If a TMDL is designed to authorize different loading levels at different flow levels (to account for different levels of available dilution), it is helpful for the TMDL document to clearly explain how the WQBEL should be expressed to account for different allowable discharges at different flow levels. A single set of concentration-based WQBELs could work across a wide range of flows; however, different mass-based WQBELs may be required for different flow tiers to ensure the WLA is met under all loading scenarios.

Example TMDLs:

- Ventura River (CA) Algae/Nutrients TMDL (2012) – The TMDL contains seasonal allocations to address the critical condition of low stream flows that are dominated by Ojai WWTP effluent discharges during the dry season (331 days). The TMDL includes dry season WLAs for total phosphorus and identifies two separate total nitrogen WLAs for summer and winter dry weather conditions, recognizing that the WWTP has lower nitrogen removal efficiency during cooler winter months. Other point sources, such as Caltrans and the local MS4, get similar dry and wet season mass-based nutrient WLAs to make it easier to evaluate attainment. Wet weather WLAs are concentration-based; the Ojai WWTP received a WLA based on "current performance" interpreted as the 90th percentile of the 12 year data record.

- Potlatch River (ID) *E. coli* TMDL (2009) – The TMDL contains clear language on translating WLAs into WQBEL: "Wasteload allocations apply as instantaneous maximum WQBEL and to any 30-day/calendar month period when effluent discharge occurs."

- Santa Monica Bay (CA) Beaches Wet Weather Bacteria TMDLs (2003) – The TMDL uniquely addressed pathogen impaired beaches by applying the reference beach approach to define WLAs as "allowable exceedance days" relative to reference/natural beach conditions. Corresponding permits have successfully implemented this approach within receiving waters.

5. Clarify Whether Mass and/or Concentration-based WQBELs Will Be Needed

To set mass or concentration-based WLAs and WQBELs for multiple water quality standards and pollutant types, the appropriate way to measure the pollutant of concern is not obvious. While we recognize that TMDLs are generally set in terms of mass per unit of time, it makes sense in many situations to complement mass-based TMDLs with other indicators and measures that are more sensitive to the type of discharges and receiving water situations and implementation activities addressed in the TMDL. For example, a mass-based TMDL can be set for official purposes but accompanied by a concentration-based expression that can be incorporated more effectively into an associated NPDES permit WQBEL or other implementation actions. Similarly, NPDES regulations and associated guidance generally specify that mass-based WQBELs are required in permits except when the standard is expressed in a way that does not make sense to take a mass-based approach or it is otherwise infeasible to do so. See 40 CFR 122.45(f). Permitting regulations provide that concentration-based WQBELs can be used to supplement mass-based WQBELs where appropriate, which may frequently arise when implementing WLAs. If WLAs suggest use of a particular approach or mix of approaches for expressing permit WQBELs that are different from conventional permitting practices, it is much easier for the permit writer to justify use of such non-traditional approaches in the affected permits.

a. Focus on timeframes in which pollutants do their work.

For pollutants for which short term concentration is the key concern in water quality protection, it makes most sense to base the TMDL and associated WLAs on concentration-based indicators (e.g. ammonia, bacteria or currently used pesticide levels downstream from stormwater outfalls). Where longer term mass loading is the key problem to be solved, the WLA and WQBELs should focus on mass (e.g. long term nutrient load to a lake or bioaccumlative pollutants like PCBs or mercury).

b. Consider discharge characteristics and ease of monitoring.

Facility discharge characteristics and ability to monitor efficiently are equally important to consider. It is easier to measure and regulate irregular, non-continuous discharges such as industrial discharges on a concentration basis, particularly in settings in which facility and receiving water flows are not routinely monitored. For example, a concentration-based allocation may make the most sense for a pollutant source such as dry weather "illicit discharges" of non-stormwater from MS4s', which are typified by non-continuous discharges that are difficult to predict or measure. It is generally easier to monitor on aconcentration basis. On the other hand, it is easier to evaluate mass-based loads from continuous discharges such as municipal wastewater treatment plants with predictable flow rates and volumes.

c. Clarify whether mass load WLAs are based on actual facility flows or design flows.

Permit regulations for WWTPs require that WQBELs be calculated based on facility design flows, which are normally greater than actual flows. See 40 CFR 122.45(b). Some TMDL WLAs may have been

calculated based on actual flows or loads, which presents a challenge for permit writers (particularly when discharge flows are substantially lower than permit design flows). The TMDL document should clarify whether flow-based WLAs should be translated directly into WQBELs or whether a correction factor needs to be included to accurately convey the intent of the WLAs in WQBELs based on WWTP design flows. If this is not done, WWTP permit limits could be set at higher loads than intended in the WLA.

d. Be careful with percent reduction allocations.

Many TMDLs express allocations as needed percentage reductions in pollutant loading. Unless these WLAs are accompanied by a very clear baseline load from which the percent reduction is to be calculated, it is very difficult to set enforceable permit limits for this type of WLA. WLAs expressed as mass-based and/or concentration-based WLAs are preferred, in addition to expressing them as percent reduction from a given baseline, since this makes it easier to evaluate whether associated permit limits are achieved. It is also important to be clear about the averaging periods in which the required percentage reductions are supposed to be accomplished.

e. Consider concentration-based WLAs for individual facilities permitted under general NPDES permits.

Many point sources regulated under general permits (e.g., industrial and construction stormwater sources) may discharge pollutants addressed by TMDLs but do not have available discharge flow measurements. In many cases, these discharges are relatively short term in duration, are sporadic, or vary substantially during storm periods. For these reasons, it may be difficult to set mass-based WLAs. In cases where flow rates are not available, it may be best to assign concentration-based WLAs for facilities authorized by a general NPDES permit to assist in setting practicable WQBELs for these discharges. This may also be the case for bioaccumulative pollutants. The lack of flow discharge rates makes it difficult to calculate mass-based WLAs, and concentration-based WLAs applicable at end-of-pipe may be more practical.

Example TMDLs:

- Los Angeles-Long Beach Harbor (CA) Toxics TMDLs (2012) – Concentration-based WLAs were identified for such sources as: construction stormwater, industrial stormwater, power generating stations, minor permits and irregular dischargers. Any future minor NPDES permits or enrollees under a general non-stormwater NPDES permit will also be subject to the concentration-based wasteload allocations. The TMDL provides both acute and chronic WLAs which are then translated to WQBELs by following state implementation procedures (similar to EPA TSD guidelines).

- Charles River (MA) Bacteria TMDL (2007) – The TMDL expresses the wasteload allocations in three ways: (i) concentration-based (equal to WQS); (ii) percent reduction for implementation purposes and (iii) mass-based in colonies/day. This multi-allocation approach was also useful for defining pollution prevention TMDLs for waters that were not currently impaired but might be found to be impaired in future assessments within the watershed.

- Bayou Cocodrie (LA) Copper TMDL (2007) – The TMDL for dissolved copper in Bayou Cocodrie has been set to the water quality criteria concentration for all sources at all effluent flow rates (criteria is met at the end-of-pipe). In the event future revisions to the criteria are made, the TMDL would remain set at the criteria at the end-of-pipe. The TMDL is set as a daily maximum value.

f. Mass-based WLAs may be more appropriate for pollutants that accumulate over time in water, sediment, or tissue.

Adverse effects of certain pollutants may occur over longer exposure timeframes or appear over longer residency within the waterbody. For example, aquatic or wildlife or humans may bioaccumulate and/or biomagnify pollutants via food web dynamics and are at risk due to long-term exposures. As nutrients can also build up in sediments and later re-solubilize to create eutrophic conditions, their long-term loading is important and mass-based WLAs are more appropriate.

6. Setting WLAs Based on Formulas

For pollutants whose environmental effects vary based on water quality factors like hardness, temperature, and pH, TMDLs and WLAs are sometimes set as a formula that makes the allocation a function of an independent water quality factor. In contrast, it is preferable for NPDES permit limits be set as firm, not floating, values since it can be very difficult to monitor and evaluate compliance with floating limits. If a WLA is set based on such a formula, the TMDL document should provide clear direction about what critical value or values should be assumed when calculating permit limits for that pollutant (e.g., default values for temperature, hardness, or pH). If no single critical value is set, the document should specify that a particular percentile value from the actual receiving water data set must be used to result in a protective permit limit. There are examples where the critical percentile value was set in the 90-99% range. Where appropriate, different critical values can be set to correspond to different seasonal requirements or variability in seasonal conditions as long as the seasonal variations are clearly articulated in the TMDL document and associated permit. Also, it is important to base a formula-based limit on data collected from the receiving water, not on data for the effluent waste stream itself.

Example TMDL:

- LA River (CA) Metals TMDLs (2005) – The TMDL applied median hardness to determine chronic WLA during dry weather conditions and 90th percentile hardness for acute WLA during wet weather.

7. Special Considerations for Stormwater WLAs and WQBELs

In setting WLAs for stormwater sources, it is important to account for the unique characteristics of these discharges and associated impairments.

a. For extremely short-duration stormwater events, acute-based WLAs may be appropriate.

If it can be demonstrated in a particular TMDL analysis that stormwater discharges and associated receiving water flows never or very rarely last for extended periods of time (e.g., 4 days or longer), it may be appropriate to base the TMDL and stormwater WLAs solely on the acute standards for pollutants that have both acute and chronic standards. Conversely, if the observed duration of storm flows exceeds critical short term thresholds (e.g., 4 days) with a frequency greater than the allowable exceedance frequencies embodied in the underlying standards (e.g. once every 3 years on average), the TMDL and WLAs should generally be based on more protective chronic standards. For example, if TMDL analysis of a long term stream flow data set showed that storm-related high flows exceeding 4 days in duration occur only once every 50 years on average in that stream, then one could consider basing the WLAs on the acute standard.

b. Concentration-based TMDLs may be easier to implement for stormwater sources.

As flows and pollutant loads vary substantially during and across different storm events, it can be difficult to develop workable mass-based TMDLs and WLAs to address certain stormwater sources. In conjunction with a load-based WLA, concentration-based WLAs are often easier to implement both from the standpoint of monitoring ease and the ability to associate changes in concentration with BMP effectiveness. It is important to collect samples at different times during the storm period and for a representative range of storms during the year in order to provide the data needed to accurately characterize stormwater-related pollutant loadings. For MS4s, in which sufficient data are collected to yield event-mean loading and concentration data, it may be desirable to express WLAs in terms of event mean concentration values. This approach allows the permit writer to assess and possibly express the WQBEL in terms of event mean concentrations.

8. Implementation Provisions in TMDLs

Actions to address the causes of stubborn water quality impairments may take many years to implement. It is helpful for a TMDL to include implementation recommendations, either within the TMDL itself or in a separate implementation plan, about the appropriate implementation timeframes and mechanisms. This information is extremely useful during the permitting process to help ensure the permit incorporates requirements necessary to implement WLAs.

a. TMDL implementation recommendations can assist in permit development and other control actions.

It is very helpful for TMDL documents to describe the state's expectations for implementation, including actions needed to implement WLAs and LAs, appropriate timeframes for implementation, responsible parties, and necessary monitoring provisions. See Section B.5 below. TMDL and permit writers should keep in mind that including time schedules in a TMDL does *not* mean that individual permit compliance schedules, enforcement orders, or variances have been legally established. TMDLs are not self-implementing; they are implemented only through subsequent permitting or other implementation actions which are then evaluated for compliance. When EPA acts on a TMDL submittal, EPA is only approving the TMDL calculations, not the State's implementation provisions incorporated in the TMDL document or a separate implementation plan. The information provided in the implementation provision portion of a TMDL or accompanying documentation should be clear that its purpose is to assist the permit writer by providing context and further information.

If a state intends to authorize compliance schedules to provide time to implement WLAs, it must establish compliance schedule authorizations pursuant to CWA 303(c) and 40 CFR 122.2 and 122.47. After EPA approves a compliance schedule authorization request that is specifically applicable to the permit(s) addressed by a TMDL, the permitting authority must then incorporate specific compliance schedule provisions within individual permits (or, if applicable, general permits) as part of an NPDES permit proceeding. As part of that proceeding, once the permittee submits a request and supporting documentation for a compliance schedule, the permitting authority would then need to establish that the compliance schedule requirements at 40 CFR 122.2 and 122.47 have been satisfied on a permit-by-permit basis.

B. INTERPRETING EXISTING TMDLs WHEN DEVELOPING NPDES PERMITS

The challenges a permit writer faces in interpreting existing TMDLs are probably as daunting as the challenges a TMDL writer faces in writing a permit-friendly TMDL. Section B offers a few basic considerations that may improve the TMDL to permit translation process. One overarching observation is that permit writers need to thoughtfully evaluate and interpret TMDLs, and avoid simply importing verbatim the WLAs into the permit unless it is clear that the WLA itself is suitable for use as a permit WQBEL. It is critical for permit writers to discuss with TMDL writers point source permitting and control issues that forthcoming TMDLs need to address. Early coordination often helps avoid many of the interpretation problems described throughout this document.

1. Check for and review TMDLs before writing the NPDES permit

In many regulatory agencies, there is no clear mechanism through which permit writers can learn about the existence of TMDLs that affect the permits they are writing. TMDL and permitting authorities should establish regular methods for keeping permit writers up-to-date on new or revised TMDLs, and make TMDL and permit writers responsible for keeping their counterparts up to date on new TMDL actions. Ideally, permit writers should become aware of and assist in the development of the TMDL well before the TMDL is completed to help ensure that the final TMDL includes needed WLAs that are written in a "permit-friendly" way. At times permit writers do not take this simple step, and miss approved, applicable TMDLs and relevant WLAs therein. Permitting specialists should ensure they are fully aware of how TMDLs apply to specific individual and general permit situations. The permit writer should carefully review the TMDL problem statement, targets, source analysis, linkage analysis, margin of safety, and TMDL assumptions in addition to the WLA section in order to develop a clear understanding of how the TMDL relates to a particular NPDES facility. Additionally, the permit writer may gain insight by reviewing the TMDL approval document which may contain language on specific items that EPA did not act on or did not approve, such as interim limits, timeframes, etc.

2. Addressing TMDLs That Do Not Include a WLA For Your NPDES Permit

Many already approved TMDLs do not include a WLA for NPDES permitted facilities or new NPDES point sources discharging to a waterbody addressed by a TMDL. How should these permits be addressed in the permitting process? One (albeit "strict") reading of this situation is that the permit should prohibit any discharge of the pollutant of concern since the TMDL made no provision for such discharges. However, in some circumstances, there may be a reasonable approach to ensure that the permit incorporates WQBELs and provisions as necessary to be consistent with the assumptions of the TMDL and ensure that applicable water quality standards are met. Doing so in the absence of a specific WLA presents some permitting challenges. In an ideal world, the TMDL would be revised to account for the unaccounted NPDES dischargers, but in the absence of that ideal, permit writers should work with TMDL developers to weigh these factors and decide an appropriate, protective course of action. Permit writers should not automatically assume that the absence of a WLA means the TMDL is not relevant for the NPDES permit. If the facility at issue discharges the pollutant of concern to a waterbody for which there is an approved and applicable TMDL for that pollutant, the permit must contain WQBELs consistent with the assumptions of "any available wasteload allocation for the discharge." See 40 CFR 122.44(d)(1)(B)(vii). As discussed in Section A.1., some TMDLs do not include WLAs for point sources discharging to a waterbody for which there is an approved and applicable TMDL. This seems to occur for three main reasons, discussed below.

a. TMDL does not recognize existing NPDES discharges.

Many TMDLs do not address all NPDES discharges that occur in the TMDL watershed. This is particularly common with respect to discharges permitted by general permits, temporary discharges, and intermittent discharges. The Region has also encountered TMDLs that mention a permitted facility within the source analysis and then neglect to assign a WLA for the facility. One interpretation of this situation when writing or renewing a permit is that no assimilative capacity is available and the NPDES discharge without a WLA must be prohibited from discharging the pollutant of concern or receive a zero permit limit. Depending upon the type of TMDL, it may also be appropriate to set a WQBEL equal to water quality criteria end-of-pipe; e.g., if the TMDL applies concentration-based WLAs for all sources, it may be possible to find that WQBEL equal to water quality criteria is consistent with the TMDL's terms and assumptions.

In rare circumstances, if the TMDL includes an unallocated explicit margin of safety created in part to account for unidentified point source discharges, it may be possible to create a numeric WQBEL consistent with the TMDL that, in essence, allocates a portion of the explicit margin of safety. It might be possible to take this approach in settings where the TMDL explicitly discusses this as a valid implementation approach, and the new discharges covered under this approach are not significant dischargers of the pollutant of concern. The preferred approach is for the TMDL to include future growth WLAs or reserve capacity as described in Section A.2.c.

b. TMDL analysis views NPDES discharge as insignificant.

In some cases, the TMDL writer may not set a WLA for an NPDES because the discharge is considered to be very small/insignificant in comparison to other pollutant sources. Presumably, the TMDL contains additional information to make clear that the decision was made on this basis. If the TMDL document does not describe this situation, it is the functional equivalent of ignoring an NPDES facility. See Section B.2.a immediately above. If the TMDL document includes an analysis of why a specific discharge is insignificant and explains how any associated permit may want to address the pollutant identified in the TMDL, the permit writer will likely have the guidance needed to develop appropriate permit provisions consistent with the assumptions in the TMDL.

Such permit provisions may include:

- Limit the discharge to current performance. See Section A.1.e.; or
- Limit the discharge based on a criteria end-of-pipe WQBEL and including monitoring requirements for the pollutant of concern to support review of the facility contribution of a pollutant of concern in the future.

c. TMDL does not account for new NPDES sources.

Some TMDLs do not incorporate WLAs for potential new NPDES sources that do not exist at the time of TMDL development. The approaches discussed in Sections B.2.a and B.2.b above may be appropriate in this situation. Other options to address this situation include:

- Revise the TMDL to include WLAs for new and/or expanded point sources and then incorporate appropriate WQBELs in the permits; or
- Set WQBELs based on criteria end-of-pipe, although this may be problematic for fully allocated mass-based TMDLs.

3. Translating WLAs into Wastewater NPDES Permit WQBELs

In the best case scenario, the TMDL tells the permit writer all he or she needs to know in order to translate the WLA into a permit. If not, the permit writer can consider following the permit derivation methods recommended in the TSD and EPA's NPDES Permit Writers' Manual (2010) to develop appropriate WQBELs based on WLAs. As discussed above in Section A, it may not be appropriate to simply copy the WLA and paste it into the permit as it may be difficult to conduct appropriate compliance monitoring and assess compliance with limits expressed in unconventional ways.

a. A separate reasonable potential analysis is generally unnecessary.

In many circumstances, where TMDL WLAs indicate the need for a corresponding WQBEL, a separate reasonable potential analysis should not be conducted if a WLA exists for a particular facility, absent any new information indicating changes in the facility's discharges of the pollutant of concern.

b. Specify duration and frequency elements consistent with WLA.

As discussed above, it may be necessary to translate WLAs so that they "fit" with how WQBELs are expressed consistent with permitting regulatory requirements. This may require some translation of WLA averaging periods to fit how short and long term permit WQBELs are expressed. The TSD provides useful guidance in how to perform these translations. For WLAs that are set on a seasonal basis or under different flow scenarios, the permit should specify the periods in which different seasonal limits or flow-based limits apply, consistent with the TMDL provisions.

c. Take care in setting mass and concentration-based WQBELs.

As discussed above in section A.5, it may be difficult to translate some WLAs into a form that makes sense for the permit (e.g., the WQBEL is clear, able to be monitored, and susceptible to compliance determination). For example, the use of concentration-based permit WQBELs may be most appropriate to address non-continuous discharges (e.g. stormwater discharges without associated flow measurements) where the underlying water quality standards and associated TMDLs are expressed in terms of concentration. Also, take care in setting mass-based WQBELs based on design flow so that the limit is as stringent as necessary to ensure an associated mass-based WLA will be met. Limits based on design flows could be inconsistent with TMDL WLAs based on actual flow or some flow assumption lower than design flows used in calculating prior permit limits.

d. Clarify where WQBELs apply.

Even in cases where WLAs do not clearly specify where allocations apply (e.g. end of pipe, in receiving water), permit limits should include clear explanation of where they apply and where monitoring data should be collected to evaluate compliance.

e. Remember technology-based limits.

Many facilities require both technology-based and water quality-based effluent limitations. Permit writers should include technology-based effluent limitations in permits if required, even if they do not address water quality impairment that TMDLs and WLAs are designed to address. Where technology-based and water quality-based requirements apply for a particular pollutant, the more stringent of the two approaches should generally form the basis for the effluent limit.

f. Respect TMDL recommendations on permitting approaches.

Particularly in cases in which TMDL developers provide specific guidance or assumptions on how permit limits should be expressed to implement WLAs, permit writers should give that guidance/assumptions careful consideration. This is particularly important in cases where the WLA is expressed in terms of multiple expression indicators (e.g., both mass and concentration-based), unless it can be shown within the NPDES permit factsheet that one permit WQBEL form serves as an effective surrogate for other indicators used in the WLA.

4. Considerations in Permitting Stormwater Sources

As discussed in Sections A.5 and A.7, developing stormwater permit provisions to implement WLAs can be challenging. Prior EPA guidance concerning TMDLs and stormwater permits provides valuable recommendations that permit writers should consider in developing appropriate permit language to address stormwater WLAs. See EPA Office of Water memo (Nov. 26, 2014) "Revisions to the November 22, 2002 Memorandum "Establishing Total Maximum Daily Load (TMDL) Wasteload Allocations (WLAs) for Storm Water Sources and NPDES Permit Requirements Based on Those WLAs", and EPA Office of Water document (June 2014) "Municipal Separate Storm Sewer System Permits- Post Construction Performance Standards and Water Quality Based Requirements- A Compendium of Permitting Practices, EPA 833-R-14-003. The following considerations are particularly relevant in developing stormwater permits.

a. Not all stormwater permits are subject to the same requirements.

In cases where WLAs are established for industrial or construction stormwater discharges, numeric WQBELs are generally practicable and appropriate for such stormwater permits. As discussed in the 2014 memo cited above, numeric limitations are often appropriate for inclusion in municipal stormwater permits as well. We have found that inclusion of numeric effluent limitations in stormwater permits improves their clarity and enforceability.

b. BMP-based approaches to implementing WLAs can work but should be supported by robust analysis.

In permits where BMP-based limitations are used in stormwater permits, it is important for the permitting authority to provide a strong basis to support the assertion that a BMP-based approach to WLA implementation will be sufficient to result in timely attainment of the WLA. This can be done if the permit is supported by detailed identification of the BMP systems to be implemented and a strong analytical documentation showing how implementation of the BMP system will result in attainment of specific WLAs in specific compliance locations. Model-based approaches to demonstrating that a BMP based approach can work may provide the necessary framework for supporting a prospective conclusion that a particular BMP-based implementation plan will provide reasonable assurance that the WLA will be implemented. Other approaches may be appropriate, but we have found it is difficult to document a robust connection between specific implementation practices and associated water quality outcomes based solely on analysis of monitoring data and BMP effectiveness studies.

As modeling and the monitoring tools used to evaluate the effectiveness of stormwater management BMPs improve, it should become increasingly plausible to establish BMP-based approaches to implementing numeric WLAs. For example, when connected with watershed scale water quality models, BMP siting and optimization models that are capable of associating BMPs with the levels of expected water quality improvements at different downstream locations can be used to quantitatively link BMP implementation plans to the attainment of specific WLAs. That is, the various BMPs would serve as

functional equivalents to the interim numeric WQBELs in the permit, provided the permittees include a robust analysis showing the BMPs 'add up' to reducing the pollutant load outlined in the TMDL. To be effective, these models need to be capable of breaking down BMP planning to a fairly small scale and then integrating BMP plans across TMDL jurisdictions to establish reasonable assurance that a selected BMP system will result in WLA attainment in a reasonable period of time. BMP-based approaches based on such model-based planning systems may be advantageous to jurisdictions as they provide a long-term framework for planning investments in stormwater controls and obtaining funding needed for BMP implementation and ongoing maintenance. This framework can also provide a basis for reviewing and, if necessary, revising TMDL implementation timeframes as the modeling results can help indicate how quickly WLA results can be attained through implementation of affordable plans for stormwater control and green infrastructure capital investments.

Example Stormwater Permits:

- Lake Tahoe MS4 Permits (2011), applicable to urban areas of two counties and one city in the Tahoe basin, implements the Lake Tahoe Sediment and Nutrient TMDL. The TMDL incorporates a model-based system for setting sediment reduction requirements and providing pollution control credits for specific BMP implementation actions necessary to meet sediment reduction needs. The credits awarded are determined based on the practices implemented, the locations of the implementation sites in the watershed, and the projected sediment reductions associated with these actions. The permit contains the WLAs based on a crediting system as numeric WQBELs and provides the framework for determining specific implementation requirements and evaluating attainment with WLAs.

- San Diego (CA) Regional MS4 permit (2013) – The MS4 permit incorporates final TMDL WLAs as WQBELs for several pollutants including nutrients, metals, and bacterial indicators, expressed as Receiving Water Limits, and Effluent Limits that are measured at outfalls. It also establishes interim WQBELs, for which compliance may be demonstrated via monitoring of receiving waters, effluent, or by implementation of BMPs in an approved Water Quality Improvement Plan (WQIP) providing reasonable assurance that interim milestones will be achieved consistent with established schedules. WQIPs are prepared for all major watersheds, and include adaptive planning processes that identify and address the highest priority water quality conditions. Permit compliance deadlines extend until 2031 for some TMDLs, and it is recognized that the state permitting authority may revise these schedules based on WQIP implementation, either via the permit's re-opener provisions, or through future renewal of the permit.

- Arlington County (VA) MS4 permit (2013) – This permit implements the Chesapeake Bay TMDL that includes mass-based WLAs for TN, TP and TSS. (The TMDL also provides underlying information about pollutant specific loading rates (lbs./acre) for sub-watersheds within each State.) The MS4 permit includes many different measureable requirements to reduce nutrient and sediment loads, including:
  - Reduce by 5 percent the loadings from existing developed lands within the first permit term. The permit includes tables with required pollutant-specific loading rates to calculate required loading reductions from existing sources.
  - Offset (by 5 percent) increased loads from new construction projects disturbing one or more acres and from grandfathered projects.
  - Identify and submit to the state at least 7 retrofit projects within its watershed retrofit plans that will be implemented within County rights-of-way or on County property within 60 months of permit issuance.
  - Plant a minimum of 2,000 trees on County lands and develop a program to distribute a minimum of 2,000 trees to private property owners.

- Fund a minimum of 200 participants in the StormwaterWise Landscape program which provides cost-sharing and technical assistance for the installation of small scale best management practices to reduce stormwater runoff from private properties.

- Washington State Department of Transportation (WSDOT) MS4 permit (2014) – The permit includes specific mandated action items for WSDOT that are associated with individual TMDLs. For example, to implement the Hangman's Creek Fecal Coliform and TSS/Turbidity TMDLs, WSDOT is required to reduce fecal coliform loads by 72% and sediment loads by 80%. Here are a few action items outlined in Appendix 3, Applicable TMDL Requirements:

  - Within 6 months of permit issuance, WSDOT will update the initial inventory findings report with updates on potential TMDL concerns, and follow-up actions taken and/or notification to others where a concerns has been identified but occurred outside WSDOT's right-of-way and control.
  - If stormwater discharges that transport bacteria over natural background levels to listed receiving waters are found from sources within WDOT's right-of-way and control, WSDOT will apply BMPs from their SWMPP or perform remediation to correct bacteria discharges. To address TSS/turbidity, WDOT will work to prevent sediment from entering area waterways along SR 27 (in upper watershed) and SR 195 right-of-ways. WSDOT will prioritize problem areas and work with individual property owners to prevent sediment from entering area waterways via WSDOT's MS4.

- Western Washington Phase II Municipal Stormwater Permit (2012) includes tables that establish additional watershed-specific actions that are required of each named MS4. The actions are differentiated based on the applicable TMDL and pollutant of concern. See Appendix 2.

(Note: More stormwater permits are presented in EPA Office of Water document (June 2014) titled, Post-Construction Performance Standards and Water Quality-Based Requirements in Municipal Separate Storm Sewer Systems Permits: A Compendium of Permitting Approaches.)

c. Numeric WQBELs can be easier to measure and enforce.

The Region has conducted several dozen stormwater permit audits over the past 5 years, and we have concluded that inclusion of non-numeric permit requirements, especially those based on relatively vague iterative BMP requirements, are difficult to evaluate for enforcement purposes and have not demonstrated significant water quality improvements. This is particularly true in cases where insufficient monitoring was required in permits to support evaluation for compliance with permit requirements. Numeric WQBELs can be much more straightforward to implement and monitor as long as permit writers are sensitive to the variability of stormwater discharges in determining how they should be expressed and monitored. As discussed above in Section A.7.b, concentration-based WQBELs are often easier and more sensible to implement than mass-based WQBELs for stormwater discharges. If, however discharge flow measurements exist and long-term pollutant exposure is of concern, then mass-based WQBELs can be applied if sufficient data are available to evaluate long term loads.

5. Providing Time to Implement Needed Controls

Some TMDLs can take decades to implement as many sources will need to implement changes in facility operations and treatment technology. There are several mechanisms for providing time needed to implement new permit requirements based on TMDLs. The key point to remember is that permit writers must follow the permit regulations. While implementation timeframes can be identified during the

TMDL process, compliance schedules can be established only through permit proceedings following establishment of compliance schedule authorization in applicable water quality standards. Similarly, variances can be established only through water quality standards proceedings. The TMDL document can provide much of the needed information on why extra time may be needed and the interim steps that should be taken. The three main mechanisms used to provide extra time for TMDL implementation in the NPDES permit are variances, compliance schedules, and enforcement orders.

a. Variances provide a time-limited designated use and criterion for the purpose of NPDES permitting that are less stringent than the underlying water quality standards.

Variances can be used to delay the effective date of water quality standards applicable to some NPDES discharges. Variances are intended to provide short term relief from immediate responsibility to comply with a water quality standard that may be unattainable in the short term, but are not intended to be permanent. Variances must be authorized under applicable water quality standards and approved by EPA. Although specific variance procedures vary and are specified in applicable water quality standards and associated implementation procedures, and their supporting documentation, variances generally:

- Are pollutant-specific;
- Specify the highest level of water quality that can be attained during the time period the variance is in effect;
- Explain why the variance is needed;
- Provide appropriate justification for temporary use attainability changes per WQS regulations; and
- Specify interim controls and studies that will be conducted to improve water quality and/or determine whether underlying designated beneficial uses are attainable.

b. Compliance schedules provide extra time but must contain interim steps to attain limits ASAP

While NPDES permits can contain compliance schedules if additional time is needed to implement new water quality based control requirements, NPDES regulations require that findings must be made at the time of permit issuance that demonstrate that the specific compliance schedule requirements are met for the permitted facility. See 40 CFR 122.47. The permit record must demonstrate that: (i) the compliance schedule is needed, (ii) the compliance schedule will result in compliance with the limit as soon as possible, and (iii) interim compliance schedules and action-based milestones are incorporated to ensure reasonable progress is being made toward completing control actions needed to meet the new WQBEL. The permit included the final WQBEL and its deadline which may be beyond the five year permit cycle. See EPA Office of Water memo (May 10, 2007) from Jim Hanlon, EPA OWM to Region 9, titled "*Compliance Schedules for Water Quality Based Effluent Limits in NPDES permits.*" Compliance schedules cannot be used to provide time to develop or revise a TMDL, a water quality standard, or a mixing zone analysis.

c. Enforcement Orders provide greater flexibility but must be outside permit.

In some cases where variances and/or compliance schedules are not authorized under state water quality standards or are inappropriate for a specific case where a new WLA is considered when a permit is issued or re-issued, it is often possible to issue a companion enforcement order with an NPDES permit. These orders can provide additional time for a facility to come into compliance with a permit requirement and specify interim actions that are needed to result in eventual compliance.

C. CONCLUSION

EPA and state TMDL and NPDES permitting authorities are still learning how to best link the TMDL to NPDES permits process to effectively guide implementation of controls on point source discharges as needed to help restore impaired waters. The Region is seeing significant improvements in comparison to the "early days" of TMDL development and implementation. This document provides some initial lessons learned and examples to illustrate these concepts. The Region anticipates that as NPDES permits are reissued in the coming years, the information in this document will need to be updated to include new examples or modified information. The Region has an interest in improving and ensuring the accuracy of the information contained in this document and therefore welcomes any comments, corrections and examples on any aspect of this document at any time. The Region may update this document as needed based on the comments received as well as any TMDL or permit examples to strengthen it. Please provide any feedback to David Smith (smith.davidw@epa.gov) and Peter Kozelka (kozelka.peter@epa.gov).