United States
Environmental Protection
Agency

Office of Water
4304T

EPA-820-R-23-002
February 2023

PROCEEDINGS FROM THE
EPA FREQUENCY AND DURATION EXPERTS WORKSHOP
SEPTEMBER 11–12, 2019

February 2023

U.S. Environmental Protection Agency Office of Water

Office of Science and Technology

Health and Ecological Criteria Division

Washington, D.C.

Disclaimer

The information in this document was funded by the U.S. Environmental Protection Agency (EPA) and was subjected to agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Furthermore, this document is a summary of the views of the individual workshop participants; approval for publication does not signify that the contents reflect the views of the agency, and no official endorsement should be inferred.

Acknowledgments

EPA would like to thank the speakers and others who participated in the 2019 *Frequency and Duration Experts Workshop*. Their contributions to the workshop and dedication to produce these proceedings are greatly appreciated.

Workshop Participants: Kevin Brix (EcoTox, LLC), Molly Colvin (Naval Information Warfare Center), Russ Erickson (EPA, Office of Research and Development [ORD], Duluth Laboratory), Bryson Finch (Washington Department of Ecology), Andre Gergs (Bayer AG, Environmental Sector Department), Sarah Kadlec (EPA, ORD, Duluth Laboratory), Jeff Manning (North Carolina Division of Water Resources, Water Quality Standards [WQS] Program), Chris Mebane (U.S. Geological Survey [USGS]), Dave Mount (EPA, ORD, Duluth Laboratory), Adam Ryan (International Zinc Association), Travis Schmidt (USGS, National Ambient Water Quality Assessment project), Bill Stubblefield (Oregon State University), Chris Stansky (Wood Environment and Infrastructure Solutions, Inc.), and Ning Wang (USGS, Columbia Environmental Research Center).

Mary Reiley, Michael Elias, and Jim Justice (EPA, Office of Water) organized the workshop. Danielle Stephan and Eric Monschein, both with EPA, were invited to give presentations on the use and implications of the duration and frequency provisions of criteria in permitting and impaired waters assessments, respectively. Observers in the room included representatives from the EPA Office of Water (Kathryn Gallagher, Lareina Gunzel, John Healey, Jacques Oliver, and Laura Phillips) and the Association of Clean Water Administrators (ACWA; Mark Patrick McGuire). Observers also included EPA contractors who served the role of note taking.

Contractor Support: Susan Bowman and Alex Taylor (Cadmus Group), Kaedra Jones (ICF)

Contents

Acronyms v

Foreword vi

Executive Summary vii

Introduction 1

Day 1 – Duration 2

Presentation I: Duration History and Application (Discussion Leader: Chris Mebane, USGS) 2

Expert Group Discussion of Presentation I on Duration: History and Application 5

Presentation II: History of Criteria Duration Derivation and Underlying Assumptions (Discussion Leader: Russell Erickson, EPA/ORD) 6

Expert Group Discussion of Presentation II: History of Criteria Duration Derivation and Underlying Assumptions 7

Presentation III: Duration Application in NPDES Permits (Discussion Leader: Danielle Stephan, EPA/OW/Office of Wastewater Management [OWM]) 8

Expert Discussion of Presentation III: Duration Application in NPDES Permits 9

Presentation IV: Duration Application in Assessment, Listing, and TMDLs (Discussion Leader: Eric Monschein, EPA/OW/Office of Wetlands, Oceans and Watersheds [OWOW]) 10

Expert Discussion of Presentation IV: Duration Application in Assessment, Listing, and TMDLs 11

Presentation V: Exposure Duration and Effects Group: Part I (Discussion Leader: Adam Ryan, International Zinc Association) 13

Expert Discussion of Presentation V: Exposure Duration and Effects Group: Part I 14

Presentation VI and Expert Discussion: Exposure Duration and Effects Group: Part II (Discussion Leaders: Adam Ryan, International Zinc Association [first presenter] and Kevin Brix, ECOTOX [second presenter]) 18

Presentation VII: Exposure Duration and Effects Group: Part III (Discussion Leader: Russell Erickson, EPA/ORD) 22

Expert Discussion on Presentation VII: Exposure Duration and Effects Group: Part III 23

Day 2 – Frequency 28

Presentation VIII: Office Director Remarks (Deborah Nagle, EPA/OW) 28

Presentation IX: Frequency: History and Application (Discussion Leader: Chris Mebane, USGS) 28

Expert Discussion of Presentation IX: Frequency: History and Application 30

Presentation X: Frequency Application in NPDES Permits (Discussion Leader: Danielle Stephan, EOA/OW/OWM) 31

Expert Discussion of Presentation X: Frequency Application in NPDES Permits 31

Presentation XI: Frequency Application in Assessment, Listing, and TMDLs (Discussion Leader: E. Monschein, EPA/OW/OWOW) 32

Presentation XII: Disturbance and Recovery Group: Part I (Discussion Leaders: Dave Mount, EPA/ORD [session presenter 1], and Andre Gergs, Bayer AG, Environmental Sector Department [session presenter 2]) 33

Comments from Observers – Disturbance and Recovery Group: Part I 37
Presentation XIII: Disturbance and Recovery Group: Part II (Discussion Leader: Chris Mebane, USGS) 37
Expert Discussion on Presentation XIII: Disturbance and Recovery Group: Part II........ 38
Comments from Observers – Disturbance and Recovery Group: Part II 40
Presentation XIV: Frequency: Approaches Group (Discussion Leader: Chris Mebane, USGS) 41
Expert Discussion of Presentation XIV: Frequency: Approaches Group 42
Comments from Observers – Frequency: Approaches Group 44
Closing Remarks (Mary Reiley, EPA/OW) 44
References 45

Appendix A. Workshop Participant List and Agenda
Participants List
Agenda
Appendix B. A Review of the Nature and Effects of Episodic Water Pollution and Implications for Aquatic Life Criteria Averaging Periods
Appendix C. The Capacity of Aquatic Ecosystems to Recover from Exceedances of Aquatic Life Criteria

Acronyms

1Q10	lowest 1-day average flow that occurs (on average) once every 10 years
7Q10	lowest 7-day average flow that occurs (on average) once every 10 years
ACR	acute to chronic ratio
ACWA	Association of Clean Water Administrators
AWQC	ambient water quality criteria
CCC	criterion continuous concentration
CMC	criterion maximum concentration
CWA	Clean Water Act
ELS	early life stage
EPA	U.S. Environmental Protection Agency
FACA	Federal Advisory Committee Act
GUTS	General Unified Threshold model of Survival
HECD	Health and Ecological Criteria Division
L	liter
LC_{40}	lethal concentration required to kill 40 percent of the test population
LC_{50}	lethal concentration required to kill 50 percent of the test population
LD_{10}	lethal dose of an ingested substance that kills 10 percent of a test population
LD_{50}	lethal dose of an ingested substance that kills 50 percent of a test population
NPDES	National Pollutant Discharge Elimination System
ORD	Office of Research and Development
OST	Office of Science and Technology
OW	Office of Water
OWM	Office of Wastewater Management
SSD	species sensitivity distribution
TMDL	total maximum daily load
TSD	Technical Support Document
µg	microgram
U.S.	United States
USGS	U.S. Geological Survey
WERF	Water Environment Research Foundation
WQS	water quality standards
WWTP	wastewater treatment plant

Foreword

The goal of the 2019 Frequency and Duration Experts Workshop was to gather the latest scientific information about frequency and duration related to aquatic life criteria for the protection of aquatic communities. This workshop represents just one in a series of the EPA-led efforts to inform aquatic life criteria development with the latest scientific thinking, and ultimately, to provide the most up-to-date guidance to state and tribal partners. The workshop was designed to be a critical thinking and information gathering exercise. Therefore, the workshop proceedings provide a record of the workshop discussions and experts' opinions but do not contain official EPA recommendations. EPA will consider the information discussed during this workshop when evaluating the state of the science for frequency and duration as related to the development of aquatic life criteria.

Executive Summary

Aquatic life Ambient Water Quality Criteria (AWQC) for toxics establish short-term (acute) and longer-term (chronic) chemical concentrations (magnitudes), averaged over a given time period (duration), that should not be exceeded more than the allowable number of times during a specified time period (frequency) to protect aquatic life. While magnitudes have varied across AWQC, duration components of AWQC have generally remained consistent, being based on assumptions described in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and their Uses* (1985 Guidelines), with one hour being the *typical* acute criteria duration and four days being the typical chronic criteria duration. However, for two recently updated chemicals (ammonia and selenium), longer duration periods have been used to reflect the specific behavior of these chemicals in the environment. The frequency component of criteria has remained the same across all aquatic life AWQC, with criteria recommended not to be exceeded more than once in 3 years.

Substantial additional science relevant to duration and frequency has become available since the 1985 Guidelines were released. EPA's Office of Science and Technology hosted an Invited Experts Workshop to review the current science relevant to the duration and frequency components of AWQC. The workshop was held on September 11–12, 2019 in Arlington, Virginia. It focused on identifying and evaluating relevant science that has become available since 1985, understanding the implications of this information on frequency and duration assumptions, and identifying unknowns that warrant further consideration. The experts invited to this workshop—from industry, academia, and state and federal government—had expertise relevant to frequency and duration.

Topics discussed during the workshop included:

- Duration and frequency history elements of criteria and their implementation
- Frequency:
  - Ecosystem disturbance and recovery characterization and evaluation
  - Alternative frequency characterization approaches
- Duration:
  - Chemical exposure duration considerations and effects assessment
  - Experimental and modeling approaches

Scientific information presented by the individual presenters and discussions during this workshop are summarized in these meeting proceedings. The information discussed during this workshop will be considered by EPA when evaluating the state of the science for frequency and duration related to the development of aquatic life criteria.

Introduction

EPA's Office of Science and Technology (OST) convened a *Frequency and Duration Experts Workshop* regarding Aquatic Life Water Quality Criteria. Invited technical experts in ecological toxicology and a range of technical areas relevant to frequency and duration, and representing academic, state, federal, and international institutions met in Arlington, Virginia, on September 11–12, 2019. The group of invited experts and affiliated observers were both present in the room. The goal of the workshop was to discuss and capture the state of the science about the effects of frequency and duration of exposure on the toxicity of chemicals to aquatic organisms. This workshop was not a Federal Advisory Committee Act (FACA) meeting. Participants did not reach consensus or give advice or recommendations to EPA; rather, participants discussed the science relevant to aquatic life criteria frequency and duration. This information was provided to EPA for further consideration as it relates to aquatic life AWQC.

The group was tasked with discussing the science that is relevant to frequency and duration of exposures to pollutants and the effects on aquatic organisms associated with those aspects of environmental exposure. The workshop was designed to provide an opportunity to share and listen to ideas, not to reach consensus on any particular topic; therefore, relevant discussion is included in this document. This discussion reflects expert opinion.

The invited experts included:

- Kevin Brix (EcoTox, LLC)
- Molly Colvin (Naval Information Warfare Center)
- Russ Erickson (EPA, Office of Research and Development [ORD], Duluth Laboratory)
- Bryson Finch (Washington Department of Ecology)
- Andre Gergs (Bayer AG, Environmental Sector Department)
- Sarah Kadlec (EPA, ORD, Duluth Laboratory)
- Jeff Manning (North Carolina Division of Water Resources, Water Quality Standards [WQS] Program)
- Chris Mebane (U.S. Geological Survey [USGS])
- Dave Mount (EPA, ORD, Duluth Laboratory)
- Adam Ryan (International Zinc Association)
- Travis Schmidt (USGS, National Ambient Water Quality Assessment project)
- Bill Stubblefield (Oregon State University)
- Chris Stansky (Wood Environment and Infrastructure Solutions, Inc.)
- Ning Wang (USGS, Columbia Environmental Research Center).

Two papers were developed by Chris Mebane in support of the workshop: *A Review of the Nature and Effects of Episodic Water Pollution and Implications for Aquatic Life Criteria Averaging Periods* (Mebane 2019a) and *A Review of the Nature and Effects of Episodic Water Pollution and Implications for Aquatic Life Criteria Averaging Periods* (Mebane 2019b). These papers (included in Appendix A and B, respectively) were distributed to the invited experts for consideration prior to the workshop to provide early and foundational information relevant to the topics being discussed.

Day 1 – Duration

Mary Reiley, EPA OST Health and Ecological Criteria Division (HECD), welcomed participants to the workshop and introduced herself as the facilitator. Background materials were distributed before the workshop to summarize the key science issues to be discussed. M. Reiley reminded the group that this is not a Federal Advisory Committee Act (FACA) meeting, which means that the group was not giving advice or recommendations but instead discussing the science emerged over the years relevant to frequency and duration. Present in the room were both the group of invited experts and affiliated observers.

Betsy Behl, Director of HECD, thanked the experts for their participation and provided background. She explained that OST is responsible for developing and finalizing aquatic life and human health criteria under section 304(a) of the Clean Water Act (CWA). These criteria are recommendations only, not rules, and are made public along with technical support documents. States review these criteria every 3 years and can adopt the criteria into their WQS, at which point they begin to be implemented for CWA purposes. They are used in permitting and impairment decisions. There are 47 aquatic life criteria. Although some are newer, many were developed in the 1980s. The guidance (USEPA 1985) used to derive these criteria was developed in EPA ORD's Duluth Laboratory. She noted that, historically, discussions have focused on the magnitude component of criteria and have spent less time on the frequency and duration components. The science that has been developed since the 1980s has prompted EPA to reexamine the science underlying the frequency and duration elements of the criteria. She reminded the group to keep in mind that these criteria are national recommendations that are intended to represent the country as a whole. States can still develop state and site-specific values. B. Behl stated she looks forward to the science-based discussion that will take place during this workshop and pointed out the diverse representation of perspectives that were present in the room. The decision to hold this workshop was partly inspired by questions that the state of North Carolina raised at Association of Clean Water Administrators-sponsored meetings.

M. Elias and J. Justice (OST) reviewed the workshop agenda and noted that magnitude, frequency, and duration are intertwined, making it difficult to separate them out for individual discussion. They noted the importance of the group fluidly navigating the areas on which they are focusing, while identifying their interrelationships, when necessary.

Presentation I: Duration History and Application (Discussion Leader: Chris Mebane, USGS)

During this presentation, several issues were discussed, including variability in flow and chemical concentrations. The first presentation began with a discussion of variability in flow in rivers and streams. The presenter noted that for most waterbodies, especially mid-sized ones, flow conditions are episodic, and many facilities discharge into small and mid-sized waterways. Exposure magnitudes, frequencies, and durations are intertwined, but magnitude attracts the most attention within the scientific community. Duration has received some attention in the past, while frequency has received little attention.

Selenium data from the Blackfoot River, Idaho, collected over 17 years (Zinsser et al., 2018), were discussed as an example of annual patterns of chemical concentration variation. In this dataset, selenium concentrations generally increased in the spring due to snowmelt and generally decreased in the fall. In terms of inter-annual patterns, some years had much higher selenium concentrations than others. Increased stream flow did not always correlate with increased concentrations of selenium, and vice versa. This raises the question of how organisms respond to short-term, high-concentration exposures as compared to longer-term, low-concentration exposures.

Water pollution can also be episodic on hourly timescales, particularly in smaller streams. Significant changes can be driven by respiration and dissolved oxygen. The presenter showed a graph comparing zinc and pH in a stream in Montana (Nimick et al., 2003), where the pH was highest in the afternoon when the zinc concentration was lowest; the opposite was true at dawn. In this case, the potential for zinc toxicity would vary depending on the time of day, based on fluctuations in both zinc concentration and potential bioavailability as influenced by pH. Another example of episodic water pollution noted was that of wastewater influents and effluents, which can vary widely throughout the day.

It was noted that other factors that modify chemical bioavailability and/or toxicity also can be episodic. For example, organic carbon tends to bind with copper and mitigate its bioavailability. A runoff event may release both; therefore, if these releases are synchronous, the increased organic carbon may mitigate the toxicity of the increased copper, while asynchronous releases could amplify toxicity. An example of a stream in Montana (Balistrieri et al., 2012) was discussed. In this example, organic carbon and copper spiked during a rainstorm and initially decreased together, but copper spiked again without the mitigating effect of carbon. However, it was stated that it is generally found to be the case that modifying factors are roughly synchronous with chemical concentrations. One instance where that is not the case is with ammonia and pH.

EPA is one of the few regulatory entities that recognizes that the exposure duration element matters in determining protective exposure scenarios. In the European Union, for instance, some substances have acute and chronic criteria, but most are focused on chronic exposures. EPA, on the other hand, adopted a "two-number" criteria approach in 1979. This included a 24-hour average and a "not to exceed at any time" criteria. EPA's 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and their Uses* (1985 Guidelines) (USEPA, 1985) revised these averaging periods and gave more explicit rationale on durations and return frequencies for criteria. These include a 1-hour average for short-term acute episodes, because some substances can have rapid toxicity, and a chronic 4-day averaging period for longer-term exposures. A key concept was that the averaging period for both criteria should be shorter than the typical test used to derive them, because a test is trying to capture the most severe concentrations. In practicality, toxicity tests often have a minor degree of fluctuating exposures.

In 1991, EPA released the *Technical Support Document for Water Quality-Based Toxics Control* (TSD) (USEPA, 1991), which is a foundational document on the implementation of water quality criteria, pairing hydrologic design flows with criteria durations and exceedance frequencies for setting effluent limits. The lowest 1-day average flow in a 10-year period (1Q10) is roughly equivalent to the "acute" 1-hour average concentration with a one-in-3-year exceedance frequency, and the lowest 7-day average flow in a 10-year period (7Q10) is roughly equivalent to the "chronic" 4-day average concentration with a one-in-3-year exceedance frequency.

The presenter reviewed the example of cadmium's criteria history, with a focus on criteria duration. EPA's 1980 "chronic" cadmium criterion was a fixed 24-hour average with a hardness-varying not-to-exceed acute concentration (USEPA, 1980). In 1984, there were 1-hour and 4-day averages (USEPA, 1984). In 2001, without explanation, the criteria reverted to a 24-hour acute average, along with a 4-day chronic averaging period (USEPA, 2001). The 2007 copper criteria also used 24-hour and 4-day averaging periods (USEPA, 2007), though the 1999 and 2013 criteria for ammonia (a faster-acting toxicant) used 1-hour acute and 30-day chronic durations (USEPA, 1999; USEPA, 2013). The 2016 acute and chronic cadmium criteria returned to 1-hour and 4-day averages, respectively (USEPA, 2016).

The presenter discussed why the issue of acute criteria duration was of interest, asking "who monitors more than once a day?" It was suggested that dischargers facing a dynamic model-based permit may be monitoring more frequently than once per day. The presenter outlined hypothetical scenarios to demonstrate the influence of different averaging periods on allowable contaminant concentrations in simulated stormwater pulses. In each hypothetical scenario, the "stormwater" pulse magnitude was constrained by setting the pulse amplitude so that the maximum average concentration, averaged over different durations, just reached the criterion maximum concentration (CMC). The presenter noted that he did not know whether there really are dischargers that have dynamic model-based permits with short timesteps, and he hoped the workshop discussions would addresses this.

The presenter noted that there are different possible modeling approaches to time-varying concentrations. Some of the issues include consideration of the speed of toxic action. There are many different models, some mechanistic and some based on direct empirical observations, but the simplest model is likely the Mancini (1983) model. The presenter considered how much time had to elapse between two toxicity events for the events to be considered independent. For example, if there was residual toxicity from a previous event when the next one began, then the two events would not be independent. The Mancini model includes a *k* coefficient, which is the rate of detoxification. The inverse of *k* is the detoxification time, which is roughly the speed of action. The presenter then discussed an example toxicity test in which almost all the cadmium toxicity had ended after 48 hours. In this example, the 1/k value was 32 hours (not very fast-acting). If the toxicity data are available, then one can calculate toxicity at any point along a toxicity test curve. He then showed example graphs for sodium cyanide in fathead minnows, which is fast-acting (1/k = 1.8 hours), and zinc in cutthroat trout, which is slow-acting (1/k = 98 hours).

In reference to latent mortality (i.e., delayed mortality), the presenter warned the group to "beware the ghost of exposures past." Mancini's model assumes organisms die when exposures produce a critical accumulation in or on the body. However, organisms sometimes die after a critical accumulation occurs, even if the toxicant exposure has ceased. This is known as delayed toxicity. The presenter discussed a study by Brent and Herricks (1998), in which the authors briefly exposed organisms to pollutants and then transferred them to clean water for observation. For *Ceriodaphnia dubia*, 15-minute exposures resulted in mortalities for up to 24 hours later in clean water. Delayed mortality was observed in every cadmium and zinc test that was run, though it was increasingly less important in longer exposures, up to the longest exposure of 4 hours. Delayed mortality was not important, however, in tests with phenol, which is fast-acting. The zinc speed of action for the Brent and Herricks (1998) study ranged from 3.1 to 19 hours. The presenter also discussed R. Erickson's (EPA) work on delayed mortality, in which Erickson

(2007) tested copper on fathead minnows and observed delayed mortality following short-term exposures (defined as less than 12 hours).

USGS researchers have explored whether the Brent and Herricks (1998) latent mortality results were reproducible. A USGS-EPA team that recently designed similar testing with zinc on *Ceriodaphnia* and rainbow trout. In their study, organisms were exposed to zinc for between 1 and 96 hours and then moved to clean water. In total, they observed 272 test chambers. They included relatively high concentrations to elicit a response at short exposures. The control for this experiment was a traditional 48-hour continuous exposure test. The study was very intensive and involved seven researchers working from 8 a.m. to 12 p.m. The presenter showed the *C. dubia* cumulative mortality curve at 125 micrograms (µg) zinc per liter (L), which is the acute criterion for zinc. After a 1-hour exposure and transfer to clean water, they observed 65 percent mortality after 48 hours. After exposures of 3 or more hours, there was more than 90 percent mortality by 48 hours. *C. dubia* is quite sensitive to zinc, and effects were more pronounced with *C. dubia* than they were for rainbow trout. USGS researchers noted rainbow trout were not killed with just an hour's exposure to zinc in this experiment.

The presentation was concluded by noting that short-term episodic exposures appear to be environmentally important in some settings for some organisms and substances. For example, short-term exposures seem to be important for substances such as copper and zinc, polycyclic aromatic hydrocarbons, and for complex mixtures such as highway runoff. The presenter left open the question of whether this was true for other organics.

Expert Group Discussion of Presentation I on Duration: History and Application

Following Presentation 1 on duration, the group discussion began with a participant asking for clarification about what was meant by "environmentally important." The presenter answered that short-term episodic exposures have a toxicological effect and occur often enough to matter in the environment. Exposures and effects come together in a complex way. A participant also asked if short-term meant less than 24 hours, and the presenter explained that he was just using it as a relative term and had single events like storms in mind, but that less than a day was probably correct. Another participant questioned how the laboratory set the hardness levels used in the experiment led by USGS. The presenter, a co-author, said that they diluted well water to reach a hardness consistent with the toxicity literature.

A participant then asked about the presentation slide on the Brent and Herricks (1998) study, which included a curve fit to the data during the recovery period. This participant also asked whether that was really fitting the model since it does not show what exposures created the toxicity. The presenter responded that the graph was of final toxicity and that it was displayed that way to question the concept of an averaging period. This participant further commented that one would need a very complicated model to capture that phenomenon or else collapse the toxicity down to the original duration and look at it from that perspective. The Brent and Herricks (1998) study had little actual toxicity during the exposures. This participant added that one could either have a more complicated model that tries to model the delayed mortality, or else one could take all the delayed mortality and relate it to the exposure duration and then create the model. The presented graph was noted to be for presentation purposes, to demonstrate delayed mortality following brief chemical exposures.

A participant had questions about the *Ceriodaphnia* and rainbow trout delayed mortality tests, asking if the organisms were kept at different temperatures because temperature can strongly influence the uptake rates of the contaminants. The presenter stated that the rainbow trout were tested in colder water than the *Ceriodaphnia*, but no tests of varying temperature for the same organism were conducted, and that this would be interesting to explore further. In one of the original papers, the organisms were kept at the same temperature, but because *Ceriodaphnia* are so small, they were still the most sensitive. The presenter noted that he does not know if any more information about the Brent and Herricks (1998) study is available.

A participant noted that additional complications may arise from the interplay between total organic carbon and metals, because in stormwater they sometimes ameliorate each other to some extent. The possible co-variance between substances, and with pH, raises some questions about how laboratories should measure these variables.

A participant said that the experiments he has participated in, with pH and other substances, agree with the results shared by the presenter. For example, phenol mortality is very quick, but materials such as cadmium show relatively slower time to effect and delayed mortality. For example, it took over 150 hours for their cadmium LC_{50} (lethal concentration required to kill 50 percent of the population) data to change, and that differed from the 96-hour data by about a factor of 10. Other substances, such as copper, produce toxic responses more quickly and do not show as much variation with time.

A participant pointed out that most of the examples discussed in this session were metals, and that these are just a subset of all the pollutants of concern. He asked whether latent toxicity was primarily a metals issue, or if it also applies to organics. The presenter responded that he did not know, and some examples such as phenol showed little latent toxicity, but that latency could matter for organics as well, and that this would be a good topic for further discussion. A participant said that many of the adverse effects of pesticides and organic contaminants are irreversible; therefore, frequency and duration are not as relevant as they are for metals, which are more reversible. Another participant asked if there were data to support this, since metals are one of the few groups of toxicants being taken up by active transport as opposed to diffusion across cell membranes, to which this participant responded that there are fewer data for metals than for pesticides because the registration process for pesticides requires such data to be generated.

Presentation II: History of Criteria Duration Derivation and Underlying Assumptions (Discussion Leader: Russell Erickson, EPA/ORD)

The presenter began by recalling acute criteria/CMC averaging periods. In 1979 EPA specified a maximum concentration based on acute toxicity, but later determined this wasn’t the best approach, largely because duration was required for National Pollutant Discharge Elimination System (NPDES) permitting because permits are based on probabilities. For example, the 7Q10 low flow is applied in permits to be representative of chronic criteria duration. During development of the 1985 Guidelines, it was determined averaging periods needed to be substantially shorter than the test durations to preclude the possibility of fluctuations in time series. For example, “Do you have a 96-hour test with the fast-acting toxicant?” The CMC for 96 hours implicitly allows a much higher concentration for several hours or even one day. This can elicit mortality.

The participants were asked to consider an example for ammonia. One hour was selected to be a default duration to cover any fast-acting toxicant. Some kinetic models can be used in the calculations. One-hour averaging with a 24-hour period should be viewed in the context of the worst hour in a longer exposure time series. It is not an isolated one-hour exposure. Storms and spills at worst generally represent several hours of high concentration exposures. The 1985 Guidelines were oriented toward calculating simulated events. With a 4-day averaging period and 3-year return frequency, it created a disconnect for application in NPDES permitting. If the criteria were to be an expression of the toxicological and biological condition, then that would translate into application. However, there was not a desire to tie the criteria expression directly to implementation at that time. The chronic criteria/ criterion continuous concentration (CCC) averaging period had a similar concern in that it had to be shorter than a test duration, largely to capture sensitive stages that may occur for a brief duration within a longer test (e.g., swim up of larval fish).

Expert Group Discussion of Presentation II: History of Criteria Duration Derivation and Underlying Assumptions

A participant recalled the late 1980s, when most sampling was 24-hour composite samples. He asked if in his recollection it was correct to assume a design oriented towards a 1-hour average was actually being measured as a 24-hour average. Another participant replied that regardless of whether a runoff event lasts 1 hour or 24 hours, some NPDES permits in California (e.g., the Navy's NPDES Industrial Stormwater Permit) only require one grab sample to be collected. This approach presents challenges; the grab sample is less representative of the entire storm event. Where feasible, flow-weighted composite samples taken throughout the storm duration are recommended, but this is expensive and typically not required. Even better is pollutograph sampling with multiple grab samples tested throughout the storm duration. Other NPDES permits in California require composite stormwater sampling, which requires more effort, but increases confidence that the storm event data represent mass loading and average chemical concentrations. Watershed programs that relate to wet weather or dry weather are trying to include composite sampling.

A participant asked if defining averaging periods based on toxicology might be backwards in the sense that you might want to start with the available data. There are reasons to use composites, but if you are going to evaluate compliance using a 24-hour composite sample then perhaps a magnitude interpretation should be designed in the context of how compliance will be evaluated. A 24-hour average might be needed because the available data supports this approach rather than because toxicology dictates this is the relevant exposure period. A toxicological interpretation is then made. The presenter affirmed that he thought this connection should be made, but that the criteria should not be expressed with that connection. The presenter noted that he thought a criterion should address the toxicological perspective and then include translation as part of the implementation. The toxicological implications of the 24-hour composite sample would need to be considered somewhere, but the decision was made that it would not be considered in the expression of the acute and chronic criteria.

A participant asked where magnitude is defined and where the interpretation happens in permitting applications and wondered if the group needed to consider where that boundary is and defining magnitude in a way that better reflects what will happen downstream in the permitting process.

Presentation III: Duration Application in NPDES Permits (Discussion Leader: Danielle Stephan, EPA/OW/Office of Wastewater Management [OWM])

The presenter provided an overview of the NPDES permitting program and noted that duration is slightly more important in calculating permit limits than frequency. Duration and frequency are typically applied in permits through the selection of critical low-flow conditions that states have selected to use for acute and chronic water quality criteria.

The presenter described how NPDES permit limits are determined through several steps. For WQS, the water quality criteria that is going to be permitted is first determined. This step includes consideration of the magnitude, duration, and frequency components of the criteria of the standard for the effluent. The question "What are the critical conditions we make sure we are going to protect when we write the limits?" is asked and the information goes into a model to determine whether a limit needs to be developed for the permit. If the answer is "yes" based on this set of conditions, then the water quality-based effluent limit is calculated. Duration is considered as part of this calculation. Reasonable potential analysis is considered the starting point before water quality based-effluent limits are developed. Factors such as critical low flow for the water body are considered. Additional questions are asked, including, "What is the critical effluent pollutant concentration?"; "What is the background concentration in the receiving water?"; and "How much already exists before the effluent comes into play?" All the receiving water conditions are then put into a mass balance model to project the potential to exceed the criteria rather than whether the criteria are exceeded. The goal is to protect against the worst-case scenario.

The presenter noted that permits reflect a situation where in-stream concentrations of a pollutant will almost never exceed criteria magnitudes. The 1991 TSD specifies a model to generate a number to compare to the criteria. One must consider if there is the potential for an exceedance and if a limit must be calculated. If a limit is needed, the duration component is considered in the permit derivation. In the permitting program, limits are developed for the end of the pipe, not necessarily for the water body. These do not represent limits but rather the maximum allowable pollutant concentration in the effluent from a facility. These values can be based on a variety of sources, including criteria magnitudes and total maximum daily loads (TMDLs). The waste load allocation is typically the worst-case scenario but does not always directly factor in duration (as it is expressed in a criteria statement). For example, permits include daily, weekly, and monthly limits, while acute and chronic criteria durations are expressed as 1 hour and 4 days, respectively. The lowest limit for the permitting program is 1-day duration.

The presenter stated that, given the NPDES permit limit derivation process, acute 1-hour averaging periods (or any averaging periods less than 24 hours) are all treated as a 1-day averaging periods in permit calculations. Waste load allocations are calculated from the 95th or 99th percentile of a long-term average that is assumed to be a log-normal distribution. The coefficient of variation, representing the peakedness or width of the curve is calculated. This value is then used to develop an acute or chronic multiplication factor to transform the waste load allocation into a long-term average that reflects the duration component of the criteria. The acute multiplication factor typically is based on a 24-hour averaging period, while the chronic multiplication factor is based on a 4-day averaging period. The presenter provided an equation from the 1991 TSD that could be used along with the developed multiplier for a 1-day equation with no adjustments for any data at intervals shorter than 1 day. Chronic multiplier tables are developed for intervals of less than 30 days. Beyond 30-day intervals, adjustments are needed. This exercise is meant for short-acting pollutants, like toxics, that have shorter durations. For nutrients, it is unclear if this approach is appropriate. Some states are utilizing this approach.

In summary, duration is used when calculating water quality-based effluent limits. The TSD procedures are used for durations ranging from 1 day to 30 days. Less than 1 day and beyond 30-day durations, there are differences in the data and other methods that should be used, where they exist. Without explicit duration components in criteria, the default is 1-hour and 4-day to be conservative. Each pollutant in each pipe is considered using this approach and they are not combined into one assessment. A consistent approach is needed so the permits are not vulnerable.

Expert Discussion of Presentation III: Duration Application in NPDES Permits

A participant noticed that if a pollutant had a 1-hour averaging period and another pollutant had a 6-hour or up to a 24-hour averaging period, the NPDES permit calculation would not be impacted because the intervals are all 24 hours or less. The presenter agreed with this summary and noted that permittees are generally not sampling on an hourly basis. If hourly samples are available, they are averaged to get a daily value. If a permittee samples more frequently than required, the data must be provided. If a sample is taken during an event at a peak and no additional samples were taken, the permittee could face a violation even if there is no actual violation. Here it would be to the permittee's benefit to sample more frequently than required. The presenter added that EPA advises permit writers to be specific about the when, where, and how many samples will be taken, and that it is the permit writer's responsibility to be protective of the criteria. The permittee only must meet the limit in the permit.

A participant asked how these issues are considered for criteria that vary with conditions like pH. The presenter replied that an understanding of the waterbody is needed before the permitting value is established.

A participant asked if durations over 30 days or less than 1 day are impractical because it is overly burdensome to the discharger or if it is because the underlying calculations no longer work out correctly. The presenter offered that an appendix in the 1991 TDS might clarify this issue. Another participant replied that, with respect to the bioavailability factors discussion, it comes down to the permit writer needing a value, whether a low flow or high flow. The presenter agreed and noted that models are generally steady state and not dynamic, but a number to permit is still needed. Complicated science is being distilled down into one number that dictates how much the permittee can discharge into the receiving water. In EPA Region 4, a dynamic model is used for nutrients.

A participant noted the duration terms of 1 hour and 4 days never directly come into play in a permit. The presenter replied that a 4-day average is translated into either weekly or monthly limits within a permit. This translation is embedded in the equation. A participant stated that when you take a 1-hour average that is toxicologically relevant and where you want to preclude fluctuations, that translation is not done. The question is whether it should be done as part of the criteria. A 1-day average can be translated into a weekly or monthly requirement. He wondered why a 1-hour average cannot be considered if there is some idea about the variability of a system. A participant responded that the issue is the consideration of that frequency in the calculation, and to capture variability, rather than simply monitoring at that frequency.

The presenter noted that variability is taken into account in the effluent concentrations. There had been some argument that EPA should factor in frequency and duration if they are not already considered in the effluent variability. A question was raised about whether EPA should look at the 80th percentile instead of the 99th percentile or if there is somewhere else where variability can be factored in.

Presentation IV: Duration Application in Assessment, Listing, and TMDLs (Discussion Leader: Eric Monschein, EPA/OW/Office of Wetlands, Oceans and Watersheds [OWOW])

The presenter discussed a simplified CWA framework to monitor and assess water quality status and develop TMDLs for impaired waters. Section 305(b) of the CWA notes that states should report on the quality of their waters to EPA every 2 years. Section 303(d) refers to a state's impaired waterbodies. EPA has recommended for the last 20 years that these be merged into one integrated report. EPA defines impaired as the failure to support WQS. Threatened, although not a regulatory definition, is defined as currently meeting WQS but not likely to meet one or more WQS by the next listing cycle. If a waterbody is impaired, then a TMDL must be developed. The language states these designations be made "from time to time" with no further specification in the regulation, but subsequent guidance indicates designations be made every 8–13 years.

The presenter noted that a TMDL is the calculation of the maximum amount of a pollutant that a waterbody can receive and still meet the WQS with an allocation to point and nonpoint sources with a margin of safety. There are different considerations based on the specificity for monitoring and assessment. CWA sections 305(b) and 303(d) do not specify specific monitoring requirements for ambient water. This leads to discretion for the states to implement their monitoring programs. There are not enough monitoring assessment resources to cover every water body in the United States, so states make hard choices with limited resources to address their priorities and to collect monitoring information. There are some recommendations to states on what they might want to consider as they develop and modify their monitoring strategies. These were divided into core and supplemental indicators. The core indicator is the health of the biological community; supplemental indicators are used less frequently. If a core indicator specifies that a water body is not meeting a designated use, then the state may look to a supplemental indicator (such as toxic parameters).

Monitoring and frequency vary for permitting, but ambient water quality monitoring sampling generally occurs quarterly. Monthly sampling is considered good and weekly sampling is unusual. There are few examples of continuous monitoring. Generally, sampling is done via a single grab sample and compared to both the acute and chronic criteria magnitudes.

The presenter stated that magnitude, duration, and frequency are important components of a typical assessment framework. Terms like digression, excursion, and exceedance are used. Just exceeding the magnitude within a specified averaging is a digression, so long as the average concentration over that averaging period does not exceed the criteria magnitude. If the in-stream concentration averaged over the criteria duration exceeds the criterion magnitude, then two components are exceeded, and an excursion has occurred. When all three components of criteria are exceeded (i.e., more than one excursion has occurred within the specified exceedance frequency period), the waterbody is deemed impaired. States typically make an assessment determination using a limited amount of data. Quarterly grab samples are used to represent an average concentration. The averaging period of both the acute and chronic duration is considered as part of the WQS. CWA section 303(d) does not prescribe which TMDLs need to be completed by a specific time or specific approaches for developing TMDLs. States consider a variety of factors to determine timing and approach, including resources and complexity of the approach. For example, for a dynamic model the internal capacity will apply a simple or dynamic approach. In the past, litigation drove the development of TMDLs. There are over 30 states with TMDL-related lawsuits.

The presenter noted there are both simple and complex approaches for addressing TMDLs. Some states use simple approaches due to resource issues related to data availability, and historically, there have been timing constraints. A simple assumption for duration is one grab sample can represent a 1-hour and a 96-hour average. The amount of reduction needed to meet the standard is the existing load minus the TMDL. Magnitude is usually determined with the 96-hour chronic component to provide a conservative approach.

The presenter noted that states sometimes employ complex models that have daily or hourly time steps. This approach can be used to help evaluate the frequency component or calculate a running 96-hour average. A 1-hour output equals a 1-hour average in the model. The complex models are very data intensive and are not commonly used in the TMDL program. The presenter suggested that, in the future, states should be provided with more time and flexibility to contemplate scaling the tool to the problem and perhaps using more complex models with more time to collect a more robust dataset to calibrate that model.

Expert Discussion of Presentation IV: Duration Application in Assessment, Listing, and TMDLs

A participant asked how the complex models are parameterized and what actual measurements are included in the sample set. The simple approach includes monthly grab sampling and a comparison to the 1-hour and 96-hour averages. The participant wondered if the complex models are calibrated to allow for interpolation. The presenter suggested that detailed questions related to modeling be collected so he could provide more follow-up later. The presenter was uncertain if application of a complex model that was not calibrated would move forward through the approval process. A participant shared that one of the models he works with is complex and requires a lot of field data for calibration, and that historical data are also used. The participant indicated that in his case the supporting monitoring program is very intensive with a great deal of water quality sampling to capture spatial and temporal changes.

A participant recalled that existing criteria were developed with an eye toward implementing them with permits. This is reflected in language in the criteria derivation procedures that assumes that criteria exceedances would generally be small (i.e., less than a factor of two). He wondered about the use of permits for stormwater events or other events that are beyond that original mindset and how these applications would affect the development of future criteria. To address more episodic events like stormwater flows, criteria approaches would need to be expanded beyond the original conceptual association with wastewater permitting. Additional/different science would be required to meet the conceptual protection goals stated in the criteria derivation methods if more temporally variable scenarios (e.g., exceedances much greater than a factor of two) are to be addressed. The practical applications of criteria are an important design parameter for criteria derivation approaches.

A participant said that capping the magnitude of exceedances (e.g., at two-fold the criterion) has an influence on designing an appropriate averaging period. Effects of exceedances are easier to envision if one has an idea of how high it could go. If the peak is limited, then longer averaging periods might be supportable even for a short-term event for a fast-acting toxicant. A participant noted that the number we come up with must be protective of the designated use and, if the designated is aquatic life, then that is the focus.

A participant pondered if it would be useful to have criteria associated with multiple averaging periods (e.g., 1-hour average, 6-hour average, "X-hour average," etc.). Users could then choose which of those numbers to focus on in their analysis. For example, permits could be based on the conditions they are trying to assess (e.g., stormwater discharge versus an industrial wastewater process discharge).

A participant stated that complex models are sometimes applied. He asked how one can digress from a magnitude without exceeding the averaging. The presenter surveyed EPA regions about their use of models before this meeting and learned that complex models are used infrequently. For a digression, one sample is examined. If there is an instant maximum "never to exceed" value, then one sample tells the entire story. As soon as duration is added and there are multiple samples, excursion becomes a consideration. Sampling across multiple years introduces consideration of frequency and a waterbody can then be listed as impaired. This terminology was developed in the early 2000s to help manage the discussions around these issues. These terms were created to allow for productive discussions and prevent the use of terms interchangeably. A participant noted that if sampling is only conducted once every 3 months, a digression is an excursion because results cannot be separated in the absence of more frequent or continuous monitoring.

A participant asked if a grab sample is treated the same as a 96-hour average and if it should be considered an excursion. The difference between an excursion and exceedance is understandable because of multiple samples taken over many years. It is unclear what the difference between a digression and an excursion is if any measurement is treated as being the 1-hour and 96-hour average. The presenter replied that if you think in terms of aquatic life, it means a sample concentration is above the criteria value more than once every 3 years. If there is one sample and its concentration is above the criteria value and the value is interpreted to cover the duration, then there is an excursion. The participant asked where a digression would be observed. If any exceedance of the criteria is considered an excursion, he wondered how there can be an excursion before there is a digression with a grab sample. The presenter replied this terminology is general and not specific to toxics. Flexibility is needed with continuous monitoring.

A participant noted that biological data in most states are only collected during the summer, in which case the sampling event is annual by definition. He asked how this applies to the biological data when the chemical data are being collected more frequently. He also asked if one could decide a waterbody is impaired if the biological data support this determination, but the chemical data do not. The presenter replied that impairment is failure to support one or more of the criteria, so an exceedance using either type of data would trigger impairment. If only biological data are available and support impairment and chemical data do not show an exceedance, the waterbody is still categorized as impaired. The data types are individually applicable.

A participant noted that, in most cases, the biological criteria are for aquatic life. The logic and kinetics behind the models used to derive aquatic life criteria should be consistent with the biological monitoring endpoints. This is currently true for fish models, but not for the crustacean models. A participant pointed out that there can be an annual biological assessment for parts of the aquatic community that have been perturbed but recovered by the time the assessment is conducted. If a fish community is significantly perturbed, this would be evident in an annual survey. A participant added that if the biological community is disturbed and all criteria are being met, things would not be considered “okay.” If the numeric criteria are exceeded and the biological community is not perturbed, things would also not be considered “okay.” The way the criteria are currently expressed might not be effective for diagnosing biological community impairment because of some of the disconnects this group has discussed. However, that does not mean that the listing will not be correct. A participant noted there is a need to figure out what the impairment is and what is causing it. Another participant added that if one waits until a biological impairment has occurred, it is too late and an excursion and/or exceedance event has been missed. The goal is to protect 95 percent of the taxa most of the time. She suggested flipping the question from consideration in terms of application to consideration in terms of science. A question to consider is how much the 1-hour duration can be exceeded while still being protective. Another participant pointed out that the 95 percent protection is just deemed adequate and is not an expectation that 5 percent of taxa are impacted. A participant responded that the original assumption was that the exceedances would be small. Another pointed out that unlike other arenas, safety factors or uncertainty factors are not present. A participant added that the statement implies the longest averaging period that is supported by the science should not be used. However, there is a practical desire that the average periods be the same.

Presentation V: Exposure Duration and Effects Group: Part I (Discussion Leader: Adam Ryan, International Zinc Association)

The purpose of this session was to discuss the observed impacts of exposure duration and fluctuations on toxic effects. The presenter began by sharing examples of phenomena seen in the data, some of which had already been shown to the group in the pre-workshop materials and in the previous presentation during the Workshop Discussion Initiation, in Presentation I. The presenter posed the following questions for consideration:

- How different are observed effects from exposures of different durations versus from standard constant exposures, based on average concentrations over the same duration?
- How much of an impact does latent response have on acute and chronic effect concentrations?
- What testing protocols or modifications to existing protocols are needed to better document the impact of exposure duration and variability on toxicity?

The primary source of material presented was from the Water Research Foundation (formerly Water Environment Research Foundation [WERF]) reports authored by Diamond et al. and written in the early to mid-2000s. The presenter had conducted some of the episodic exposure modeling associated with their work on copper.

Discussion began with the dictionary definition of duration, which is a continuance in time or the time during which something exists or lasts. In the context of criteria, this presents the questions “over what period of time (averaging period) should the instream concentration be averaged for comparison with criteria concentrations?" and "what is the duration of exposure at a given concentration beyond which adverse effects are expected?"

In short-term (i.e., acute) exposures, there are many examples of toxicity apparently increasing with exposure duration. As examples, the presenter showed two *speed of exposure*, or *time to effect* graphs that showed effect concentration decreasing as exposure duration increased. For comparison, he included two graphs that showed the same thing (Bailey et al., 1985), but presented data as the ratio of LC_{50} at each time point in a test over the LC_{50} calculated in the exposure. The ratios were much higher for short durations of exposure for copper sulfate and acrylonitrile than for ammonia, chromic acid, chlorobenzene, or xylene. This observation provided significance to the question of making generalizations.

These graphs did not tell us what the effects would be over the duration of a standard test if the exposure duration was limited to a fraction of the test period. These were all continuous-exposure tests and LC_{50} values were calculated at different time points where the organisms or responses were observed. He asked the participants to consider experiments in which organisms were exposed to a brief pulse and then moved to clean media. The presenter reviewed the results from Zahner et al. (2009), which indicated that the LC_{40} (lethal concentration required to kill 40 percent of the population) for a 96-hour continuous exposure of copper at 50 µg/L is roughly the same as the LC_{40} for 9- or 18-hour pulsed exposure at the same concentration.

This raises the question of how best to express the concentration during a pulsed exposure. The LC_{40} is similar to the continuous exposure test only if the maximum 1-hour average is used. He asked the participants to consider what might happen if a different average was used. If the pulse duration was averaged over the 96-hour test period (e.g., 9 or 18 hours at 50 µg/L over 96 hours), a lower LC_{40} would be observed, and the chemical would appear much more toxic. The presenter asked whether it was even appropriate to use a 4-day average, or if organisms are responding to the average.

Expert Discussion of Presentation V: Exposure Duration and Effects Group: Part I

The presenter discussed the results of Ivey and Mebane (2019). The authors used a test with a higher resolution in the first 24 hours. Is this important for answering the question "are finer timescales better, even though they are less practicable?" A participant pointed out that it is feasible to conduct an experiment with more observation periods. However, running multiple tests—like the protocol used in Ning Wang's study of zinc exposure in *Ceriodaphnia* and rainbow trout with multiple observation times—is more challenging.

The presenter noted that the results from the Ivey and Mebane (2019) study of rainbow trout are generally consistent with the results Zahner et al. (2009) observed in fathead minnows; however, results from studies of *C. dubia* indicated a different pattern of response with a significantly higher mortality rate. A participant stated that the number of replicates should increase as the number of samples increases. The presenter added that the WERF reports from Diamond et al. recommended increasing the number of observations to increase the statistical power. Another participant countered that if you increase the number of observations that one is looking at using a finer time scale, then more replicates might not be needed. The participants discussed different scenarios for interpreting experimental results based on the number of individuals and number of observations. A participant pointed out that the uncertainties for survival curves are different from uncertainties for LC_{50} curves. Survival curves need more test organisms.

In addressing the presenter's question about resolution in experimental design, a participant emphasized the importance of having observations that adequately populate the response curve. Observations early in the study period are needed to fill the early part of the curve (with "early" being a relative term), otherwise the curve will not provide information on kinetics. The presenter added that the term "early" is dictated by what the experimenter considers important for different chemicals, organisms, and endpoints. He wondered what practical approach could be devised to observe effects at four different nonmortality endpoints. A participant pointed out that these endpoints might have similar kinetics depending on the mechanism.

The presenter asked the group to consider what the Ivey and Mebane (2019) results suggest about the differences in kinetics between rainbow trout and *C. dubia,* given that their different responses are for just one chemical (zinc). A participant asked whether an immobile *C. dubia* should be considered dead while another participant responded that the mortality definition is problematic with this species. Immobilization may precede death.

A participant asked if others were considering ionic exposure because rainbow trout and *C. dubia* have different surface areas and surface area to volume ratios. Kinetics could be impacted by these differences at higher temperatures. Another participant agreed that these differences could affect the kinetics of uptake, but this discussion also considered the kinetics of the mortality process. A participant noted that it is the ion turnover rate (i.e., the surface area to volume ratio) that is important.

A participant suggested there might be value in discussing these issues for chemicals that are well understood, but this is and will not be the case for many chemicals. One of the strengths of the original 1985 Guidelines is that an understanding of how the specific chemical worked was not needed to develop criteria values. However, this presents a corresponding weakness, in that criteria derivation can be insensitive to additional knowledge of how a toxicant acts, because it relies only on empirical data. One should deviate from the basic derivation procedure to incorporate compound-specific information. This commenter asked the group to consider the following questions: Should the averaging period for contaminants with known toxicokinetics be different from those with unknown mechanisms? Or is a one-size-fits-all approach better? A participant shared that delayed mortality in *C. dubia* was more pronounced in experiments with zinc than in studies using cadmium and phenol. Another participant asked why the delay of death would be relatively higher for *C. dubia* than for fish, to which another answered that this observation might be due to the slope of the time curve. The participants discussed interpretation of the study results and it was pointed out that rainbow trout are more resilient whereas *C. dubia* are more sensitive to zinc exposures.

The presenter asked the group to refocus on his question of "is it reasonable to make generalizations based on a single chemical, given that it affects these two species differently?" A participant replied that incorporating size and temperature may be the best approaches, and he cited some of his previous work in which fish size proved to be important in determining mortality. Even small changes in size had an effect. Size is also closely related to age. Another participant shared some of his research in which juvenile trout were found to be less sensitive than adults.

A participant stated that for criteria you are trying to generalize about communities that can have significant variation within them. A longer duration test may capture that variability and produce a meaningful result, but short duration tests may produce results which are difficult to compare. Another added that the steepness of the concentration slope for a toxicant can impact what is observed in these experiments, while another shared that, in her research, she struggled to set a pulsed concentration that would produce a target amount of mortality. Results were inconsistent. Independent of concentration and test-to-test variability, there is uncertainty when the concentration-response slope is very steep. Even small details and the characteristics of the target chemical can affect how experiments play out. A participant agreed that it is very difficult to come up with concentrations that produce these partial effects over long durations. Another participant added that the shape of the species sensitivity distribution (SSD) flattens out if there is extrapolation across a large community. When studying chemicals with steeper toxicity slopes, the specificity across taxa is higher. This indicated there is wider interspecies variability in sensitivity.

With respect to the concentration-response curves, it was noted that species with high inherent sensitivity, such as *C. dubia*, tend to have a steeper concentration gradient, and a shorter time to action. These issues run together. The presenter wondered if this issue is important given that the more sensitive species have a shorter time to action. A participant added that reconciling the differences between ceriodaphnids and fish may not be feasible, while another asked if there were situations in which fish were more sensitive than ceriodaphnids. A participant noted he has conducted research in this area. He compared several species of ceriodaphnids with fish and found that the lower the threshold, the higher the mortality and the faster the kinetics. If the metabolic rate of a species is known, other kinetic parameters can be predicted to gain a sense of the SSD. He agreed to provide the reference for this study (Gergs et al., 2015).

A participant, M. Colvin, stated that their work with marine species related to the discussion on the timing of exposure. This work involved changing the duration of the short-term pulse and varying when the pulse occurred within the first 24 hours. Within the experiment, organisms were exposed to a chemical during the first six hours and then moved to clean water. In another scenario, organisms were exposed only from hours 6 to 12 of a 24-hour period. Significant differences in the sensitivities of sea urchins, varying with the timing of exposure, have been observed. The mode of action or why there was increased or decreased sensitivity in that first 24 hours was unknown. However, pulsed exposure appeared to be an important consideration for short-term durations. The participant clarified that this work was funded by the Department of Defense's Environmental Security Technology Certification Program, the Navy Environmental Sustainability Development to Integration Program, and Navy Region Southwest. The overarching goal of the project was to find an environmentally relevant stormwater compliance testing method. The group considered rainfall and discharge analyses to understand the episodic behavior observed in southern California. For example, rainfall in industrial use areas with little to no permeability results in a situation where rainfall essentially equals the discharge of interest. As part of this work, rainfall analysis was conducted across the United States at different locations to determine the 50^{th}, 75^{th}, and 95^{th} percentiles for rainfall duration, which they are using as a proxy for discharge over a 96-hour period (the typical acute test exposure duration). These rainfall durations were then applied to episodic exposure durations in the toxicity tests. A complete refinement of the pulsed exposure protocol is underway for four species and several analytes, including copper, zinc, and organic bifenthrin. Multiple laboratories are involved, including M. Colvin's, C. Stransky's, University of California at Davis, and Loyola University in Chicago. With the help of these researchers, the protocol is being refined and a standard operating procedure is being developed for the new method. This method will be validated using real world samples: they collect from their dischargers, test them in the laboratory, and concurrently do standard toxicity tests (i.e., the standard continuous exposure test). In-situ exposure duration tests are performed using a technology called the C-ring, which mimics a laboratory test. It is a caged protocol that allows the same animals used in the laboratory to be placed in the receiving water environment. They are still evaluating this data and will conduct an inter-laboratory calibration study with commercial laboratories. Another goal of this project is to conduct outreach and to understand the state of the science, which this meeting partly satisfied.

Participants acknowledged the need for toxicity testing, even if just to obtain a baseline level of understanding. Including every chemical or every interaction between chemicals in this testing might not be needed. It was noted that the use of toxicity testing for permitting, end-of-pipe monitoring, and in receiving waters has increased in California. The state has been grappling with how to interpret chemical concentrations measured at the end-of-pipe during a storm event and how to compare that data with water quality criteria and toxicity. Data for storm discharges and receiving waters were considered. The dynamics between tides and currents were considered and data were compared with what is measured at the end-of-pipe. The ultimate goal of the work is protecting the beneficial uses of receiving waters. One of their major considerations is modifying their testing methods to mimic these storm events and episodic exposures. A participant stated that current static Whole Effluent Toxicity methods are not appropriate for episodic exposures such as stormwater discharges.

A participant noted that in-situ work is being added into compliance monitoring as a means of endpoint validation. When toxicity is measured at the end-of-pipe but not in the receiving water, there is uncertainty about whether the critical period has been missed or whether the receiving water is representative of the most critical conditions during a storm event. The in-situ method is being refined as a means of validation, provided there are no confounding factors (e.g., loss of test animals in the field environment due to predation, storm events that damage equipment). The goal of protection is thought to be met if a pulsed exposure is conducted.

It was asked if part of the laboratory protocol was to transfer the test animals to clean water after the pulsed exposure to look at the rate of mortality. M. Colvin responded that the animals were transferred to either clean water or receiving water. Observations were carried out to 120 hours to allow for observation of latent effect associated with continuous exposure. For this experiment, purple sea urchin and mysid shrimp acted as the test animals.

A participant asked how pronounced the latent mortality was. M. Colvin responded that little latent mortality has been observed following pulsed exposures to copper and zinc. Some latent mortality has been observed with continuous exposures. Another participant added that other tests are currently being conducted with *Ceriodaphnia* and *Hyalella* and that some latent mortality also has been observed amongst those species. M. Colvin clarified that with the sea urchins, the 96-hour test was conducted, and that the discharge samples used were grab samples that were co-collected as part of the NPDES monitoring program at this naval base.

Another participant recalled an investigation (Tobiason et al., 2003) on which he collaborated that studied runoff from the Seattle-Tacoma airport. Analytical zinc concentrations were spiked because of dry spells. The participant noted that a first rainstorm may only last an hour or two but could have a huge impact on the area. Another participant added that during the first rainstorm event, a spike in copper, zinc, cadmium, nickel, and even mercury (in a few cases) has been observed following the first flush. Spikes vary among storms and the activities performed at a specific site. Another participant shared that samples have been collected at different points in time during a storm, and variations in the chemicals and toxicity were observed. A "first flush effect" is typically observed at some industrial sites. In some of the watersheds and off the highways, however, there can be a delayed effect and it takes some time for certain chemicals to mobilize. Passive sampling has also been added to examine time-average concentrations. If one thinks of toxicity as integrating the effect over time, passive samplers are great at doing the same thing. The pulsed/variable nature might be missed, and the only way to capture that is to conduct pollutograph sampling to have measures taken over time. A participant pointed out that their pulses have all been with a single grab sample.

Presentation VI and Expert Discussion: Exposure Duration and Effects Group: Part II (Discussion Leaders: Adam Ryan, International Zinc Association [first presenter] and Kevin Brix, ECOTOX [second presenter])

The first presenter for this session discussed modeling strategies for exposure. The presenter noted a paper by Jager et al. (2011) that described the framework of modeling exposure and duration, and that Ericson et al. (2007) provided a great overview of the issues in his discussion about exposure duration and effects. He referenced several key models including single compartment lethal accumulation models, damage repair models, multiple mechanism of action models, multiple compartment models, and a potential combination of these models. He noted that Jager et al. (2011) provided a framework for how to effectively use these models in conjunction with one another. This presenter discussed a schematic from Jager et al. (2011) that described the General Unified Threshold model of Survival (GUTS) framework, which relates an exposure concentration and external concentration to an internal concentration via toxicokinetics. Kinetics dictate uptake and elimination. Within GUTS, there is an optional component called "damage" that may not be necessary if internal concentration can be directly related to a threshold. This can be tied to hazard rate and lead to a survival function. However, the damage component is available as an option to link internal concentration and survival when necessary.

A participant shared that there are two different views about how mortality occurs. The first view is that each individual has a susceptibility, either damage or accumulation, which determines individual tolerance. The other view is that an assemblage of organisms is treated as a population with a certain probability of death in time, which is more tied to the hazard format. The Jager et al. (2011) model combines the two views. He noted that use of this combined approach has not been demonstrated because there are too many parameters. Another participant added that there are examples where users have tried to combine these mechanisms, but the addition of parameters presents an issue, and it was discussed that the multiple correlations between parameters presents challenges to interpretation, and that many datasets are not well suited for use in the combined modeling method.

The presenter stated Jager et al. (2011) provides more detail on toxicodynamics and describes assumptions related to the parameters. For the purposes of this presentation, only the idea of a linkage between external concentration and survival and associated difference considerations are addressed. Notably, if the time course of survival cannot be described by the time course of internal concentration, something is needed to link them. This could be the case where there is a latent effect. If there is latent effect, a damage pool or a two-compartment model could be useful to allow for damage or accumulation to occur on a different time course or at a different location after the pulse has stopped. This allows damage to continue even after the external concentration has ceased and the internal concentration begins to decrease, based on the relative kinetics of the internal concentration and accumulated damage.

It was pointed out that this cannot be compared to compartment bioaccumulation models because there is no feedback in terms of concentration. Another participant confirmed that, in this case, waterborne exposure can cease but an internal concentration capable of producing damage is possible. He presented a scenario involving copper exposure where the gills of an organism have been partially dissolved, but it takes time to leak enough ions to reach a critical loss. Therefore, there is not necessarily ongoing damage, it just takes time for the damage that has already been done to be expressed. The presenter replied that damage is an integrated component of physiological and biochemical processes. Therefore, there may not be a direct link between modeled damage and observable physical damage to the organism. It was added that the damage would be the ion leakage that leads to death. Another questioned whether the assumption is trying to link the damage directly to residual internal concentration, or if the assumption means that damage is continuing. In many of these cases, internal concentration is not measured. Damage occurring in the organism may not be due to a “whole-body” metal accumulation within the organism, it may just be metal accumulation at the gills causing damage. Therefore, the internal concentration may not be linked to damage. A participant noted that, with this framework, nothing prevents the addition of more components. Another participant recalled a recent paper that measured isotopic concentrations of zinc through the development of an aquatic insect to differentiate between what is toxicologically relevant and what is inert. The study demonstrated that one reason why this threshold does not apply to aquatic insects is because they do not show a large amount of metal upon emergence. This disconnects survival rates between adult and larval phases. Another participant replied that the model is just an approximation and may not capture all the details implicitly imbedded within these scenarios.

The presenter provided a conceptual framework to think about the model: If a threshold is exceeded, an individual has an increased probability of death. Thus, the exposure concentration can be related to survival while also considering the kinetics of the internalization and possibly the kinetics of damage. Models can be combined or customized if there are multiple mechanisms involved. He presented a model and associated results that he and a colleague implemented to explain episodic exposures of copper. Probability of death increased after the threshold was exceeded. With respect to these results, a participant asked how he dealt with growth dilution. The presenter explained that growth dilution was not considered but it would get buried in the fitted constants. In both cases, the internal and damage pools were scaled to the external. A participant noted that growth dilution will not matter if the kinetics are fast. The presenter confirmed that the kinetics in this example were very fast.

The presenter shared modelling examples that considered a higher concentration of copper with a 24-hour exposure. As the internal concentration increased, the damage increased above a threshold. The model also included a bioavailability component. The response was very fast, and the observed effect occurred by day two. The presenter discussed several more models with higher concentrations that resulted in more damage. Generally, the results were similar. Another model/experiment showed a larger variation in response later in exposure times. Unlike the previous single-pulse experiment, this experiment was a double pulse. Model I showed that survival increases, then decreases, then increases again. Damage increases, decreases, then increases again. The presenter discussed a series of models with increasing concentrations of copper, stating he completed the 50–60 tests as part of this work. The model and parameters were developed using all the available data at once. The model can predict 70 percent of variation. He considered his model complete, according to Jager et al. (2011), but wanted to reduce the model parameters because he does not believe the damage component is necessary. The model pools are similar to the damage model and he believes the damage component is irrelevant. The model might be more complex than necessary.

A participant asked if there is a need to routinely include frequent observations (e.g., not just daily observations) in constant exposure experiments and raised the question: "What would be the characteristic of the response to constant exposure that would cause the model to predict latent toxicity?" A different participant added that this model can only predict latent effects if it is due to some process such as damage dynamics or bioaccumulation, while another participant asked if anything can be inferred about latent mortality if one only has data from continuous exposures. A participant responded that his gut reaction is that it would be difficult to definitively make such an inference. One of his datasets shows a kink in the LC_{50}, suggesting one would need to determine if there are two mechanisms of toxicity and if the situation is a two-compartment situation. It is hard to discriminate between possibilities, which could include latent toxicity. He did not think one could address latent toxicity without a longer short-pulse exposure. Latent toxicity would need to be observed.

A participant asked whether anything could be determined from the incipient level curves in terms of the slope of the effect during the initial phase of exposure. They wondered if in the event that the time is longer whether the possibility of a systemic toxin or a potential for accrued damage at a site is inferred; they offered an example where a chemical was being metabolized and excreted, but an adverse effect remained sequestered at the site of action. They also wondered if the curve could explain this. The participants were unable to answer this question at this time.

A participant asked the presenter why they were using 12-hour gaps in the study, and the presenter replied that there could be other examples that have more than 12-hour gaps, but he could not recall. Another participant offered another example focused on latent effects. If something like feeding inhibition with an initial acute effect is considered, followed by an internal concentration that causes an effect that is not actually seen, there could be latency occurring in 5–10 days. The presenter asked if, in this case, this observation would be treated as a second mechanism or as damage. A participant replied that, in this example, the focus was only on the decline in energetics over time. Another asked if exposure had to be withdrawn to see that effect, to which he responded that exposure did not need to be withdrawn, but observations on these effects via growth or feeding assays were needed. Another asked if a survival curve could be used, to which he replied that a survival curve on its own could not be used because one needs to know why the individual is dying. A participant inquired about survival and growth data, and whether these data could be used to obtain the same results with a 7-day fathead minnow survival and growth test; he answered that a longer study would be needed and noted that 28-day insect experiments have been used in other cases.

A participant described the results of one of his studies that suggested two mechanisms attributed to responses because the model flattens out and then dips down again. There may be alternate explanations because what is happening in the organism is extremely complicated. These models should be treated as a quasi-mechanistic tool that helps interpret the data.

A participant asked if, from a practical standpoint, there is a way for latent toxicity to be detected when it has not been explicitly evaluated. Another noted that, while they do not know the answer offhand, they are doubtful. They further emphasized the complexity of the organism and the simplicity of the model; however, with enough data on latent mortality, there is the potential to come to some conclusions about whether a chemical or organism expressed latent toxicity.

It was asked what would happen if a 7-day only trial was run that reduces biomass. Theoretically, a two-phase result would be observed. A participant pointed out that an issue with that experiment is the organism is undergoing so much development during that time, one is likely to see many phases that serve as confounding variables.

The presenter showed another slide related to the application of the model the participants had been discussing to demonstrate that these models can be applied to an exposure scenario to predict a response. These models can be used to determine what concentration, over a given exposure scenario, results in a certain level of response. He presented a table that demonstrated this application by predicting the concentration of copper that resulted in a certain percentage mortality for a given exposure duration.

A participant stated that the issues of latency and other extinctions only apply to short, sharp pulses of exposure. A broader pulse would allow for a stronger relationship between the type of data that are available and the expected outcome of the exposure.

A participant asked if, when talking about the most sensitive life stage where a lot of development is occurring, participants are considering different models for different life stages, or if they are just targeting the most sensitive life stages. Another replied that his initial thought is that the 95th percentile species is being modeled, which does not actually exist, and its kinetics is based on other information. Nothing else is being explicitly monitored except the lower end of what is believed to be the sensitivity distribution based on the tested species. For certain taxa, that distribution almost never includes the most sensitive life stage. That remains unknown unless studies have shown that the distribution is representative.

The second presenter for this session, K. Brix, reviewed the three key discussion questions for this session and asked the participants to consider the question: “What additional data requirements might be needed in constant concentration toxicity tests to support modeling?” A participant stated that the complete data set from 24–48 hours of observation should be reported. If the mortality at each concentration and at each time is reported, that would go a long way. Another participant added that the toxicity test should be long enough to assess full toxicity, so a 48-hour test is probably insufficient, while another agreed with this need but raised the issue of feeding when longer time intervals are employed in a study.

The second session presenter summarized that longer tests and reporting of all intervals are likely needed. The group added that more information is needed on latent mortality with tests short enough to be meaningful and clarified that his suggestion is not intended to imply it should be standard test protocol, but instead that more information is needed to know whether it is important or not. This presenter asked if this issue should be the focus since these criteria are being used for other applications. A participant noted that the purpose of criteria is to protect ambient water.

A participant tried to summarize the discussion. The 1985 Guidelines have as an explicit underlying presumption that most exceedances are small (i.e., less than 2-fold). This assumption is hugely impactful in determining how one might establish appropriate averaging periods. Another participant cautioned that this point presents the possibility that there is a whole world outside that box, which may not be the case. Another asked what a pulse might look like from an industrial wastewater facility.

Presentation VII: Exposure Duration and Effects Group: Part III (Discussion Leader: Russell Erickson, EPA/ORD)

The purpose of this session was to discuss the application of toxicity models to toxic effects from time-varying exposures. The presenter stated that applications of the model are affected by uncertainty and variability. Models are approximations of reality and are sometimes useful; discretion should be used to determine when the model is useful. Toxicity model theory indicates an exponential decline from an infinite LC_{50} at time zero down to a plateau of LC_{50} at time infinity. The kinetic constant explains the exponential decline. This toxicity model was used to set an averaging period of 20 hours for earlier criteria. The final equation is a general form of the equation. “T” is the duration of the test which was used to set the criteria. If an LC_{50} of infinity is used, this is 1/k. Toxicity models can be used to set chemical-specific averaging periods. Delayed mortality affects “k”; specifically, if this model is based on the standard concentration test, “k” will be underestimated.

The presenter discussed a study of acute toxicity of copper to fathead minnows. Pulses of the same concentration were run on the minnows at hours 1.5, 4, 8, 12, and 24. At the study’s completion, exposure was lethal to the fathead minnows, both at the time of and after exposure. The LC_{50} from this study was compared to the LC_{50} from the constant exposure test. The study generated two sets of data: observations corrected and not corrected for delayed mortality. The uncorrected observations generated a “k” value of 0.68 per day, producing an averaging period of 1.4 days. The delayed mortality did not seriously impact the averaging period for this dataset. In both datasets, the averaging period was one day, which was the rationale for deciding to use a 1-day averaging period for acute exposures for copper.

In using these models to determine averaging periods, latent mortality should be considered for correction. A sufficient number of datasets should be analyzed to determine the uncertainty of the averaging period due to latent mortality.

Extreme value statistics are needed to determine the exposure level that should not be exceeded in an averaging period for a hypothetical concentration for a 1-year exposure time. Calculating this for permitting is easier because streamflow data and effluent variability are known; calculating this for compliance monitoring is difficult because a rare event should be detected.

The presenter noted that models can be applied to a time series to compute the level of effect for any day. For a particular average concentration, the probability of exceeding the mortality rate of 1 percent and 10 percent per day can be calculated given the variability. The average concentration can then be assessed, which would be more amenable to compliance monitoring and would be related to the risk of certain levels of effect. The copper data appear to be bi-phasic; one variation was a two-mechanism/stochastic model. The individual tolerances of the stochastic models deviate slightly.

The presenter noted that it is worth considering the extent to which toxicity models could serve purposes other than setting an averaging period. He presented the following questions for discussion:

- How can toxicity models be applied to the development of averaging periods to address the effects of time-variable exposures?
- Is an averaging period the most informative way to use these models to apply effect concentrations from standard constant exposure tests to time-variable exposures?
- What are other ways model calculations can be applied to characterize risks for time-variable exposures?
- Rather than the averaging period and frequency, how often are the predicted effects exceeded without an averaging period? How often are effects predicted by the model above an acceptable level? The model can be judged by these answers.

Expert Discussion on Presentation VII: Exposure Duration and Effects Group: Part III

A participant commented that concentrations presented for copper are high compared to a state standard. This dataset is almost irrelevant because the concentrations presented do not fall within the range of what could be permitted. The presenter agreed that the copper concentrations were high. There is a lot of information on acute toxicity and states regulate based on chronic toxicity. There is a lot of information on organisms that are not sensitive. Data for fathead minnows are more illustrative of approaches; lessons can still be applied from relationships in data for how to handle sensitive endpoints, which drive standards. Research is primarily focused on chronic toxicity, and acute toxicity is typically only used to establish relationships that are relevant to chronic toxicity. Toxicity modeling has been used for analyzing chronic toxicity and growth; although the equations are different, the principle is useful.

A participant asked for confirmation on the conclusion that the resulting LC_{50} from a discontinuous exposure was never lower than the continuous exposure in the presented examples. The presenter clarified that it depends on how the data are analyzed. If observed latent toxicity is presumed to have resulted from a 3-hour duration, instead of 20 percent mortality at hour three, there is 80 percent mortality once latent mortality is accounted for, which lowers the LC_{50} values.

A participant asked how an instantaneous concentration could exceed an average concentration if all averaging period approaches used continuous data from 48-hour or 96-hour exposures. If the instantaneous maximum is capped at the LC_{50} at the 50^{th} percentile, it is not clear that there could ever be a problem, unless averaging periods approach the total exposure period and mortality occurs over 14 days. There might be a problem if there are more extreme short-term exposures. A 3-hour averaging period would not use a toxicity test that uses 3-hour data; instead, the ending data would be averaged over the three periods. If 3-hour LC_{50} values are used to set a criterion magnitude, then the latent mortality would be critical. The presenter responded that for this organism, the goal was to avoid reaching the LC_{50}. If the 96-hour LC_{50} was one period, if it is assumed that no latent mortality occurred afterward, and if the averaging period is 96 hours, the context still has an effect; there is logic to setting the averaging period. A participant thought that the presenter's point emphasized the reason that a 2X limit to instantaneous chemical concentrations (relative to a criterion magnitude) within an averaging period is important. Longer averaging periods are dangerous if the exposure distribution within the averaging period has allowable large spikes because of long compensating periods. The presenter responded that the 3-hour LC_{50} could be lower because of latent mortality. The concern is to protect shorter averaging periods. The context does have an effect; a context of a 2X limit to instantaneous chemical concentrations (relative to a criterion magnitude) within an averaging period would cause a shift. The extent to which latent mortality is an issue depends on how much it affects the sensitivity.

A participant expressed concern that the presenter's example is extreme. The presenter responded that he showed two examples where this occurred. A 3-hour averaging period not counting latent mortality was only 50 percent higher, which is why a shorter averaging period was implemented. Latent mortality might make mortality even more extreme.

A participant mentioned that a paper (Gergs et al., 2016a) was published on the scientific opinion of the Toxicodynamic-Toxicokinetic model, in which it was proposed to calculate measures independent of the actual concentration and exposure time by integrating the entire profile of measurements or simulated exposure profiles. From this, a margin of safety value could be calculated instead of actual risk concentrations. It was also proposed that at the end of profile of a certain exposure with multiplication factors, the multiplication factor was needed to calculate LD_{50} (the lethal dose of an ingested substance that kills 50 percent of a test sample) or LD_{10} (the lethal dose of an ingested substance that kills 10 percent of a test sample) and the margin of safety. This information revealed how close a population was to being killed.

A participant asked if the types of models used for the averaging period could be used in a different context. He wondered if, for a frequency-based approach in which no more than 5 percent of observations can exceed criteria, these sorts of calculations would define the cap for the magnitude (instead of 2X the chronic criteria). The presenter responded that yes, hypothetically. If an exposure time series is being assessed, to the extent the models are deemed accurate, it can be determined how frequently effects would be exceeded.

Some participants discussed whether it would be easier to deal with variable averaging periods if the description of an exceedance criteria included exceedance magnitude. If the exceedance magnitude has a quantitative cap, an average can be calculated over a certain period. The presenter responded that averaging periods depend on what is being considered.

Some participants pointed out that it is more difficult to go longer than a 1 day averaging period; there is a 24-fold difference between 1 hour and 1 day. There may be room to do something more than 1 hour and less than 1 day, not just by fiat but by toxicity modeling results. The presenter responded that if assumptions are made about limiting excursions, then longer averaging periods could be determined.

A participant mentioned that the concentration function slope typically uses a factor of two, but this could be involved in describing the amount of expected impact from an excursion. Another participant stated that assuming the most sensitive species have the shortest time to effect and steepest dose-response curve, this is not an issue. He added that if the CMC is half of the LC_{50} and the curve is shallow, then exceeding the criteria will do nothing. Effects will not be expected until well past the criteria value. If the slope is steep, effects will be expected immediately. If the LC_{50} is exceeded, then a steep curve will lead to a higher mortality at a faster rate. These examples discuss LC_{50} relationships. The way to analyze these data are to fit the model to the entire dataset to predict any level of mortality. For a 24-hour averaging period, even for fast-acting chemicals, if exceedances are capped, the limits on “k” can be determined before problems occur.

A participant shared that they worked on sites where exceedances went well above the LC_{50}. NPDES permitted outfalls in small streams in watershed drainage areas with sufficient mining can experience high concentration spikes during times of flushing. This also occurs in industrial and Superfund sites. Another participant emphasized that the exceedance magnitude is rarely exceeded in wadable streams.

The presenter asked the group to discuss how to place bounds on episodes. How bounds are determined is dependent on the cap. The cap-level impacts what errors should be considered. When considering model uncertainty, the tendency is to analyze the dataset and the “k” value, but the uncertainty of the “k” value should also be considered. The uncertainty effects what averaging periods are used.

A presenter discussed toxicity tests that she conducted. She conducted Test A, which was a constant 7-day exposure of 17-day post-hatch fish to five concentrations of carbaryl followed by 7 days in clean water. The four other tests, Tests B through E, were standardized 6-hour tests with 30-day old minnows reported in the Fathead Minnow Acute Toxicity database in 1988 (USEPA 1988). These fish were less sensitive than the fish used in Test A. Concentrations ranges were higher in Tests B through E. She fit a simplified GUTS models to Tests A through E and compared predictions generated with the GUTS models for pulse exposures. Pulse exposure tests were conducted to use as a validation dataset; eight pulse treatments were conducted for either 24 or 48 hours using either one or two pulses. The model-fitted results and the calibration results reveal that the model predicted survival well within the calibration dataset. She stated that if the model is used to fit Test A results to predict effects from the pulse test, results vary widely because the fish in Test A were more sensitive than the fish in tests B through E. The model over-predicts toxicity. The challenge with the observed toxicity in pulse tests was that there was not a measurable effect from the second pulses; thus, the model does not predict any effect.

A participant stated that the focus should be on the effect concentrations, especially with steep curves, because a small shift in effect concentrations can predict a very different mortality. This is important to criteria. Another participant responded that this is an extreme example in terms of variability; the variability in input datasets makes it reasonable to expect more extreme examples. He asked if the discussant tried to fit datasets together to get better information on actual toxicity and a better estimate of total uncertainty of the dataset. The discussant clarified that she did this to an extent. She fit the model to Test A and to one of the other tests and saw that predictions did improve, but this seemed to be because sensitivity was averaged; the pulse tests coincidentally had the same sensitivity.

The group asked if only continuous exposure was put into the model, and if the model provided predictions about discontinuous exposure. Because concentration ranges from 0 percent to 100 percent mortality, which was small compared to the tight LC_{50} range, the most impact on variability comes from the potency estimate, not the toxicokinetic-toxicodynamic model. The experiment intended to determine response to repeated pulses changes at the known potency.

A participant asked why contaminants other than ammonia do not have a 30-day cap. Another participant responded that ammonia has this cap because there is enough data. A different participant asked if the approach could be applied more broadly. A separate participant emphasized that the objective is to protect designated use and water quality standards. A longer duration period can capture longer potential offsets.

A group member mentioned that fish early life stage (ELS) tests are an example of a test where ongoing test data (e.g., daily survival) may not always be included in the test reports. Another participant responded that he cannot speak for contract laboratories, but he has never done an ELS test without daily survival data. Many decisions made on appropriate averaging periods analyze the end effect of the aggregate of long-term, low-level exposure of windows of high sensitivity, potentially for 30-day data. A participant expressed concern that the 30-day data ignore important maximum daily data.

A participant stated that data collected from toxicity tests include survival data for certain species. These data are used to do toxicity identification evaluations to determine treatment effectiveness. Entire datasets are rarely reported. Total mortality within 24 hours is rare. Generally, there is a decline over a couple of days. Thus, data over time rather than solely endpoint data would be a useful addition. Short-term chronic embryo development determination tests do not allow the ability to look at data over time, only at endpoints. This should be considered when analyzing pulse/episodic exposure in addition to how more information can be obtained from different tests.

Participants noted that, with respect to assessments of use attainment and water quality monitoring, a longer duration of chronic criteria could get states over the hurdle of wanting to collect more data if one sample is a small exceedance of acute or chronic criterion. More data could reveal if all exceedances are the same. A longer chronic duration could be helpful for resource-limited community monitoring. Data could also determine if a 2X exceedance is equivalent to a 10X exceedance and what this reveals in respect to use attainment and impact to the aquatic community.

A participant asked if states always conduct grab samples for CWA compliance monitoring. Another participant answered that the sampling approach is dependent on the parameter. For example, North Carolina adopted dissolved standards and definitions of acute and chronic criteria for copper. For acute and chronic compliance monitoring, the state collects two samples per hour and averages the two results to obtain a single sample value. A participant noted that some coalitions of dischargers are disincentivized to collect more data because if they have two exceedances, regardless of how big the dataset is, they are impaired, because of the greater-than-one-in 3-hours approach.

A group member asked if a chronic 4-day averaging period would require at least two samples to be collected. Another participant confirmed that a chronic 4-day averaging period would require at least two samples, while another participant pointed out that a utility could sample once in 30 days and use it as their 30-day average if the sample meets criteria.

A participant stated that, based on the models, effects are well-predicted for exposure concentration “X” and duration “Y”, but questioned if this can be applied in a proactive manner to set appropriate magnitude and averaging periods. Another participant responded that this relates to incorporating the upper end of instantaneous maximum averaging period and maximum concentration. Bounds should be put on the range of episodic patterns that might occur within the averaging period to better tailor the averaging period. Toxicants rarely have averaging periods longer than 1 day unless there are caps on the exceedance magnitude. The interplay between modeling and recommending an averaging period depends on the top end of the pulse. Other issues (e.g., latent mortality) present challenges in the absence of a cap. This could be incorporated quantitatively using the models.

A participant pointed out that the implications of implementing a 24-hour average should be considered along with the universe of exposures to which this could be applied. A participant wondered if the 2X guideline assumption could be applied to higher X factors. Another participant responded that system variability may determine different averaging periods.

A participant asked if acute and chronic ratios could help identify contaminants as having a higher potential for latent toxicity. Another participant responded that this is possible. For chronic concentration, the acute criterion (i.e., CMC) caps the characteristics of a longer-term average. The idea that acute criterion caps the maximum exceedance magnitude is conceptually true, but the way it does so is highly variable.

Day 2 – Frequency

Presentation VIII: Office Director Remarks (Deborah Nagle, EPA/OW)

Deborah Nagle welcomed the participants to Day 2 of the workshop. She noted that EPA often hears from implementers of the NPDES program and the TMDL program that magnitude alone does not do them any good when they are trying to write permits or wastewater allocations. Frequency and duration are important for protecting aquatic life. She noted that having the conversation was not where the work stops; rather, it is what comes out of the conversation that is key, and it is important for EPA to understand where we do the research and how we can advance the science to more confidently protect aquatic life. She thanked the participants for taking time out of their busy schedules to attend the workshop and share their expertise on duration and frequency and their application to aquatic life criteria.

Presentation IX: Frequency: History and Application (Discussion Leader: Chris Mebane, USGS)

The presenter noted that the discussion of the frequency component of the criteria has historically been quite different than the discussion of duration. The rationale behind the frequency component of criteria is based on the recovery of aquatic life after an exceedance event. He proposed the participants consider some fictional, randomly generated data for four scenarios. In each scenario, there was one exceedance in a 3-year period. In scenario 1, the mean in-stream concentration was below the criterion magnitude and there is one slight exceedance. In scenario 2, there was only one slight exceedance, but the mean in-stream concentration was close to the criterion magnitude. In scenario 3, the mean in-stream concentration was well below the criterion magnitude, but there was a major exceedance. In scenario 4, the mean hovered right below the criterion magnitude and there was a major exceedance. Scenario 4 presented the most serious scenario.

The presenter described how a biological population or community cannot be affected until an individual is affected. Recovery starts with individual organisms. If an organism has a near death exposure, the organism will either make a full recovery or suffer from reduced fitness. The presenter provided an overview of several published papers that discussed recovery (Ashauer, 2015; Landrum, 2004; and Zahner, 2009). In reviewing the available literature related to recovery of organisms, there were several high-quality studies out of EPA's Duluth laboratory. These studies, conducted in the 1970s, examined the exposure of brook trout for three generations to lead, mercury, zinc, cadmium, and copper. At the end of each of these studies, the fish were put back in clean water and the effects were documented. Fish exposed to zinc recovered fully, but different results were observed among fish exposed to other contaminants. This leads to questions: "Is bioaccumulation a factor? What about residual contamination?"

The presenter discussed an unpublished example on exposure of bluegills to selenium for 1 year (Hermanutz et al., 1996). Following exposure, the reproductive effects were studied for 1 year. When exposed to treatments of 2.5 µg/L, 14 percent of the progeny were deformed and hemorrhaging. At the highest exposure concentration, 96 percent of the progeny were deformed with edema. After 1 year, whole body and ovary concentrations of selenium dropped, and the reproductive effects went to zero.

Individual fish recovered within 1 year. This means that the bluegill lost one reproductive year. These fish have a lifespan of 4 years with two or three reproductive years.

The 1985 Guidelines used the one-in-3-year exceedance frequency to be protective; however, concerns for long-living fishes remain. A walleye in Canada may live 10 years, but their lifespan is dependent on where they live geographically. The bigmouth buffalo fish can live to be about 100 years old, and Greenland sharks can live to be 500 years old. The presenter wondered if there should be different exceedance frequencies in waters with long-living fish species. Selenium significantly reduces the lifespan of the bluegill, which presented the question, "Should the exceedance frequency for bioaccumulative pollutants with criteria magnitudes expressed in terms of tissue residue be longer than one for water-based criteria concentrations like ammonia?"

He noted that accidents happen, and they are not just a result of variation in wastewater treatment plants (WWTPs). Upsets can kill a lot of fish. Examples included the warehouse fire where thousands of barrels of bourbon were lost in the river, which resulted in a large fish kill, and an upset at a Louisiana pulp and paper complex, which resulted in the release of toxic effluent to a river. The magnitude of the disturbance matters in defining effects and recovery rates. There can be an alternative stable state if the disturbance is not too great.

In the 1985 Guidelines, there is an assumption that if waters are not subject to other stresses, and if exceedances as large as a factor of two are rare, there can be a one-in-3-year exceedance and aquatic communities will remain adequately protected. This does not take spills into consideration, as spills are not part of normal operation of an effluent discharger. The one-in-3-year exceedance from the 1985 Guidelines clearly addresses that "spills and similar major events are not what is meant by an exceedance." Most U.S. waters are subject to anthropogenic stress. Many U.S. waterbodies receive nonpoint sources of pollution like agriculture and urban and suburban runoff, and compliance records are not always great. In Idaho, there were more than 100 violations in a 3-year period. It's not only the pristine water bodies that experience minor exceedances.

The presenter discussed the science of recovery, which depends on the physical, chemical, and biological characteristics; the severity and duration of exposure; and the proximity to an undisturbed area. An important consideration is colonization. One very informative experiment was done in the 1970s in six mangrove islands in Everglades National Park (Simberloff and Wilson, 1969). These mangroves were tented, fumigated, and then observed to see how the area was recolonized. Although the focus of this experiment was on insect recolonization, the concepts hold for other species. Early colonizers take hold and make it more difficult for other species to return. They found that the islands that were furthest away from another undisturbed island took the longest to recolonize. In 1989, several EPA water programs held a major workshop pertaining to case studies of recovery of aquatic ecosystems. The workshop produced at least 18 articles, some of them related to chemicals, geomorphology, and other topics. The studies found that in a short-term, nonpersistent disturbance, about 85 percent of fish recover in less than 2 years (Niemi et al., 1990). Three years is generally considered enough time for most ecosystems to recover.

The presenter discussed a literature review of the ecological traits of species in 3,900 papers including where they live (Kattwinkel et al., 2015). For invertebrates, five generations was a good estimate for recovery time. The available data supported the idea that a species with a lifespan of 7 days has a shorter recovery time than a species with a lifespan of 3 years. Another recent literature search looked at recovery in the context of the ecosystems around the disturbance (Stanford et al., 2018). It was more difficult for a previously undisturbed ecosystem to recover than a historically degraded ecosystem where there may not be much living there to begin with. He acknowledged A. Gergs's recent paper review (Gergs et al., 2016b). Although the context of the paper was focused on pesticides, it covered a lot of applicable topics.

The presenter considered the criticisms of the past work looking at catastrophic disturbances like massive forest fires and drought that killed nearly everything. The kind of exceedances one would observe associated with a water quality criteria exceedance are anticipated to be less severe, so a faster recovery would be anticipated. One paper out of the 1989 workshop (as cited in Gergs et al., 2016b) discussed a classification with a level one effect, which is a large disturbance on the landscape that kills everything with the colonization sources lost. Level 2 describes reduced severity but is still a very large disturbance (e.g., train in Northern California loaded with fungicide went into a river and affected 100 miles of that river). Any continuing release would also qualify as a level 2. Level 3 is less severe and may leave behind some sources of recolonization (e.g., the Kentucky bourbon spill). Level 4 is the least severe (e.g., a deliberate pesticide application to river to control black flies).

The presenter noted that recovery timing is an inexact measure. He reviewed a study that looked at recovery of damage from an Idaho mine to a set of streams over 25 years (Mebane et al., 2015). In 1995, the sources of contamination were reduced, and the copper contamination came down very quickly. He wondered where the timeline should begin for recovery. As the copper concentration decreased, more species appeared in the biomonitoring data. In 2008, the water quality criteria for copper were largely met. The question remains, "Should the clock have started in 1995 when the initial reduction began, or in 2008 when the copper concentration returned to acceptable amounts?"

In preparation for this workshop, the presenter shared that he reviewed about 70 studies of level 3 and level 4 non catastrophic events. In about 67 percent of these events, recovery took place within 1 year. In 5 years, recovery was observed for 94 percent of events. This observation supported the one-in-3-year exceedance criteria. The workshop in 1989 concluded that a 5-year recovery interval was too long, but the available studies suggested that 3 years is robust.

Expert Discussion of Presentation IX: Frequency: History and Application

The group wondered if recovery is generally just looking at populations or tissue concentrations. The presenter replied that it varied depending on what he could find. With a careful choice of endpoints, a faster or shorter recovery could be observed. Some studies report as time of first occurrence and others look at a population census. Most studies report multiple measures.

A participant asked for clarification of the 67 percent recovery. C. Mebane suggested reviewing the background report that gives a cumulative distribution figure of invertebrates and fish. Aquatic invertebrates recovered faster than fish, but invertebrates can be a zooplankton or a mussel. Only one case study was specific to mussels. In general, fish live longer than most invertebrates, so the invertebrates with shorter lifespans and more rapid reproductive cycles came back more quickly. It is possible the 67 percent came from mostly invertebrates, but there were a lot of fish in the studies.

Presentation X: Frequency Application in NPDES Permits (Discussion Leader: Danielle Stephan, EOA/OW/OWM)

The presenter discussed frequency application in permits and noted that frequency comes in when identifying the critical receiving water low-flow conditions that go into a reasonable potential analysis. The program identifies applicable WQS, determines where the discharge occurs and what water quality criteria exist that are applicable to that receiving water, and identifies any implementation policy that is associated with applicable standards of the criteria.

States that have developed criteria should have implementation policies and procedures in place to inform permit writers about the critical low-flow conditions. The next step is to characterize the receiving water in effluent. This is done by identifying the pollutants of concern and evaluating each one for reasonable potential. Next, mixing and dilution can be factored in. Generally, a steady-state model is used.

The 1Q10 and 7Q10 flows are typically used for acute and chronic criteria, respectively. The 1Q10 flow is a lowest 1-day flow in a 10-year period and the 7Q10 is lowest 7-day flow in a 10-year period. The 1Q10 and 7Q10 flows are hydraulically based. The permit writer does not make a decision on which low flow they will use, but rather the states make this decision. Generally, the 1Q10 and 7Q10 are used for toxics and a 30Q2 (i.e., lowest 30-day average flow that occurs, on average, once every 2 years) can be used for nutrients because longer averaging periods are considered for nutrients. As long as the state can say that a value is protective of their standard, it can be used. Modeling needs to be done for tidal flows to use these models in bays. Dilution and mixing are factored into the model, but this is just one piece of the larger picture. A participant added that it would be simple to report 1Q10 on a website as it is a rolling calculation. Calculations would be easier if the statistic was on a website.

The presenter stated that the 1991 TSD does not provide much information on the background but does present the conclusions of what was determined in the 1990s to be protective at a low frequency of exceedance. It is important to understand that the low flows established in implementation procedures are intended to be protective of the frequency component; this is the only place that frequency is factored into a permit. If a criterion does not include a frequency, one interpretation could be "never to be exceeded." From an enforcement perspective, a *never exceed* value might affect the statistics used. For nutrients, no dilution or mixing is allowed.

Expert Discussion of Presentation X: Frequency Application in NPDES Permits

The group asked if frequency is considered in reasonable potential analysis when effluent variability is addressed. A participant replied that it depends on whether data are available or not. A single data point can be projected on a lognormal scale. The use of more data is recommended to provide a better indication of what the facility is actually discharging over time. The likely distribution of the effluent will be shown (i.e., What are you discharging? Where are you going to exceed the criteria on a projected basis?). Another participant recalled that most permits have discharge information, and a 7Q10 or 1Q10 is used to take into account the frequency and flow. He wondered if maximum design flow is used for NPDESs discharge. The presenter replied it is based on design flow but could not recall if it was maximum or average.

A participant stated that it sounded like criteria are implemented on downstream ecosystems that have not experienced an exceedance for a period longer than a 3-year recurrence, at a 1Q10 or 7Q10 flow interval, at the maximum concentration, or that was otherwise not influenced by any limiting factors. The presenter agreed with this summary. The maximum concentration at low flow is being projected in permit derivation. Important questions to consider when evaluating if criteria will be met include considering the relationship with background concentrations, the possible effect of dilution and mixing, and maximum discharge concentrations.

A participant clarified that this analysis is not focused on setting permit limits, but rather on deciding whether the effluent should receive limitations for this parameter. The worst-case scenario is to determine if there is reasonable potential for this effluent to ever exceed water quality criteria. If the answer is yes, there is another set of similar calculations that gets back into where the limit should be set. The presenter agreed and added that a conservative approach is then employed. Calculations are needed to ensure there are no exceedances.

Presentation XI: Frequency Application in Assessment, Listing, and TMDLs (Discussion Leader: E. Monschein, EPA/OW/OWOW)

The presenter reviewed how frequency components of criteria are used in the assessment, impairment listing, and TMDL programs. The presenter started by indicating there are implications of limited datasets and making inferences about what those limited data mean for the acute and chronic duration. This also applies in terms of frequency, with quarterly sampling being most common. States are interested in having robust datasets, so in addition to their own sampling, they are looking for partnerships and data from other stakeholders to increase the sample sizes available for an assessment. In practice, the one-in-3-year returned frequency can be a disincentive. Once there is one excursion, there is a disincentive to find the second excursion that would lead to an exceedance over a 3-year period. If there is a frequency of zero and a 90-day average, and that exceedance is not allowed to occur more than once, there would be some incentive to go back the second year to collect more data to try and show a different condition. The issue is that, even though monitoring frequencies are low, if there is one exceedance then there are increased odds of having a long-term problem and possibly an exceedance if one grab sample is above the limit. The goal is to increase the size of the sample set to determine if this was a rare or isolated event.

The presenter noted that in their standards many states include a phrase dictating an exceedance once in 3 years “on average.” That statement is confounding to assessors because it is unclear what “on average” means. There is value in determining if this statement should be included in the expression of frequency. A participant replied that the way the “once in 3 years on average” is phrased, it is not always clear what it takes to get the average. It sounds like 9 years of data are required. Another participant replied that, unfortunately, this is the way it was conceived when looking for a long-term average. The focus was on permitting where one needed calculations based on probabilistic considerations. It was intended that if there was an exceedance in one 3-year period there could even be two exceedances if there had been no exceedances for several years, but this is not known from a monitoring standpoint. It creates a problem from a compliance standpoint if there is limited sampling. The presenter stated that, in practice, users disregard the “on average” expression.

For TMDLs, simple assumptions are common (e.g., set to never to exceed). Complex assumptions are sometimes used. The conservative nature of the assumptions was not investigated further or whether they should be set at "never to exceed." The presenter noted having all three components are important to an implementation program. These are not always included in state-approved criteria, and it is important that criteria are implemented consistently in every program. This practice is extremely valuable and reduces legal vulnerability when the duration and frequency are clearly expressed in the state standard.

Presentation XII: Disturbance and Recovery Group: Part I (Discussion Leaders: Dave Mount, EPA/ORD [session presenter 1], and Andre Gergs, Bayer AG, Environmental Sector Department [session presenter 2])

The first presenter started his discussion by presenting the acute species sensitivity distribution from the ammonia criteria document. The Genus Mean Acute Values were on the Y-axis and the portion values were on the X-axis. The values presented were acute values, but this would be essentially the same for chronic values because the assumption is that the chronic sensitivity distribution is the same except that it is pushed down in concentration by a standard value for each genus. The green line indicated the general assumption that when at half the LC_{50} there is a low level of mortality (i.e., limited acute effect). The bottom solid line was the CMC and it crossed the green line at about 5 percent. The red lines indicated where the exposure was. He assumed that the 1-hour average causes the full acute duration mortality. Thus, if at the 5-fold value, then 50 percent mortality is projected for about one-third of the organisms. If at the 10-fold value, that equates to 50 percent mortality for about three-fourths of organisms.

This presenter also discussed the copper acute species sensitivity distribution showing similar principles as the ammonia example. He presented another curve for cadmium, which looked quite different. The four most-sensitive genera were considered sensitive at 2X to 10X the criterion magnitude. He noted that the only thing this species sensitivity distribution curve dictated in the derivation is the percentile, but that information and those differences are not used in any other way. However, he believed that this suggested that the consequence of a given fold exceedance is very toxicant dependent. He emphasized that the consequence of the 2X assumed exceedance was consistent, but above that the consequences vary.

A participant commented that he agreed with all of this information for the acute values, but he does not quite agree that the chronic sensitivity is simply pushed down. His observation is that SSDs are much more compressed along the axis. Therefore, the change to 10X has a more dramatic effect on the percent of species mortality. Another participant responded that he does not think that is consistent with pesticides. The group engaged in a discussion of these issues. Following that discussion, this participant concluded that there could be multiple correct conclusions. A participant stated that there is a smaller range in concentration as one goes from very sensitive to very resistant. There remains variation and the nature of the variation depends on the chemical.

A participant asked if the acute to chronic ratio (ACR) gives some indication of what the difference is among chemicals. Another participant clarified that, when applying the ACR, the whole distribution is brought down by the factor of the ACR. The shape of the curve is irrelevant to the current derivation process. This participant emphasized the importance of considering the varying effects of different chemicals on species. Another participant added that the Europeans take this into account by fitting the distribution to the entire curve and suffering the consequences on the tails. Approaches used in both Europe and the United States have their own consequences. Another participant clarified that in Europe it is done in a way where included species are just those that are within one order of magnitude in terms of sensitivity. The presenter added that the impact from an exposure that is above the 2X criteria is not equal across chemicals. Participants agreed with this scientific statement.

The second presenter explained that he is in the process of working on, revising, and publishing the ecological recovery paper referenced in an earlier presentation. The presenter described the literature review he conducted in support for this workshop presentation. The review included studies published between 1990 and 2010. He reviewed these studies to determine if there was an improvement in the way ecological recovery studies were done as a result of the reviews in the 1990s. Five different aspects of recovery and pesticides were considered in addition to different ecosystems, taxonomic groups, and types of disturbances. Field and semi-field studies were compared. Another objective was to determine if magnitude had an effect on recovery time. Information on various recovery endpoints including abundance, biomass, diversity, and community composition were collected. Studies covered a wide variety of disturbances from pesticides to flood-related disturbances of metals. A total of 600 potentially useful case studies were identified from the abstracts. Only the studies that included a defined disturbance and reference data, and that defined what the recovery was related to, were further evaluated. Nearly 150 publications were considered in scope.

Presenter 2 discussed a figure summarizing the taxonomic groups covered in the literature review and stated that there are different ways to assess recovery based on the literature. Recovery measures were reported in terms of total abundance or biomass of macroinvertebrates. Other measures included taxa richness, diversity indices, and calculated community composition. A participant asked what sort of taxa richness and diversities indices were used. The presenter answered that taxa richness is often just referred to as number of taxa while the diversity indices also accounted for abundances. Another participant asked if there were redundancies in the figure, and if there was sequential evaluation of recovery for a few studies. The presenter responded that some redundancies did occur. Sometimes it was only possible to derive a single endpoint from a paper. Often the studies were not conducted for a long enough time period to show the full community recovery and they reported that they could not see recovery in some cases. Recovery was typically monitored for about for 3 years; however, many studies monitored recovery for less than 1 year while others monitored recovery for nearly 10 years.

A participant asked if the chart presented was limited as far as the taxa it pertained to—whether the chart encompassed a breadth of taxa or if it was restricted in any way. The presenter confirmed that it was not restricted in any way and confirmed that studies were identified from multiple continents, but most of the studies came from North America. Studies from Europe, Africa, and Australia were also analyzed.

The presenter asked the participants to consider how the magnitude of disturbance can affect recovery intervals. Most studies included high-effect levels, but in the few studies where the disturbance resulted in low to no effect, the recovery was relatively quick. The higher the effects, the more diverse the recovery times were. Therefore, the only thing that could be concluded was that the variability increased with the size of the effect. Stream macroinvertebrates were a strong example of this. A participant asked if the data used to draw these conclusions were collected in the field or the laboratory. The presenter replied that these data were from field studies and aquatic mesocosms. A participant stated that mesocosms imply internal recovery and wanted to clarify if that is what is occurring here. The presenter responded that it can also be external and the mesocosm does not explicitly imply internal recovery.

The presenter asked the participants to consider how recovery periods vary with different types of water bodies. The general conclusion from the presenter's review was that lotic systems recover faster than lentic systems due to the connectedness of lotic systems. He pointed out that stressor types in these systems were different and should also be considered. When focusing on pesticides, it was more about the use pattern, if you think of pesticides as being pulse disturbances or press disturbances. They were statistically somewhere in between.

The presenter asked the participants to also consider how much chemical properties affect recovery. Depending on chemical properties, a chemical stressor can quickly disappear from the water column, but it could persist in the sediment. In such cases the aquatic organisms showed a quick recovery; however, benthic organisms in the sediment may show long-term effects and delayed recovery. Environments that showed frequent disturbances also had more resilient and adaptive communities. He raised the hypothetical question: "Can these systems serve as a reference for pesticide exposure?"

Finally, the presenter asked the participants to consider how ecological characteristics affect recovery rates. Habitat connectivity and spatial distribution of undisturbed sites played a major role for colonization, recolonization, external recovery, and drift (in lotic systems). He also stated succession of recovery or recolonization might play a role along with the availability of food. Patterns in streams of early colonizers having success to short generation times and food availability were observed. Later in the sequence of succession, grazers and shredders also began to recover. The presenter noted that comparisons of recovery times throughout studies were difficult because most studies were not designed in a way that focused specifically on recovery. Other issues with comparing recovery times between studies also arose. For example, sometimes studies focused on different taxa and/or different trophic levels, so it was hard to find a common ground. Moreover, connectivity of habitats was not always clear. Sometimes species that were not the most abundant were disregarded. The completion of recovery was not always recorded accurately or clearly.

A participant asked the presenter if he examined any sort of pre-exposure and if that had any impact on recovery. The participant suggested a tolerance effect on pre-exposed individuals. The presenter stated that the history of exposure was not always clear, and these studies mostly focused on long-term exposure events. Another participant asked the presenter what his thoughts were on *Trichoptera* recovering in less than 1 year when their life cycles are 9 months to 1 year. The presenter explained this phenomenon with drift entries from streams.

A participant asked if it was possible to differentiate between accumulative and nonaccumulative compounds in terms of recovery times. A presenter responded by saying he found very few studies on that. Many case studies that focused on recovery were from longstanding sources, thus it was not clear when the disturbance ended and the recovery started. A presenter discussed bioaccumulative chemicals versus nonbioaccumulative chemicals. Juvenile daphnids can eliminate triphenyltin hydroxide (a chemical used in industrial settings that is not a pesticide), but adults cannot. A two-compartment bioaccumulation model was parameterized based on these data. Data were added for adults and juveniles, a full GUTS model was parameterized, and this was integrated into a population model to study population-level dynamics to determine what would happen under different assumptions and scenarios. If one assumed that bioaccumulation in adults was not possible, a very fast recovery was observed, which was not predicted by the model. If one assumed that bioaccumulation was possible in adults, the recovery of the population was delayed and more or less covered by the prediction of the model. He noted that it was also interesting that the population was already recovering before exposure, so the last two peak exposures did not have an effect in this experiment. This was also due to the nutrients available that compensated for the simultaneous effect. Twelve different population dynamic examples were tested to see if the peaks were toxicologically independent of population dynamics. The frequency of the peaks did not appear to have an effect on recovery time after exposure.

The presenter asked the participants to consider what kind of population models exist for different species with different life histories. If the model is tested on different species, it can be applicable as long as the model parameters are changed to accommodate the life histories of the target species. He explained that this model has been used to predict daphnids, snails, polychaetas, and isopods, among others.

The presenter discussed development of an individual-based community model to consider mayflies, some snail species, and mussels. He focused on a mayfly species in his presentation and provided a model of population dynamics in relation to pesticide exposure. The effects on dynamics of different mayfly species varied. He emphasized that this may illustrate that things were more complicated than the model showed. The addition of trout as a predator completely changed the dynamic of the model. A participant asked how this worked within the mayfly model and what influenced that dynamic. The presenter replied that he had every single individual in the model represented by the Dynamic Energy Budget Model (Murphy et al., 2018); then, he linked every species with a trait database that informed where the species lived and what they ate. Species also move around in their environment and feed and occupy their niche according to that linked trait database. Switches to other food sources can be made based on availability.

A participant asked what caused the better performance of the mayfly after the third pulse in the predation situation. The presenter believed it was due to reduced competition. Another participant clarified that this was relative to a control population. The control population would be lower in the predation scenario and there is less competition.

The presenter asked the participants to consider how ecological characteristics effect recovery rate. This was from the study by a colleague that focused on *Daphnia* spp. (Gergs et al., 2016b). Different scenarios with different food sources were considered and a pulse exposure effect was mimicked. When a competitor was added, the model was more complicated. A participant noted that this example illustrated how it is not easy to infer what would happen in different ecological scenarios. The presenter noted that all the models he presented are published, but one model is owned by his former employer, and he cannot use it anymore. The models discussed are individual-based models he developed.

Comments from Observers – Disturbance and Recovery Group: Part I

An observer asked about case studies relating to ecological characteristics and their effects on recovery rates. He wondered if Presenter 2 identified papers where predator-prey interactions or keystone species were studied, noting care is needed when considering the recovery of keystone species given their role in the ecosystem. Presenter 2 responded that he did find studies relating to phenomenon like trophic cascades, but he did not find any studies that related to a keystone species; however, a few studies did follow indicator species.

Presentation XIII: Disturbance and Recovery Group: Part II (Discussion Leader: Chris Mebane, USGS)

This discussion focused on the question: Is population modeling informative? In 2004, a paper published on pesticides (Landrum et al.) looked at different recovery scenarios for different organisms and how that could be used in decision-making. Using a modeling approach, different organisms with different life histories were considered. The same stress was then applied to all the organisms to see how they would recover and to identify the mortality scenarios. The baseline population lifespan was about 12 years. The scenario that was used fell into the severe category, where half of all life stages were killed.

Considering population modeling, the presenter noted an example where certain fish species might have life histories that would allow for an acute effect every other year, or even every year, without having an effect on the overall population. In this example, however, the concern remains that the effect could be severe enough to cause a local extinction event. Density dependence is often considered in these models, which can create confusion when thinking about an effect. Density dependent models consider how the remaining members of a population have more resources following a disturbance event. Therefore, removing 20 percent of the juveniles may not translate to a 20 percent reduction in the adult population. Moreover, population modelers should consider the carrying capacity and how strong of a density compensation should be used in the model. The density dependence relationship can have a large influence on the outcome of a population model run. For example, a density dependence relationship can be strategically selected, especially when the relationship is uncertain, such that population modelers or risk assessors can develop whatever extinction rate they wish.

Another example discussed was from the University of Manitoba. In this study, researchers observed that when the lakes froze, most of the fish would die; however, subsequent recovery was also observed. Because the fathead minnow can reproduce very quickly the extinction risk is not any different from the baseline. These concepts are something fisheries managers consider when developing fishing limits. Different species can provide varying results. The critical limiting factor for salmon is whether they can get out to the ocean and back. Shortnose sturgeon in the Connecticut River live to 60 years and do not reproduce until they are 10 years old. When half the population was killed off, the risk of decline was 100 percent. Shortnose sturgeon are rare, so the assumption is they are density independent. Thus, no compensation for density was made to the example model. If one sturgeon is killed in an early life stage, it will cascade through to the population. A baseline chronic stress was added and resulted in a 5 percent loss of juveniles instead of a 50 percent loss. The impact of a 20 percent loss every 3 or 5 years was also considered in the example model. In this case, the risk of a loss or extinction event was not very different from the baseline. The presenter noted this was an interesting exercise that could be done forever with a variety of different species under different conditions.

Expert Discussion on Presentation XIII: Disturbance and Recovery Group: Part II

A participant asked about differences in complexity when looking at individual species and wondered if this could be achieved with metrics. The presenter replied that this scenario would have to be run for different organisms and those scenarios would then need to be put together. The population models he used are based off the life histories of the specific organisms. The models he used were constructed to not have lingering effects.

A participant noted that, when talking about risk of dropping below a certain abundance, these are often below the top curve. He asked what this observation means. The presenter replied that the sturgeon model began with a population of 400 females. The first risk curve is the risk of decline below the population abundance level at any time during the simulation, which, in this case, was 33 years. Over the course of a different simulation that was for 21 years, there was a 100 percent chance the abundance would drop below 1,500 (including the initial population). A population at carrying capacity has an R maximum of one. However, the R maximum was around five for the fathead minnow.

A participant shared that, except for the low effect on the sturgeon, the message seemed to be that in some cases events every 3 years could be tolerated without a great decline in the population. The presenter agreed, but cautioned that if density dependence is not included, there might be different results.

A participant described a real-world example where the population of oysters in the Chesapeake Bay was reduced because of a pollutant. This led to the collapse of the fishery followed by a rise in competing species. Disease was likely always present but had not been previously noticed. Population modeling could work for certain kinds of organisms. The presenter noted this modeling is for ecological risk assessment to determine how “alternative A” compares to “alternative B.” There is skepticism because a level of sophistication to the model is needed to account for competition. At high densities, the population suffers evenly, but with competition the rich stay rich and the poor stay poor because resources are not distributed equally. Fish are territorial. There is a difference between contest and scramble models and changing that assumption changes the outcome.

A participant noted that density dependence is affected by the carbon input in the model. Not all fish can grow with limited energy sources. A smaller population would have more resources available. It confines the system with a fixed behavior. A river may not have the same constraints. Another participant agreed with this statement and noted the need to get the scenario right. It is very easy to manipulate the model, and detailed information on the driving factors is needed. One should consider how this could be transferable to other species.

A participant said that every model needs to define what is an unacceptable change or effect. This has not been defined. Quantitative tools do not help if you cannot identify the unacceptable threshold in quantitative terms. One should understand where the threshold is between acceptable and unacceptable. If that is not understood, arguing about the time frame is not productive. It does not matter what tool is used if the units to classify measurements above or below the threshold are not available.

A participant noted the one digression in 3 years is not saying that the recovery period is 3 years. It is premised on the recovery being 1 year or less, with the second and third year being an unperturbed situation to allow for recovery. Another participant acknowledged that there are no single water quality criteria that can be protective of everything. A weight of evidence approach is used. Toxicity testing is a great tool, but the community is also considered. Impairment at the community level is important to determine what is acceptable and what is not. He wondered how that tool could be applied elsewhere and how the science could be brought together.

A participant summarized that there were two levels of issues on the table. The first issue is what we know about frequency and duration. Questions to consider are: “For frequency, is 3 years a reasonable timeframe?” and “What is the time scale of recovery?” This participant stated the answers to these questions were “it depends,” with a wide variation rooted in all the factors. Three years is a reasonable place in that range, but perhaps on the longer side. He wondered if there is something compelling to argue that longer or shorter is supported by the information we have and if recovery is understood well enough. Because the criteria are national criteria, they are thought of as more chemical-specific rather than system-specific.

A participant pointed out that mussel regeneration occurs in 5 years, so 3-year criteria are much shorter. If this is considered along with endangered species, the conversation may be different. Another participant replied that the participants are not proposing it should be zero or that mussels are sensitive to all toxicants. The expectation of national criteria related to endangered species is worth mentioning in this context. Another participant shared that his experience with endangered species is different and uses a whole different equation.

A participant reminded the group that “most species, most of the time” is the typical expression. Exactly what that means is unclear, but it is not the same as “all individuals, all of the time.” It circles back to the point where if a clear target for the analysis is not defined, the analysis becomes challenging.

The group noted that previous evaluations of frequency included similar opinions. When there are so many variables to consider, it is relative to which input parameters are used. These models give a sense of what more specific considerations need to be factored in. National criteria use a broad brush. The general understanding of the sensitivities is one aspect of the considerations for specific uses.

A participant pointed out that assuming mussels are not threatened or endangered, A. Gergs’s slides indicated mollusks had a relatively slow recovery time. He asked if it is logical to have a longer recovery time knowing there is a specific species that takes longer to recover. A. Gergs replied it might make sense to look at the most sensitive species and look at how long it takes them to recover. The presenter wondered if mussels are very sensitive to ammonia and if a 5-year recovery should be used. He was also unsure about frequency components for tissue-based criteria. A participant added that the mussels may be a specialized case. The questions are if freshwater fish are sensitive and what makes them sensitive. Answers to that question could be because the fish are long-lived or large. Mussels are considered to be long-lived, but this is not clear for other related species. For example, what if snails recover quicker and mussel data are not available? There are limits on what can be concluded. Implementation can be a challenge (e.g., How much effect is enough? What is an outlier?). For some circumstances, there might be enough evidence to support a sensitive species. Another participant noted mussels are not sensitive to all chemicals, but they are sensitive to many metals and other pollutants.

A participant said there is a push and pull when tailoring analyses. The 1985 Guidelines are explicit enough. To a large extent, deriving a criterion does not suffer from user bias to the same extent. Less time is spent arguing about process. It is not just transparency; it is consistency and the relative lack of ambiguity. Guidance should not become more subjective.

A participant asked if it makes a great deal of difference if one says the frequency of exceeding is two or five or seven. He wondered if that actually changes the criteria values. The reality comes in on the permitting side because the designs of the WWTPs are different and the treatment methodology needs to meet the criteria value. Another participant clarified that if a state regulator is concerned about life expectancy of mussels and they represent a higher risk, and the state imposed a factor of two and comes up with a slightly lower number, EPA would be okay with this approach because it is lower than the regulation.

A participant was concerned about regulation of water when some processes are not taking place in the water. There have been proposals to create integrated criteria in the past. Sediment or other parts of the system other than the water column may have a level of contamination that is causing an effect. Another participant pointed out that if growth is reduced, there is more food available for the rest of the population. If the organism is not killed but growth is reduced, they feed less, and a higher population can be supported. Finally, a participant shared that S. Kadlec's group is looking at growth effects over a whole lifecycle with discontinuous exposure. For some species during established lethal exposure, the growth slows or stops during the exposure. When it ends, growth picks up where it previously left off and the species continued to get as big as expected had there never been exposure. It was if there was just a time delay. The growth delay results in a lower final adult body size for the other species. In some cases, reproduction is related to body size.

Comments from Observers – Disturbance and Recovery Group: Part II

An observer asked if there was concern that a model could be built to generate the desired outcome, creating a sort of bias. The presenter replied that this might not happen on purpose, but there is bias to pick the parameters to fit the perception of the problem. A participant agreed and added that it is important to choose the data so that one can understand if the assumptions in the model are reasonable. The presenter added that there is value in training and testing the datasets. A participant agreed with this statement so long as there is confidence that the model is reasonable for use with a given scenario. The participant commented again, clarifying that the magnitude of the effects was 60 to 80 percent mortality for the models.

An observer asked if the one-in-3-year approach holds if looking at nonlethal effects. The presenter replied that, for population biology, only survival and reproduction can be considered—not growth. A decline in reproduction can be plugged into the model. Density dependence is also a factor.

Presentation XIV: Frequency: Approaches Group (Discussion Leader: Chris Mebane, USGS)

The presenter shared that the purpose of this session was to discuss episodic water pollution and implications for pollutant exceedance frequencies for aquatic life criteria. He stated that, if a stream has a goal to protect 95 percent of taxa 95 percent of the time, and if 20 representative samples are taken over a 3-year period and none of these samples are exceedances, then the stream is in compliance. However, if a stream has a goal to protect 95 percent of taxa 95 percent of the time, and if only one representative sample is taken over a 3-year period, and if that sample is an exceedance, then the stream is out of compliance. Thus, the 95th percentile approach can incentivize additional monitoring.

The presenter provided an example of a data-rich system from the Blackfoot River, Idaho. The river has an autosampler that collected more samples during high flow periods than low flow periods. The purpose of this example was to illustrate that the use of different methods to calculate averages impacts the resulting data. Selenium data were evaluated using rolling averages over 30-day periods for 6 years. The data were split into two 3-year periods and criteria were expressed by generating the 95th percentile of 30-day maximum values. In this case, the 95 percent of the time approach gave equivalent protection as did the 30-day rolling averages.

The presenter then discussed a separate example of a zinc-contaminated river in northern Idaho where a more typical sampling effort was employed. This river acts as an example of a press disturbance with contamination from groundwater. He regressed the zinc concentration data over paired stream flow measurements and found that the lowest flows had higher concentrations of zinc and the highest flows had lower concentrations of zinc. There were approximately six samples per year which were event-driven; the lowest and highest flows were intentionally sampled. Because the samples were not taken quarterly, no two samples were collected within a 4-day period. Thus, a 30-day and 4-day maximum average could be the same since both time periods might have only one sample. In this example, the 95th percentile of the measured values was considerably lower; if criteria were expressed this way, the samples that exceed the 95th percentile line get a pass. If a lot of data are available, the 95th percentile would be lower and less protective than this example where one sample is collected in a 4-day period.

The presenter then reviewed an example of two streams in New York. Sampling in these streams produced a time-dense sample set for total aluminum. The dataset included a 3-year period in which many samples were collected within 30 days of each other. Most of the time, total aluminum was below the chronic criterion value, but there are a few exceedances. He explained that he rescaled the exceedances so that exactly 95 percent were in compliance. The reason for rescaling the data was to use realistic variability from a real stream to look at what the potential consequences would be of exceedances 5 percent of the time.

Using a 95th percentile of sampled data to determine compliance is not as strict as using 4-day maximum averages. He noted that one should consider what this would mean for protecting communities. As a next step in his analysis, he scaled the total aluminum dataset so that 95 percent of data were less than the criteria factor and 5 percent of data were exceedances. He then questioned what fraction of an aquatic community would be adversely affected by the exceedances. He used SSD to see if just a few taxa could be harmed or many. For this exercise, he assumed the SSD from the criteria document represented a community and generated a regression.

By definition, an exceedance factor of one can potentially affect up to 5 percent of the community. Four exceedance events could affect up to 20 percent of the SSD twice in this 3-year period. The presenter clarified that the SSD data are points in time, not averages. If compliance were interpreted as an allowable percentage, there is no need for an averaging period. He clarified that these are chronic data in the SSD, because most data points for the stream chemistry are between 4 to 20 days apart. If the SSD is reflective of the real distribution, the SSD can be used to analyze exceedance effect. Finally, he clarified that this assumed that the SSD is relevant to real communities.

The other idea is to examine a locally important stream resident. In this example, the presenter found a study of 20-day exposures of aluminum to brook trout with a mortality function. He expressed brook trout exposure concentrations as exceedance factors and generated a curve that can be used to determine how many events would cause mortality to brook trout. In this example, one event would cause mortality. He clarified that he used scaled-down data, and that if he had used original concentrations, there would be a higher effect and the stream would not be in compliance. This methodology would make it easier to conduct monitoring and determine compliance.

Expert Discussion of Presentation XIV: Frequency: Approaches Group

A participant asked if this method would encourage more monitoring, because more sampling would increase representativeness of the stream. The presenter confirmed that this method would encourage monitoring. Another participant expressed support for this method and that the method used the exceedance frequency and toxicity models to analyze the exceedance effect. An audience member wondered if this method could consider site variability. The audience member also wondered how using averaging frequency verses exceedance frequency would encourage or discourage monitoring. The participants discussed that, because data are considered as the percentage of time criteria are exceeded, one high data point is likely to significantly change the average. This observation would hypothetically incentivize sampling if there was prior knowledge of the variability to foresee the need for additional sampling. The participants noted that averaging encourages monitoring and that setting a low average criterion might disincentivize monitoring if monitoring is done once and is below criteria.

The participants discussed implementation guidance. Implementation guidance would express criteria as a value that shall not be exceeded a certain percent of time. If samples are below the criteria, monitoring could stop unless a requirement prohibits this approach. A participant added that states discourage averaging site data in order to allow impairments to be more specifically located.

A participant shared that Virginia is developing criteria for the James River and is specifying that monitoring occur monthly during the growing season. However, because the 2-year integrated report cycle cannot be avoided, it is unclear why high frequency monitoring should occur if no steps can be taken until after the diagnosis is made in 2 years. The statute does not allow for quick response. The only reason for collecting this high-frequency monitoring data would be to report it to the public for awareness. The presenter clarified that the purpose of monitoring is typically to aid decision-making, and 2-year reports are rarely helpful for managing water quickly. However, the 3-year return interval could be kept and analyzed every 3 years.

A participant asked if current criteria could be translated to the new format and structured to encourage monitoring and compliance. Another participant similarly wondered how this would be done and noted that low flow could be made a period of focus for the compliance calculation. The presenter noted that he has not considered the extent to which permit calculations and criteria are decoupled.

A participant noted that there are comparisons of 7Q10 values to one-in-3-year period values because daily discharge fluctuation is averaged in the longer term to avoid shorter-term peak events, which may be lower than the 4-day peak. Another participant pointed out that this represents an alternative approach that needs further development and consideration of the impact on implementation and compliance.

A participant expressed support for this approach and the potential to incentivize additional monitoring. He asked if using a modified toxicity test would be an option during extreme events or 1-hour exposures. He considered the use of laboratory-based pulse studies to develop criteria. For high-pulse episodes of nonpoint source stormwater, it is difficult to determine what is necessary to be environmentally protective. A participant noted that good exposure characterization for an event is critical. Another mentioned that high-resolution data inform calculations. Instead of solely analyzing peak exceedances, the entire profile can be analyzed for an entire SSD. Exposure-specific SSDs show the level of exceedance that can be tolerated for a certain scenario.

A participant requested clarification as to why 5 percent exceedance is a viable approach. He asked if a 5 percent rate of exceedance is similar to the one-in-3-year approach based on similarities outlined in the premeeting materials. The presenter clarified that 5 percent exceedance is not necessarily less protective. There are more 4-day periods in 3 years than in 5 percent of data. If there was no data limit, and the one exceedance level was set, it would be more restrictive in the 4-day approach than the 5 percent approach. Sufficient protection is critical to consider. He mentioned using a factor of 2.5 magnitude cap on the conventional framework or a factor of 2 on 5th percentile; there is a way to balance things. He suggested this approach as a viable way to express magnitude.

A participant requested clarification on the reasoning and variability behind factor choices and how it relates to protection of the aquatic community. The presenter clarified that two was used as a protective factor. An exceedance of 1.5 could run up to 20 percent of the factor. The community experienced quasi-acute exposure to chronic levels. There should be a cap. Impacts to specific species are easier to model.

The participants clarified that a small percentage is being considered. The First-order Acidity Balance model and exceedance factor both happen to be two. This would be impacted by the concentration slopes for acute and chronic response. To standardize the effect, the slope of the SSD and the slope of the chronic response should be analyzed per species. A value cannot exceed the CMC without exceeding the CCC. A factor of two covers lethality.

A participant noted that when there are severe stresses, the “greater than one occurrence in a 3-year period” approach is still applicable. The presenter said the approach that there are no more than 5 percent exceedances in a 3-year period could also be applicable. The advantages of the different approaches should be considered further.

A participant stated that, if 95 percent of data are analyzed over 3 years, a lot of data are needed because data cannot be extrapolated. The presenter clarified that not a lot of data are needed and only one or two samples would be needed if these samples are below the criteria value. However, a lot of data (e.g., collected through continuous monitoring) would make the distribution known and extrapolation would not be needed.

Comments from Observers – Frequency: Approaches Group

An observer stated that monitoring cannot typically be adjusted as data are submitted. Monitoring programs and schedules are developed months in advance of the sampling. Thus, disincentivizing monitoring is not a concern. States look at data both in aggregate and in granular form to make a diagnosis for which they may not have enough data. States typically prefer to have more data. A participant pointed out that incentives for monitoring vary by state. Monitoring type and frequency are dependent on the parameter. The presenter suggested that a limit on magnitude would be appropriate. For example, the 95th percentile approach generates a lower cap than other approaches because if you have enough data, the 95th percentile would allow more exceedances than the one-in-3-year approach. The presenter noted that the return frequency and the state's interpretations should be considered when setting limits.

Closing Remarks (Mary Reiley, EPA/OW)

M. Reiley thanked the experts for their participation. She shared that PDFs of the presentation slides would be sent to the participants. She asked the presenters to provide redacted versions of their presentations, or instructions for redaction, if needed. EPA's NPDES website can be visited for more information on implementation of criteria in permits. A printable version of the support document would also be sent.

References

Ashauer, R., I. O'Connor, A. Hintermeister, and B.I. Escher. 2015. Death Dilemma and Organism Recovery in Ecotoxicology. *Environmental Science & Technology* 49(16):10136–10146.

Bailey, H.C., D.H.W. Liu, and H.A. Javitz. 1985. Time/Toxicity Relationships in Short-Term Static, Dynamic, and Plug-Flow Bioassays. In *Aquatic Toxicology and Hazard Assessment: Eighth Symposium*. ASTM No. STP 891. American Society for Testing and Materials, Philadelphia. pp. 193–212.

Balistrieri, L.S., D.A. Nimick, and C.A. Mebane. 2012. Assessing time-integrated dissolved concentrations and predicting toxicity of metals during diel cycling in streams. *Science of The Total Environment* 425:155–168.

Brent, R.N. and E. Herricks. 1998. Postexposure effects of brief cadmium, zinc, and phenol exposures on freshwater organisms. *Environmental Toxicology and Chemistry* 17(10):2091–2099.

Erickson, R.J. 2007. Quantification of Toxic Effects for Water Concentration-Based Aquatic Life Criteria. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Laboratory, Duluth, MN.

Gergs, A., D. Kulkarni, and T.G. Preuss. 2015. Body size-dependent toxicokinetics and toxicodynamics could explain intra- and interspecies variability in sensitivity. *Environmental Pollution*. 206:449–455.

Gergs, A., F. Gabsi, A. Zenker, and T.G. Preuss. 2016a. Demographic Toxicokinetic-Toxicodynamic Modeling of Lethal Effects. Environmental Science & Technology 50(11):6017-6024.

Gergs, A., S. Classen, R. Strauss, R. Ottermanns, T.C.M. Brock, H.T. Ratte, U. Hommen, and T.G. Preuss. 2016b. Ecological Recovery Potential of Freshwater Organisms: Consequences for Environmental Risk Assessment of Chemicals. *Reviews of Environmental Contamination and Toxicology* 236:259–294.

Hermanutz, R.O., K.N. Allen, N.E. Detenbeck, and C.E. Stephan. 1996. Exposure of bluegill (*lepomis macrochirus*) to selenium in outdoor experimental streams. U.S. Environmental Protection Agency, Mid-Continent Ecology Division, Duluth, MN.

Ivey, C.D. and C.A. Mebane. 2019. Acute and latent effects of zinc on two commonly tested species (*Ceriodaphnia dubia* and *Oncorhynchus mykiss*). U.S. Geological Survey data release, https://doi.org/10.5066/P9X4HH4T.

Jager, T., C. Albert, T.G. Preuss, and R. Ashauer. 2011. General Unified Threshold model for Survival – a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science & Technology*. 45(7):2529–2540.

Kattwinkel, M., M. Liess, M. Arena, S. Bopp, F. Streissl, and J. Rombke. 2015. Recovery of aquatic and terrestrial populations in the context of European pesticide risk assessment. *Environmental Reviews* 23(4):382–394.

Landrum, P.F., J.A. Steevens, D.C. Gossiaux, M. McElroy, S. Robinson, L. Begnoche, S. Chernyak, and J. Hickey. 2004. Time-dependent lethal body residues for the toxicity of pentachlorobenzene to *Hyalella azteca*. *Environmental Toxicology and Chemistry* 23:1335–1343.

Mancini, J.L. 1983. A method for calculating effects, on aquatic organisms, of time varying concentrations. *Water Research* 17:1355–1362.

Mebane, C., R.J. Eakins, B.G. Fraser, and W.J. Adams. 2015. Recovery of a mining-damaged stream ecosystem. *Elementa Science of the Anthropocene* (3):000042.

Mebane, C.A. 2019a. A review of the nature and effects of episodic water pollution and implications for aquatic life criteria averaging periods. Workshop discussion initiation paper no. 1 from the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019. 54 pp.

Mebane, C.A. 2019b. The capacity of aquatic ecosystems to recover from exceedences of aquatic life criteria. Workshop discussion initiation paper no. 2 from the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019. 72 pp.

Murphy, C.A., R.M Nisbet, P. Antczak, N. Garcia-Reyero, A. Gergs, K. Lika, T. Mathews, E.B. Muller, D. Nacci, A. Peace, C.H. Remien, I.R. Schultz, L.M. Stevenson, and K.H. Watanabe. 2018. Incorporating Suborganismal Processes into Dynamic Energy Budget Models for Ecological Risk Assessment. *Integrated Environmental Assessment and Management* 14(5):615–624.

Niemi, G.J., P. DeVore, N. Detenbeck, D. Taylor, A. Lima, J. Pastor, J.D. Yount, and R.J. Naiman. 1990. Overview of Case Studies on Recovery of Aquatic Systems from Disturbance. *Environmental Management* 14(5):571–587.

Nimick, D.A., C.H. Gammons, T.E. Cleasby, J.P. Madison, D. Skaar, and C.M. Brick. 2003. Diel cycles in dissolved metal concentrations in streams: Occurrence and possible causes. *Water Resources Research* 39(9):1247.

Simberloff, D.S. and E.O. Wilson. 1969. Experimental Zoogeography of Islands: The Colonization of Empty Islands. *Ecology* 50(2):278–296.

Stanford, B., H. Jones, and E. Zavaleta. 2018. Meta-analysis of the effects of upstream land cover on stream recovery. *Conservation Biology* 33(2):351–360.

Tobiason, S., D.K. DeForest, K. Marx, J. Keithly, R.C. Santore, W.A. Stubblefield, and K.V. Brix. 2003. Zinc risks from stormwater runoff at an urban airport. Platform presentation at the Society of Environmental Toxicology and Chemistry 24th Annual Meeting, Austin, TX.

USEPA (U.S. Environmental Protection Agency). 1980. *Ambient Water Quality Criteria for Cadmium*. EPA 440/5-80-025. USEPA. https://www.epa.gov/wqc/ambient-water-quality-criteria-cadmium.

USEPA (U.S. Environmental Protection Agency). 1984. *Ambient Water Quality Criteria for Cadmium – 1984*. EPA 440/5-84-032. USEPA. https://www.epa.gov/wqc/ambient-water-quality-criteria-cadmium.

USEPA (U.S. Environmental Protection Agency). 1985. *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*. USEPA. https://www.epa.gov/sites/production/files/2016-02/documents/guidelines-water-quality-criteria.pdf.

USEPA (U.S. Environmental Protection Agency). 1991. *Technical Support Document For Water Quality-based Toxics Control*. EPA/505/2-90-001. USEPA. https://www3.epa.gov/npdes/pubs/owm0264.pdf.

USEPA (U.S. Environmental Protection Agency). 1999. *1999 Update of Ambient Water Quality Criteria for Ammonia*. EPA-822-R-99-014. USEPA.

USEPA (U.S. Environmental Protection Agency). 2001. *2001 Update of Ambient Water Quality Criteria for Cadmium*. EPA-822-R-01-001. USEPA.: https://www.epa.gov/wqc/ambient-water-quality-criteria-cadmium.

USEPA (U.S. Environmental Protection Agency). 2007. *Aquatic Life Ambient Freshwater Quality Criteria – Copper 2007 Revision*. EPA-822-R-07-001.USEPA. https://www.epa.gov/wqc/aquatic-life-ambient-freshwater-quality-criteria-copper-2007-revision.

USEPA (U.S. Environmental Protection Agency). 2013. *Aquatic Life Ambient Water Quality Criteria for Ammonia – Freshwater 2013*. EPA 822-R-18-002. https://www.epa.gov/sites/production/files/2015-08/documents/aquatic-life-ambient-water-quality-criteria-for-ammonia-freshwater-2013.pdf.

USEPA (U.S. Environmental Protection Agency). 2016. *Aquatic Life Ambient Water Quality Criteria Cadmium – 2016*. Available on the Internet at: https://www.epa.gov/wqc/aquatic-life-criteria-cadmium#2016. EPA-820-R-16-002.

USEPA (U.S. Environmental Protection Agency). 1988. Fathead Minnow Acute Toxicity database. USEPA. https://comptox.epa.gov/dashboard/chemical_lists/EPAFHM.

Zahner, H. 2009. Recovery of Fathead Minnows (*Pimephales promelas*) Following Episodic Copper Exposure: A Biochemical, Physiological, Individual, and Population Perspective. *All Dissertations*. 360. https://tigerprints.clemson.edu/all_dissertations/360/.

Zinsser, L.M., C.A. Mebane, G.C. Mladenka, L.R. Van Every, and M.L. Williams. 2018. *Spatial and temporal trends in selenium in the upper Blackfoot River watershed, southeastern Idaho, 2001-16*. U.S. Geological Survey Scientific Investigations Report (5081). https://doi.org/10.3133/sir20185081.

Appendix A. Workshop Participant List and Agenda

U.S. Environmental Protection Agency
Frequency and Duration Workshop
September 11–12, 2019 | Cadmus Group
3100 Clarendon Boulevard, Arlington, VA 22201

Participants List

Table 1: Invited Experts, Presenters, and Meeting Coordinators

Name	Affiliation	Meeting Role
Betsy Behl	EPA	Presenter
Kevin Brix	ECOTOX	Invited Expert
Molly Colvin	United States Navy	Invited Expert
Mike Elias	EPA	Coordinator
Russ Erickson	EPA	Invited Expert
Bryson Finch	Washington State Department of Ecology	Invited Expert
Andre Gergs	Bayer AG	Invited Expert
James Justice	EPA	Coordinator
Sarah Kadlec	EPA	Invited Expert
Jeff Manning	North Carolina Department of Environmental Quality	Invited Expert
Chris Mebane	USGS	Invited Expert
Eric Monschein	EPA	Presenter
Dave Mount	EPA	Invited Expert
Deborah Nagle	EPA	Presenter
Mary Reiley	EPA	Coordinator
Adam Ryan	International Zinc Association	Invited Expert
Travis Schmidt	United States Geological Survey	Invited Expert
Danielle Stephan	EPA	Coordinator
Chris Stransky	Wood Environment and Infrastructure Solutions, Inc.	Invited Expert
Bill Stubblefield	Oregon State University	Invited Expert
Ning Wang	United States Geological Survey	Invited Expert

Table 2: Observers

Kathryn Gallagher	EPA	Observer
Lareina Guenzel	EPA	Observer
John Healey	EPA	Observer
Jacques Oliver	EPA	Observer
Laura Phillips	EPA	Observer

Agenda

<u>September 11, 2019</u>

- **Arrival and Registration** (8:15 – 8:45)

- **Introduction**
    - Introductions, logistics, and agenda (M. Elias, J. Justice, and M. Reiley, EPA/OW, 8:45 – 9:00)
    - Division Director Welcome (B. Behl, EPA/OW/HECD, Director, 9:00 – 9:10)

- **Workshop discussion initiation presentations and initial questions** (C. Mebane, USGS, 9:10 – 10:10)
    - Issue of episodic water pollution and toxicity of fluctuating concentrations (9:10 – 9:30)
    - Capacity of aquatic ecosystems to recover from toxic episodes (9:30 – 9:50)
        - Questions and Discussion (9:50 – 10:10)

- **Break** (10:10 – 10:25)

- **Duration: History and Application** (10:25 – 11:45)
    - History of duration derivation and underlying assumptions (R. Erickson, EPA/ORD, 10:25 – 10:45)
    - Duration application in NPDES permits (D. Stephan, EPA/OW, 10:45 – 11:00)
- Duration application in assessment, listing, and TMDLs (E. Monschein, EPA/OW, 11:00 – 11:15)
    - Duration use and application (11:15 – 11:45)
        - Question and Answer

- **Lunch** (11:45 – 12:45)

- **Exposure Duration and Effects Group Part I** (led by A. Ryan, International Zinc Association, 12:45 – 1:45)
    - Observed impacts of exposure duration and fluctuations on toxic effects

- **Exposure Duration and Effects Group Part II** (led by K. Brix, ECOTOX, 1:45 – 2:45)
    - Modeling impacts of exposure duration and fluctuations on toxic effects

- **Break** (2:45 – 3:00)

- **Exposure Duration and Effects Group Part III** (led by R. Erickson, EPA/ORD, 3:00 – 4:00)
    - Application of toxicity models to toxic effects from time-varying exposures

- **Exposure Duration Summary** (led by M. Elias, J. Justice, and M. Reiley, EPA/OW, 4:00 – 4:45)
    - Discuss items that were not resolved

<u>September 12, 2019</u>

- **Recap and Agenda** (M. Elias, J. Justice, and M. Reiley, EPA/OW, 8:45 – 9:15)

- **Frequency: History and Application** (9:15 – 10:30)
  - History of frequency derivation and underlying assumptions (C. Mebane, USGS, 9:15 – 9:35)
  - Frequency application in NPDES permits (D. Stephan, EPA/OW, 9:35 – 9:50)
  - Frequency application in assessment, listing, and TMDLs (E. Monschein, EPA/OW, 9:50 – 10:05)

- **Office Director Remarks** (D. Nagle, EPA/OW/OST, Director, 10:05 – 10:15)

- **Frequency Use and Application** (10:15 – 10:30)
  - Question and Answer

- **Break** (10:30 - 10:45)

- **Disturbance and Recovery Group Part I** (led by D. Mount, EPA/ORD, 10:45 – 11:45)
  - Characteristics of ecological disturbance and recovery

- **Lunch** (11:45 – 12:45)

- **Disturbance and Recovery Group Part II** (led by D. Mount, EPA/ORD, 12:45 – 1:45)
  - Implications of chemical exposure/effect characteristics for ecological recovery

- **Frequency Approaches Group Part I** (led by C. Mebane, USGS, 1:45 – 2:45)
  - Population, time distribution, and other parameters to address frequency

- **Break** (2:45 – 3:00)

- **Workshop Wrap-up and Summary** (led by M. Elias, J. Justice, and M. Reiley, EPA/OW, 3:00 – 3:55)
  - Discuss items that were not resolved

- **Closing Remarks** (3:55 – 4:00)

Appendix B. A Review of the Nature and Effects of Episodic Water Pollution and Implications for Aquatic Life Criteria Averaging Periods

A review of the nature and effects of episodic water pollution and implications for aquatic life criteria averaging periods

Workshop discussion initiation paper no. 1 for the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019*

Contents

Summary 1
Introduction 3
The episodic nature of contaminants exposure in freshwaters 3
The episodic nature of water quality parameters that affect toxicity/bioavailability 10
Averaging periods for criteria – background and controversies 12
Chronic criteria averaging periods 13
Acute criteria averaging periods 14
What difference do acute averaging periods make? 15
Toxicity resulting from brief chemical exposures 19
Approaches for evaluating averaging periods for criteria 23
Predictive modeling approaches 23
Speed of action calculations 24
The delayed mortality problem 32
References 38
Appendix 1: A compilation of time-to-effects in pulse or short-term toxicity testing 47

Summary

1. Water pollution is often time-varying in mid-size rivers and in streams. Pronounced variability may result from seasonal or longer patterns, daily runoff or daily-respiration-driven cycles, or stormwater-driven pulses lasting a few hours or less. Variability is dampened in larger waterbodies such as large rivers, lakes, or estuaries.
2. Factors that modify toxicity/bioavailability such as organic carbon, hardness, and pH usually vary in step with fluctuating contaminant concentrations.

* Prepared by Christopher A. Mebane on behalf of the Workshop participants. If further referenced, the following form is suggested:
Mebane, C.A. 2019. A review of the nature and effects of episodic water pollution and implications for aquatic life criteria averaging periods. Workshop discussion initiation paper no. 1 from the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019. 54 pp.

3. Short-term (< 1 day), contaminant exposures, particularly from stormwaters, can be highly toxic, are difficult to monitor for, and can be a pervasive scenario causing degraded waters.
4. The vast majority of aquatic toxicity testing designs expose organisms under constant conditions in order to develop concentration-response models for fixed time periods. Clean Water Act (CWA) section 304(a) ambient water quality criteria for the protection of aquatic life ("Aquatic life criteria") in the USA attempt to bridge this disconnect using a two-number-criteria approach, with different averaging periods for acute and chronic criteria. Acute averaging periods are intended to avoid toxicity from high-magnitude, but short-term pollution events of one day or less, and long-term averaging periods are intended to avoid toxicity from indefinite exposure periods. Considerably less research has been devoted to setting averaging periods or time-to-effect models than has been devoted to establishing concentration responses at fixed-time exposures.
5. Investigators have taken two main approaches to resolve pulse/episodic pollution episodes: experimental and toxicokinetic modeling. Various modeling approaches have been investigated for potential applicability in risk assessment or criteria statements.
6. One of the simplest approaches for modeling time-dependent effects is a modification of the Mancini kinetic model for comparing the speed of action of toxicants in conventional, constant exposure toxicity tests. A variety of example applications of this model to toxicity test data are presented. This model estimates time-independent, asymptotic $LC50\infty$ values and a mortality rate constant, k. The 1/k statistic can roughly be thought of as a half-life for the test populations to reach their time-independent $LC50\infty$. The 1/k "characteristic time of effects" metric ranged from 1.8 to > 96 hours. Most of the examples produced 1/k values of >24 hours.
7. The theoretical support for applying the Mancini 1/k kinetic model of 96-hour constant exposures to estimate effects of time-varying concentrations relies on the observation that toxicity initially increases over time for a given concentration. The increase in toxicity over time results from the uptake rate of toxicants exceeding depuration rate, eventually reaching critical accumulation levels. Later in the exposure, depuration and uptake reach a pseudo-steady state, stabilizing toxicity at an asymptotic LC50. The use of rate constants obtained from time-to-effect curves to set averaging periods to protect against short-term episodic pollution events has not been thoroughly vetted.
8. Speed of action calculations using the Mancini kinetic model are based on the premise that contaminant exposures produce toxicity at or before the time periods recorded. If organisms do not die immediately after receiving a fatal dose, but succumb hours or days later, this delayed mortality may produce highly misleadingly long 1/k speed of action values.
9. Delayed mortality appears to be more of a concern with metals toxicity than for acutely toxic organic chemicals. Among metals, the most work has been done with copper, with results differing in whether pronounced delayed mortality was observed. The more limited information available for Zn suggests delayed mortality could be more of an issue for Zn than for copper.
10. The information reviewed on the common timing of episodic pollution events and the potential for substantial toxicity following short-term episodic events provide ample support for setting acute criteria with averaging periods less than 24 hours. Exposure to concentrated stormwater runoff for only 1 hour can be highly toxic to freshwater organisms, and several studies showed pronounced toxicity from elevated toxicant exposures of ≤ 3 hours.

11. Chronic aquatic life averaging periods have usually followed the 4-day recommendation of the 1985 Guidelines. The rationale for that approach was generally supported by the information reviewed. Ammonia criteria have used a complex criteria expression with 30-day, 4-day, and 1-hour averaging periods. The rationale for this complex approach is not unique to the aquatic toxicity of ammonia, and could be expanded to other chemical criteria, at the cost of increased complexity. Criteria that were derived from extrapolating critical tissue burdens in organs other than gills (or other tissues directly in contact with the surrounding water) to water-column concentrations could be expressed with longer averaging periods than for water concentration-based criteria.

Introduction

Water pollution sources are often episodic, which results in time-varying concentrations in receiving waters. This real-world complexity presents a long-recognized problem for establishing criteria and managing discharge, because toxicity data used to establish species tolerances deliberately minimize variation in exposures over time. The need to develop aquatic life criteria appropriate to, and protective of dynamic real-world situations and the need to test under tightly controlled conditions to minimize influencing factors other than those being tested has created a lab-to-world tension that has persisted since the earliest systematic criteria development efforts in the 1970s to date. In the USA, the strategy has been to apply a two-number criteria framework to address short-term (acute) and indefinite, long-term (chronic) exposures. While the two-number criteria strategy seems broadly accepted within the USA, the optimal duration for defining short- and long-term criteria and allowable frequencies of exceedence has been debated since the strategy was introduced nearly 40 years ago (USEPA 1979).

This review gives examples of episodic or cyclical contaminant patterns in streams and rivers, summarizes literature relevant to the toxicity of short-term or episodic pollution episode, and examines approaches for addressing time varying pollution through two-number acute and chronic averaging periods.

The episodic nature of contaminants exposure in freshwaters

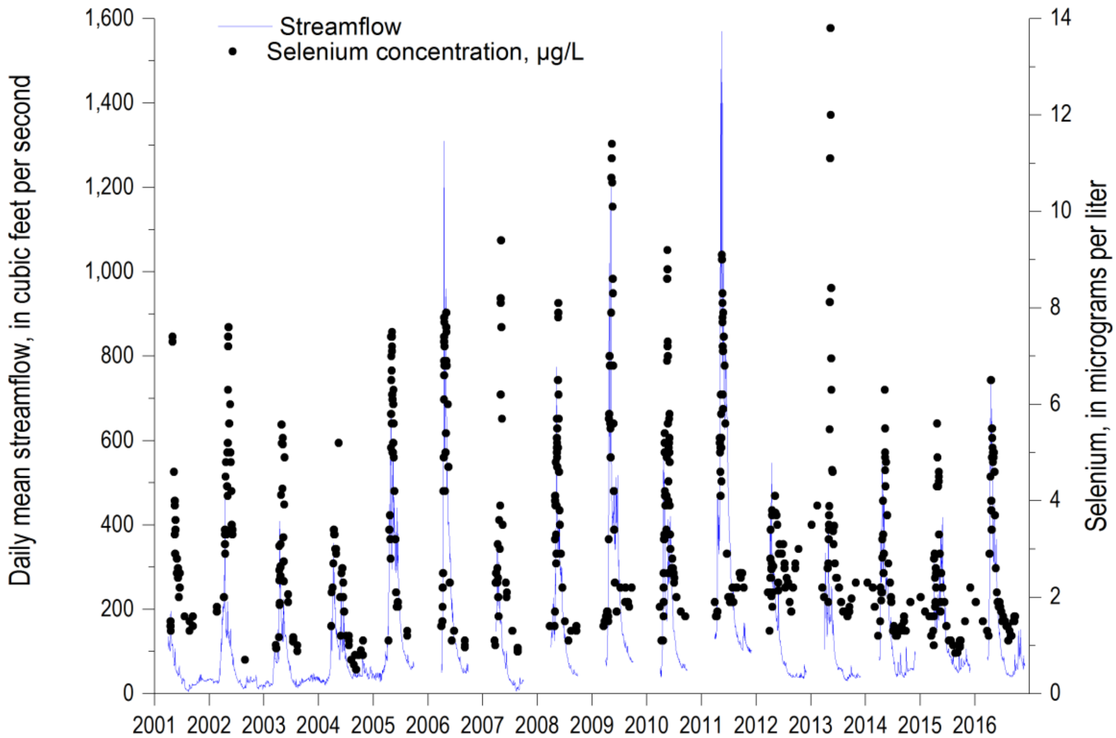
Water pollution sources are often episodic, which results in time-varying concentrations in receiving waters. Episodic pollution results from events and activities such as stormwater runoff from built up or disturbed areas including urban areas, highways, and industrial, agricultural, or mining operations. Brief pulse exposures of pollutants such as ammonia and sulfide can result from activities such as dredging and disposal of dredged materials to maintain navigational channels. Cleaning blowdowns of power plant cooling systems can result in brief discharges of biocides, among many other examples (Kuivila and Foe 1995; Makepeace et al. 1995; Burton et al. 2000; Lee et al. 2004; Corsi et al. 2010; Scholz et al. 2011).

Episodic water pollution can vary more or less randomly from spills and extreme weather events, as well as having more predictable seasonal or daily patterns. For example, seasonal variability associated with annual snowmelt is both highly predictable as to whether and when it will happen in some areas, yet is unpredictable (or at least difficult to predict) as to its magnitude. Several graphs are shown of datasets illustrating these sources of variability (Figure 1 and Figure 2).

Selenium in Idaho's Blackfoot River has a clear seasonally-cyclic pattern with low, stable concentrations during the low-flow conditions present most of the year, punctuated by high concentrations that occur regularly for a few days or weeks each year during the spring snowmelt runoff (Figure 1A). Yet while the timing of the high-flow peaks is predictable, the year to year magnitude is not. For example, flows in 2012 and 2013 were roughly similar, but concentrations were very different.

In small streams, seasonal or runoff driven pulses can be relatively large compared to baseline levels (Figure 1B). In a small stream draining old mine workings in Colorado, lead concentrations were very low until an almost imperceptible bump in flow with the first thaw of spring in April produced a pronounced spike in Pb concentrations to > 2 mg/L. This pulse of Pb declined as conditions re-froze until the onset of the main spring thaw and start of runoff. The pattern of Pb concentrations sharply increasing on the rising limb of the pulse, and declining slowly is consistent with the "first flush" phenomenon that can be an important (and disputed) feature in runoff and stormwater management in mining, urban and agricultural settings (Characklis and Wiesner 1997; Lee et al. 2004; Nordstrom 2009). The first flush concept holds that "*gradual increases in concentrations occur during long dry spells and sudden large increases are observed during the rising limb of the discharge following dry spells (first flush). By the time the discharge peak has occurred, concentrations are usually decreased, often to levels below those of pre-storm conditions and then they slowly rise again during the next dry spell. These dynamic changes in concentrations and loadings are related to the dissolution of soluble salts and the flushing out of waters that were concentrated by evaporation*" (Nordstrom 2009). In such circumstances, very severe pulses of metals or acidity may occur and dissipate after several hours, meaning that routine monitoring programs would be lucky to detect them. For example, a monsoon summer rainstorm raised zinc concentrations in the Red River near Questa, NM from a baseline of about 200 µg/L to about 4500 µg/L and back within about a day or two (Nordstrom 2009). In larger drainages, this characteristic first flush pulse (Figure 1B) may not occur or may be dampened, as the magnitude of pulses would be expected to be relatively less extreme and the duration is longer.

A. Selenium and streamflow in the Blackfoot River, Idaho



From Mebane et al. (2014, with updated data download)

B. Lead in drainage from a Colorado mine

T.P. Chapin / Applied Geochemistry 59 (2015) 118–124

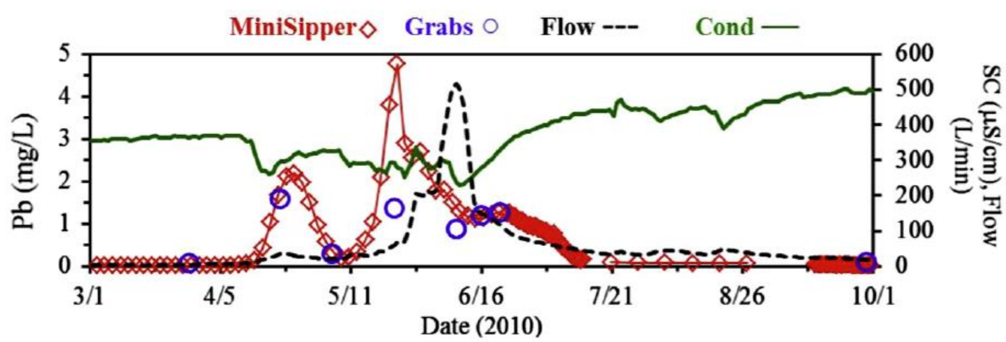


Figure 1. Examples of seasonally fluctuating contaminant concentrations in natural waters in temperate settings, monitored using remotely deployed autosamplers. The Blackfoot River (A) is characterized by persistent, low-level contamination during low flow conditions, with brief pulses of much higher concentrations during snowmelt runoff. Dissolved Pb concentrations in Colorado mine drainage are very low during low flow conditions but spike following the early spring melt (Mebane et al. 2014; Chapin 2015).

Seasonal and diel time series examples of time variable pollutant concentrations are shown in Figure 2. Figure 2A shows copper during snowmelt in a 4th order stream downstream from a mining-polluted tributary which, at the time, received uncontrolled runoff from waste piles (Mebane et al. 2015). In seasonal sampling, Cu increased sharply over the course of a few days in the early runoff, and then declined as runoff continued. In hourly sampling during runoff, copper concentrations varied by over 3X within 6 hours, as the snowmelt runoff increased with the warmth of the day, declining as the temperatures dropped again at night (not shown). Thus, these diurnal changes resulted from snowmelt driven mobilization and dilution of copper, not diel respiratory changes.

Rainfall brings faster changing conditions (minutes to hours) than snowmelt (days to weeks). Copper increased from about 5 to 25 µg/L within an hour when it started raining during a diel processes study in a 3rd order stream, Silverbow Creek, Montana. Within 22 hours, Cu concentrations had returned to baseline concentrations (Balistrieri et al. 2012). A marked peak-trough-peak in concentrations captured early in the event may be from incomplete mixing as the copper contaminated runoff entered the stream at different side drainages. In both the runoff and rainfall driven times series, the increases of copper were accompanied by a pulse in dissolved organic carbon (DOC). Rainfall on small drainages, such as urban streams where the headwaters are dominated by impervious surfaces such highways, buildings, parking lots, and turf may produce short, relatively large contaminant pulses. In Figure 2C, for example, Zn concentrations in runoff during the first hour after it started raining approached 2000 µg/L, before declining to about 100 µg/L within about two hours. Grab samples collected from the first hour of the urban highway runoff were often highly toxic (Kayhanian et al. 2008).

In some settings, most notably in smaller streams, photosynthesis and respiration driven diel cycles of O_2 and pH can cause substantial diel changes in concentrations of some contaminants. For instance, manganese, Fe, Zn, N, P, and NH_3 may show strong diel patterns, but Cu and Pb show minimal diel changes (Nimick et al. 2003; Gammons et al. 2011; Nimick et al. 2011). In Montana streams, short-term (two-hour to daily) variations in stream chemistry made up a large proportion of the variation seen on much longer time frames (Nagorski et al. 2003). Figure 2D shows daily concentrations in Zn cycling between about 15 and 70 µg/L, driven by pH changes (Nimick et al. 2003).

In contrast to these examples of variation on short-time scales, groundwater controlled stream chemistry can be relatively stable for months at a time. The South Fork Coeur d'Alene River in Idaho has elevated Cd and Zn concentrations because of groundwater inputs that transported dissolved metals from old mine dumps, workings, tailings, or disturbance. Unlike the runoff scenarios, here the maximum concentrations occur during low- flow periods when flows are sustained by groundwater. Instead of distinct peaks, maximum seasonal Zn and Cd concentrations occur in comparatively stable plateaus that may last for several months (Figures 2E, F). The contaminated South Fork Coeur d'Alene River eventually reports to the Spokane River. In that much larger Spokane River, the concentration variability is greatly dampened, varying only by about 2X compared to >10X in the smaller river. Only Cd plots are shown, but the Cd and Zn concentrations invariably track together, with Cd mass concentrations about 200X lower than Zn concentrations (Mebane 2006; Mebane et al. 2012). Metals concentrations in wastewater influenents may be highly variable, with up to a 10-fold variation in concentrations over the course of a day. Although few published data comparing metals in influents and effluents were reviewed, in at least one case study with copper, both the absolute concentrations and the fluctuations in concentrations were greatly dampened in treated effluents compared to influents (Figure 3).

Ammonium concentrations in a river downstream of a secondary wastewater treatment showed substantial variations. There was a pronounced diurnal pattern in effluent ammonium concentrations which rose sharply near 12:00 noon each day and declined to lowest levels by the following morning. These instream patterns followed effluent discharge rates (Figure 4).

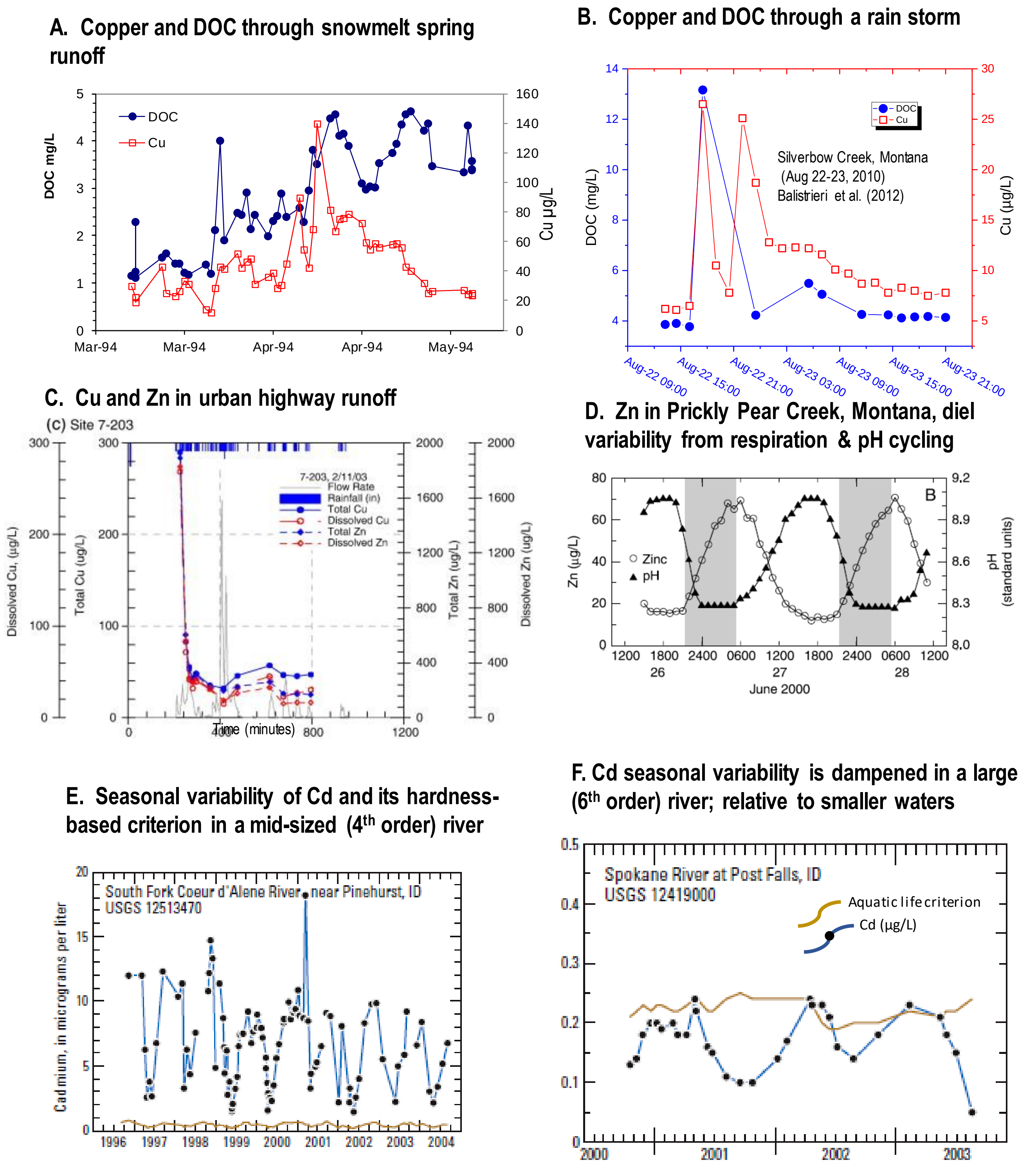


Figure 2. Examples of seasonally or daily fluctuating contaminant concentrations in ambient waters, with variability driven by snowmelt runoff or diel photosynthesis and respiration cycles (Nimick et al. 2003; Mebane 2006; Kayhanian et al. 2008; Balistrieri et al. 2012; Mebane et al. 2015).

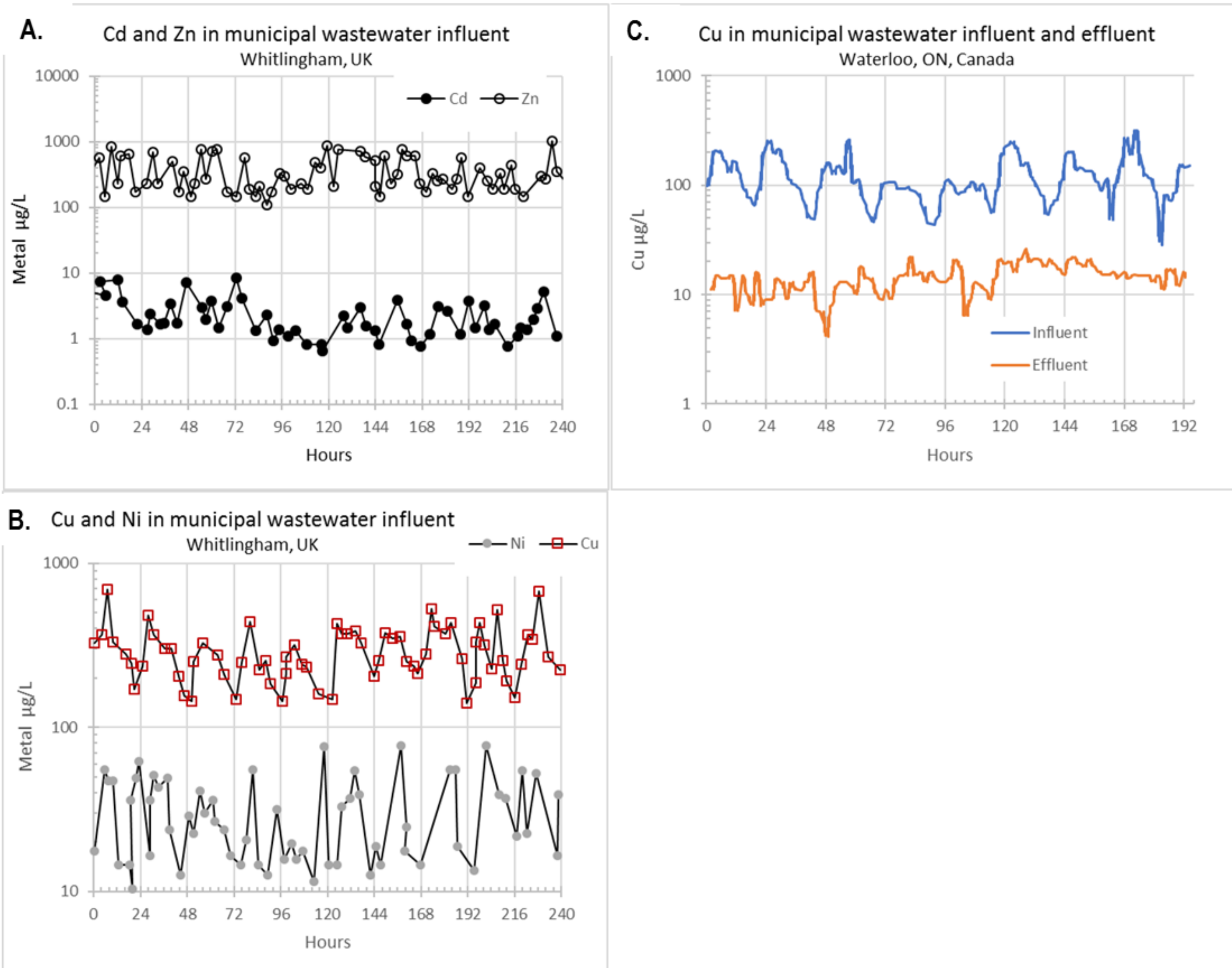


Figure 3. Examples of daily fluctuating metals concentrations in municipal wastewater influents and effluents. Values were digitized and redrawn from the published figures (Melcer et al. 1988; Goldstone et al. 1990b, a).

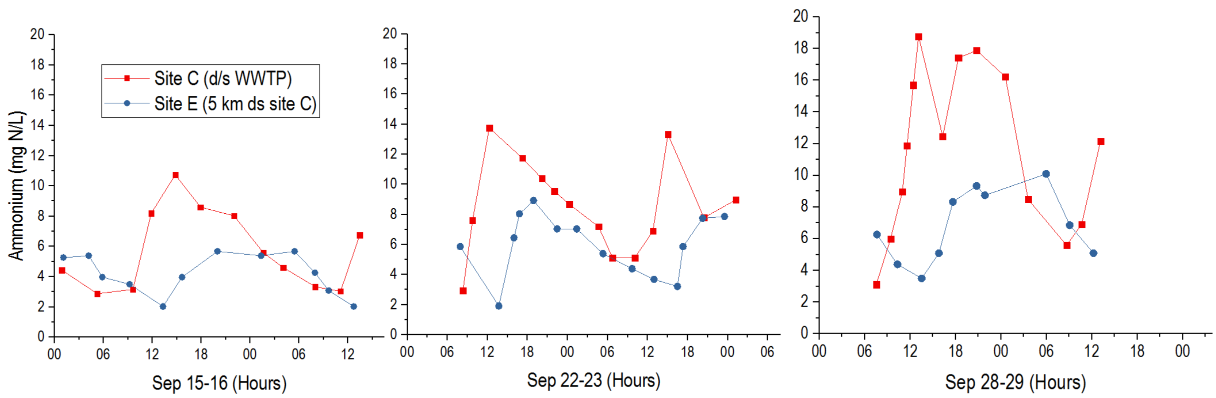


Figure 4. Diurnal fluctuations of ammonium in the South Skunk River, Iowa, downstream of a secondary wastewater treatment plant, September 1984. Data digitized and redrawn from Crumpton and Hersh (1987).

The episodic nature of water quality parameters that affect toxicity/bioavailability

The hydrologic cycles and weather of course do not only affect solutes or characteristics of regulatory interest, but rather virtually all aspects of stream chemistry. Many are controlled by the same weathering and dissolution mechanisms and are expected to co-vary (Miller and Dreaver 1977; Lewis and Grant 1979; Nagorski et al. 2003). Taking the relatively well characterized Coeur d'Alene, Idaho, region as an example, the major ions Ca, Mg, Na, sulfate, and dissolved inorganic carbon tended to rise and fall in concert with Cd and Zn (Balistrieri and Blank 2008; Mebane et al. 2012; Clark and Mebane 2014). As result, the change in risks of increasing or decreasing Cd and Zn concentrations are dampened by the concurrent increases or decreases in Ca and other ions. This can be seen in the plots 2E and 2F where the aquatic life criteria lines bend up and down in synchrony with the Cd concentrations.

In most runoff situations, major ions will tend to be diluted. These changes would tend to increase the relative toxicity of metals. The pH of streams will change during snowmelt or rain driven runoff, likely dropping if flow paths are short and contact time with rock and soil is limited. However, in long flow paths or in urban areas where the watershed land use is dominated by concrete and asphalt, the pH may rise. Declining pH would tend to increase the toxicity to aquatic animals and mobility of some metals such as Pb and Cu, but decrease the risk of toxicity to aquatic animals from other metals such as Cd and Zn or from ammonia.

The interplay between Zn and pH in diel cycles in streams is such that the most toxic conditions for Zn occur in late afternoon when pH is highest, but actual Zn concentrations are at their daily lowest at that time (Figure 2D). Peak Zn concentrations occur just before daylight as pH hits its daily low. Ammonia follows a similar pattern in situations where there are strong daily cycles (Gammons et al. 2011).

Dissolved organic carbon is expected to increase in rain or snow driven runoff, as decaying leaf litter and other vegetation is dissolved and transported to streams. In cold-temperate climates that receive a seasonal snowpack, stream dissolved organic carbon levels would be expected to increase rapidly as spring melt commences, peak before maximum discharge, and decrease quickly as melting continues (Boyer et al. 2000). In the snowmelt and rainfall examples from mining disturbed areas (Figure 2A,B), DOC generally increased in concert with copper. In the rainstorm example, the DOC dropped back to baseline concentrations within 6 hours after the rain began and remained low, although Cu increased in a second peak that was not matched by DOC (Figures 2A,B). In a Montana river, DOC was shown to change two-fold on daily cycles during stable flows, presumably influenced by uptake and release from heterotrophic bacteria in periphyton (Parker et al. 2010)

In most, but not all instances reviewed, the overall direction of change in the water quality parameters that modify toxicity/bioavailability would tend to mitigate potential risks from toxic substances. An exception was ammonia and pH changes following a rainstorm, in which ammonia concentrations increased while pH also increased (Lawler et al. 2006), thereby increasing potential ammonia toxicity.

One point of these examples is that it is difficult to generalize whether the concurrent change in contaminants along with factors that modify toxicity/bioavailability keeps relative risk to aquatic organisms about the same, exacerbates or mitigates risk. Thus, to estimate relative risk of fluctuating toxicant concentrations in the environment, it is important to evaluate the co-varying factors that may influence toxicity/bioavailability such as pH, DOC, and major ion concentrations. With metals such as Cu for which aquatic risk is estimated through biotic ligand models (BLMs), this presents a logical problem as BLMs are based on calculations of chemical equilibria, and rapidly fluctuating environmental conditions are doubtfully in chemical equilibria. In particular, depending on the experimental design, DOC and Cu complexation may take as little as 5 minutes to more than 24 hours to approach equilibrium (Ma et al. 1999; Meyer and Adams 2010, SI). As a result, some scientists have argued that aquatic toxicity tests that that did not allow 24 hours for Cu and DOC to reach or approach equilibrium, may exaggerate Cu toxicity and should be discounted in risk evaluations or criteria derivation (Santore et al. 2001; Meyer and DeForest 2018). Situations such as those illustrated in Figures 2A, B, and C caution that non-equilibrium conditions may not be unusual in lotic habitats, at least for slower reactions such as DOC-Cu complexation.

A second point of these examples is that discharge management approaches that artificially decouple naturally co-varying stream chemistry parameters can produce unrealistic scenarios. Consider an example calculation of discharge permit limits for hardness-based metals criteria that factor into the chronic waste load allocation calculations:

(1) for diluting streamflow, use the 7Q10 which is the lowest 7-day average flows occurring in a 10-year period of record, or in other words, a very dry condition which flows meet or exceed 99.8% of the time (USEPA 1991);

(2) the 95th percentile background metals concentrations in the receiving waters;

(3) the 95th percentile effluent flow volumes and metal concentrations; and

(4) the 5th percentile stream hardness value.

In real world situations, these four factors would not co-occur. In many real streams and rivers, water chemistries are highly correlated with flows, although as described earlier even the direction of the relations vary with constituents and settings. The use of upper-end percentiles instead of maximum or minimum values is undoubtedly intended to reflect a conservative but less than worst case discharge scenario and to be appropriately protective. However, decoupling naturally inextricable physical and geochemical processes can create scenarios that would never occur in nature. For example, it may be inappropriate to assume receiving streams have relatively low hardness values during low-flow conditions, while in situ monitoring data actually suggest receiving streams have higher hardness values (e.g., beyond the 5th centile hardness) during low flows, when geochemical ions are concentrated.

Averaging periods for criteria – background and controversies

The recognition that environmental concentrations fluctuate and criteria for pollution control need to recognize this has long been embodied in criteria approaches in the US. In an early strategy for developing water quality criteria the EPA defined criteria in terms of an average 24-hour concentration and an instantaneous-ceiling concentration (USEPA 1979). The incorporation of an average value was because aquatic organisms can be expected to tolerate some excursions over this mean so long as the excursions are not too high or too frequent. "*The 24-hour period was chosen instead of a slightly longer or shorter period in recognition of daily fluctuations in waste discharges and of the influence of daily cycles of sunlight and darkness and temperature on both pollutants and aquatic organisms.*" Recognizing that the usual toxicity test conducted at a constant concentration in static conditions for 96 hours bore little resemblance to real world conditions, the EPA had gone to considerable effort to compile toxicity data over time in tests, and examined over 700 acute tests to estimate toxicity ratios between 24 hour and 96 hour exposures, and ratios between flow through and static tests (USEPA 1978).

These efforts informed EPA's 1985 "*Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses*," (Stephan et al. 1985). To address the problem of time-varying concentrations of pollutants in aquatic environments, EPA's 1985 national guidelines (the "Guidelines") for deriving aquatic life criteria established a two-number framework expressing acute and chronic criteria in terms of concentration, duration, and exceedence frequency of allowable exposures. Chronic criteria magnitudes (Criterion Continuous Concentration [CCC]) were to be expressed as a 4-day average concentration not to be exceeded more than once every 3-years, and acute criteria (Criterion Maximum Concentration [CMC]) were to be expressed as 1-hour average concentrations, not be exceeded more than once every 3-years. These durations were consciously set to be shorter than the typical duration of acute and chronic toxicity tests for two reasons: 1) to address fluctuating concentrations in field conditions and 2) to account for toxicity tests that generally encompass different life stages which might have different sensitivities, so that effects might disproportionately occur during the portion of the test in which a sensitive life stage is present (Stephan et al. 1985). Quoting from Stephan et al. (1985):

> *Even though only a few tests have compared the effects of a constant concentration with the effects of the same average concentration resulting from a fluctuating concentration, nearly all the available comparisons have shown that substantial fluctuations result in increased adverse effects [citations omitted]. Thus if the averaging period is not to allow increased adverse effects, it must not allow substantial fluctuations… For the CMC the averaging period should again be substantially less than the lengths of the tests it is based on, i.e., substantially less than 48-96 hours. One hour is probably an appropriate averaging time because high concentrations of some materials can cause death in one to two hours. Even when organisms do not die within the first hour of so, it is not known how many might have died due to delayed effects of this short of exposure. Thus it is not appropriate to allow concentrations above the CMC to exist for as long as an hour (Stephan et al. 1985).*

Following the 1985 publication of the Guidelines, all criteria documents that were written from 1985 until 2001 appear to have included a 1-hour averaging period for their acute criteria expressions. However, the 2001 acute Cd criteria document set the acute averaging period to 24-hours. No explanation was included in the document for the deviation from the Guidelines (USEPA 2001). The (2007) Cu criterion followed suite with a 24-hour average criterion, again without discussion. As described below the acute cadmium criterion was again updated in 2016 with a 1-hour averaging period on the basis that it was more appropriate in the absence of specific data supporting a change to a 24-hour averaging period (USEPA 2016a). The 1-hour average acute criterion recommended in the Guidelines was retained in acute ammonia criteria (USEPA 1999, 2013).

Chronic criteria averaging periods

Chronic criteria averaging periods have generally retained the 4-day average concentration recommended in the 1985 Guidelines. The rationale for the time period, which is shorter than the typical chronic test, is two-fold. First, concentrations in the field are typically much more variable than concentrations in laboratory tests, and in some instances, variable concentrations of toxicants sometimes have been more toxic than constant concentrations when the comparisons are based on average concentrations during the exposure (Thurston et al. 1981; Seim et al. 1984). By shortening the averaging period to which the criterion applies, the average concentration over the entire exposure will be below the criteria concentration, increasingly so as the variability of the concentration increases. Secondly, chronic tests generally encompass different life stages, which might have different sensitivities, so that effects might occur only, or disproportionately, during the fraction of the test in which a sensitive life stage is present, rather than cumulatively over the whole test (Stephan et al.

1985). For example, in a follow-up to the Seim et al (1984) study, EPA's Western Toxicology Testing Facility tested the responses of steelhead trout to copper, with tests initiated with fish of different stages of development. Sudden increases in mortalities occurred as fish aged into sensitive stages (Chapman 1994). The same phenomenon was observed in early life stage tests with rainbow trout and Cd, Zn, and Pb, with survival remaining stable for weeks, punctuated by brief mortality events (Mebane et al. 2008).

The exceptions to the recommended 4-day averaging period for chronic aquatic life criteria appear to have been limited to ammonia and selenium. In the case of the chronic selenium criterion duration, Se toxicity occurs through bioaccumulation and food web transfer. These processes are thought to function on longer time periods than 4-days (USEPA 2016b).

The chronic ammonia criterion was updated in 1999 with a duration of 30-days that was further constrained by the provision that no 4-day period within the 30-days should exceed 2.5X the 30-day average. This same criterion duration was also retained when the chronic ammonia criterion was again updated in 2013 with the explanation that "*an averaging period of 30 days could be used when exposure concentrations were shown to have limited variability*"(USEPA 2013). How limited the variability in exposure concentrations needed to be to be considered of "limited variability" was not explained. Rather, the basis for the provision that no 4-day period within the 30-days should exceed 2.5X the 30-day average was similar the concept of acute-to-chronic ratios (ACRs). Ratios of quasi-chronic 7-day toxicity endpoints to 30-day early-life stage toxicity endpoints from across different studies tests were compared, and the 7-day values were found to be about 2.5X higher than the 30-day values. Other data reasons led to judgements limiting the time period of the cap on the 2.5X fluctuations within the 30-day average to 4-days (USEPA 1999).

While the analyses of toxicological data supporting the ammonia criteria chronic averaging periods clearly explain *how* the criteria deviate from the national guidelines averaging period recommendations, the documents are silent on *why* the deviation was considered more appropriate for chronic ammonia toxicity than for other criteria substances. The daily fluctuations in ambient ammonia concentrations downstream of a wastewater outfall can indeed be substantial (about 8 fold in figure 4; (Crumpton and Hersh 1987; Gammons et al. 2011)), but it is not clear that the variability in fluctuating concentrations or toxicology was the primary consideration for the different approach with ammonia.

Acute criteria averaging periods

The USEPA (2016a) update to the Cd criteria appears to be the only criteria document that explicitly discussed the acute averaging period, beyond just repeating the criteria guidelines wording. Because of the limited nature of time-to-effect investigations and absence of additional supporting information, EPA set the acute duration in the 2016 cadmium criterion to be consistent with the 1-hour duration with the 1985 Guidelines (USEPA 2016a).

One comment letter, focusing on Cd, argued that few published studies included time-to-effect data showing toxicity occurring prior to 24-hours, and Cd's mode of toxic action made it unlikely to be a fast acting toxicant (UWAG 2016). Santore et al. (2016), focusing on copper, argued that "*a 1-hour averaging period is overly conservative, and that the 24-hour averaging period recommended in the 2007 copper criteria document would be suitably protective for sensitive invertebrates.*" Their argument was based on the observation that acute LC50s are a function both of concentration and time, and longer exposure durations usually produce lower LC50s until approaching an asymptote when toxicity no longer increases with increasing duration of exposure. The time to approach this time-independent LC50 was called the "averaging period" and was estimated by non-linear regression analysis of a dataset of LC50s calculated at different times during tests, such as survival at 24,48, 72, and 96 hours. There was no mention of the issue of mortalities following exposures. The time-independent LC50 "averaging periods" for invertebrates ranged from 17 – 240 hours, averaging 66 hours (Santore et al. 2016). The "averaging period" calculations summarized in Santore et al. (2016) built upon an USEPA collection of the "speed of action of metals acute toxicity to aquatic life" calculations (USEPA 1995). This collection, however, largely did not contain any contextual description or explanations, and therefore, the "speed of action" calculation approach is examined in more detail later in this review.

What difference do acute averaging periods make?

Depending on the situation, whether an acute criterion was based on a 24-hour, 1-hour, or some other averaging limit might have no meaningful difference or a profound difference in implementation. Water quality monitoring can be costly, a small minority of water bodies in the United States are systematically monitored, and of those, the frequency of discrete monitoring might be annual or seasonal. In such settings, such as for State's bi-annual compilation of waters not meeting standards, averaging periods are nearly moot. The available point samples are compared to criteria concentrations without regard to averaging period, because water concentration measurements over the course of an averaging period is seldom available.

Effluent discharge monitoring requirements for major municipal or industrial wastewater facilities typically have weekly or more frequent monitoring for parameters that have limits and for which sampling and analysis is relatively simple and low cost (e.g. total suspended solids, biological oxygen demand, ammonia, chlorine). Parameters such as flow and temperature may be continuously monitored. Parameters for which sampling is more expensive or require specialize procedures, such as metals or chlorinated organics, may be considerably less frequent[1]. Requirements, if any, for receiving water discrete sampling are quite variable at the discretion of the permitting authority, such as monthly, 4X a year, annual, once per 5-year permit cycle, or none.

[1] Brian Nickel, EPA Region 10, written communication, 29 January 2018

Debates over acute criteria averaging periods will become increasingly important as sensor and automated micro-chemistry technology as well as automated methods for high frequency sampling improve (Chapin 2015; Blaen et al. 2016). Secondly, real-time estimates using site-specific surrogate relationships have been increasingly used in remote locations or when direct analyses of the parameters of interest are expensive or slow ("Water watch - what is a surrogate?" website). For example, characteristics such as pH, conductivity, flow, turbidity, fluorescence, dissolved gases, and acoustics can be measured directly in real-time and can be correlated with results of discrete water quality sampling for substances as diverse as mercury, arsenic, and oil and gas byproducts in water (Dittman et al. 2009; Etheridge 2015; Son and Carlson 2015). The resulting regression-based surrogate models can be predicted with reasonable accuracy as a function of real-time measurements to allow decision making.

In some instances, criteria can vary widely over the course of a day, such as an example of the acute ammonia criterion in a productive, meandering, slow-moving stream. The criterion varied between about 1 to 14 mg/L ammonia (Figure 5). If a discharger were to be given limits for such a receiving water, should the limits be based on the daily average or perhaps something close to the minimum?

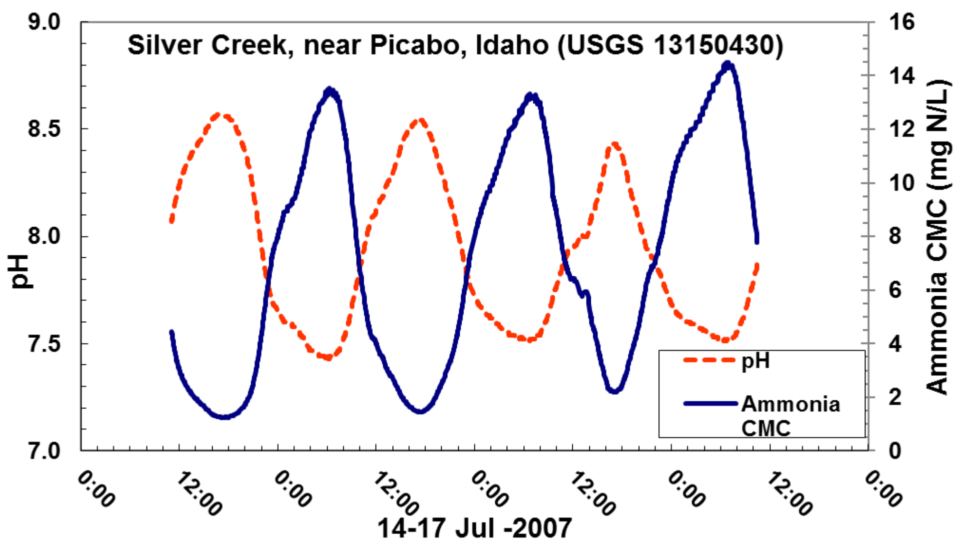


Figure 5. As with contaminant concentrations, criteria values can vary widely over short time periods due to daily cycles or weather and flow events. In this example, the ammonia CMC varied 6X over the course of a day. In a study in which actual diel ammonia concentrations were measured, ammonia concentrations were lowest during late afternoon when pH was highest and ammonia would be expected to be most toxic (Gammons et al. 2011).

Let us next consider two scenarios comparing hypothetically allowable pulse concentrations under different criteria averaging schemes. As with any hypothetical discharge scenario, these are a bit contrived, but have some similarities to real-world rainfall driven pulse scenarios.

Scenario 1: Short stormwater pulse events occur in the City of Gaussberg, running into Gaussian Creek, which is a flashy, urban, headwaters stream with a very short travel time from pavement to stream.

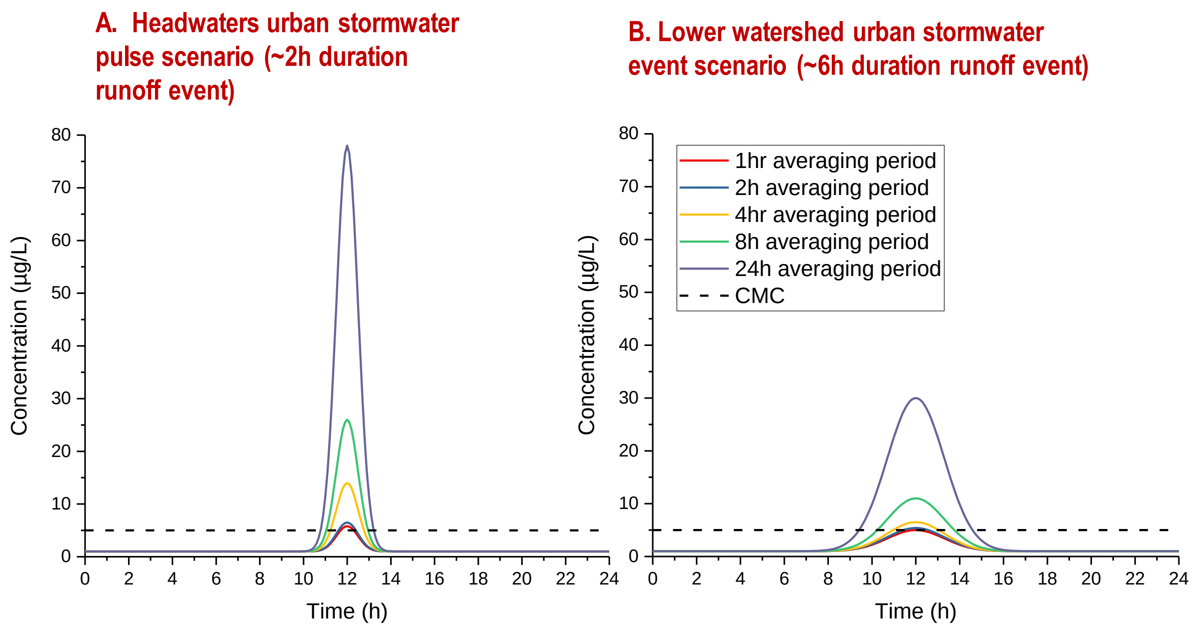
The event duration is about 2 hours, baseline to baseline. This neat and predictable city has best-management practices that limit their stormwater runoff to streams by applying a Gaussian function to limit the magnitude (amplitude) of their stormwater pulses to just meet the aquatic life criteria magnitudes for the duration of the event. The City Planning Commission is exploring the effect of different averaging periods on pulse magnitudes. The Criterion Maximum Concentration is 5 µg/L and the stream has a background concentration of 1 µg/L (this tight range between background and criteria can occur with some real criteria, such as copper). The deterministic weather patterns affecting this stream are such that storm peaks always occur at noon, and the storm duration is always about 2 hours.

Equation 1: Gaussian pulse curves: $y=B+A\cdot e^{\left(\frac{-(x-\mu)}{2\sigma^2}\right)}$

> Where y is the pulse concentration, B is the background concentration, A is the amplitude of the peak, x is time in hours, μ is the time at the peak, and σ is the standard deviation of the function, which defines the width of the pulse.

In this scenario, when the Gaussian stormwater concentrations are limited to 1-hour maximum average concentration of 5 µg/L, the 24-hour average concentrations are much lower, only 1.2 µg/L, which is barely above the background concentration of 1 µg/L in this scenario (Figure 6A). If the City Planners changed their rules to allow the CMC of 5 µg/L to be met as a 2-hour maximum average, the 1-hour maximum average would still be constrained to only 5.7 µg/L and the 24-hour is still only 1.3 µg/L. Exploring the consequence of changing their rules to only constrain the 24-hour average to the 5 µg/L CMC, the planners note that the 1- and 2-hour maximum concentrations would be allowed to reach 65 or 45 µg/L respectively. They decide to consider another scenario.

Scenario 2: Further downstream, Gaussian Creek becomes the Gaussian River, where the longer travel times from the many headwaters results in longer stormwater pulse events. The event duration in this larger water body is always about 6 hours, baseline to baseline, and the peak still somehow always occurs at noon (Figure 6B). In this longer pulse event scenario, in order to comply with the 5 µg/L concentration, the magnitude of the pulse is lower under all averaging periods. The curves for the 1- and 2-hour averaging periods are almost indistinguishable, and both would constrain the 24-hour average concentration to 1.6 µg/L. Conversely, our intrepid planners notice that if they were to adopt a 24-hour averaging period, this would allow 1- and 2-hour maximum average pulse concentrations of 30 and 28 µg/L, respectively. Our hypothetical city planners decide that this is getting complicated and perhaps they should consider toxicity resulting from brief chemical exposures before making a recommendation, which is where this review goes next.



Metric (concentrations in µg/L)	1h CMC averaging Period	2h CMC averaging Period	4h CMC averaging Period	8h CMC Averaging Period	24h CMC Averaging Period
CMC	5.0	5.0	5.0	5.0	5.0
Background	1.0	1.0	1.0	1.0	1.0
Amplitude	4.8	5.5	13	25	77
Peak	5.8	6.5	14	26	78
Max. 1h average	5.0	5.7	12	22	65
Max. 2h average	3.8	5.0	8.5	15	45
Max. 4h average	2.5	2.7	5.0	8.6	25
Max. 8h average	1.7	1.9	3.0	5.0	13
24h average	1.2	1.3	1.7	2.3	5.0

Metric	1h CMC averaging Period	2h CMC averaging Period	4h CMC averaging Period	8h CMC Averaging Period	24h CMC Averaging Period
CMC	5.0	5.0	5.0	5.0	5.0
Background	1.0	1.0	1.0	1.0	1.0
Amplitude	4.1	4.4	5.7	10.0	29.5
Peak	5.1	5.4	6.7	11	31
Max. 1h average	5.0	5.3	6.5	10.7	30
Max. 2h average	4.7	5.0	6.1	10.0	28
Max. 4h average	3.9	4.1	5.0	8.0	22
Max. 8h average	2.6	2.8	3.3	5.0	13
24h average	1.6	1.6	1.8	2.4	5.0

Figure 6. Influence of different averaging periods in the allowable contaminant concentrations in simulated stormwater pulses. In each scenario, the "stormwater" pulse magnitude was constrained by setting the pulse amplitude so that the maximum average concentration, averaged over different durations, just reached the criterion maximum concentration (CMC), blue shaded boxes.

The short pulse (Scenario A) is intended to be reminiscent of a stormwater pulse in a flashy, urban, stream setting where the impervious parking lot & highway catchment sends runoff quickly to the headwaters channel. The more sustained pulse (Scenario B) is intended to be reminiscent of a larger, 3rd order urban stream where the travel time storm to stream is greater, causing a lower concentration, broader duration pulse.

Toxicity resulting from brief chemical exposures

The speed of action of toxicants varies tremendously and is a function of the intrinsic mode of action of the toxicant, the organism, and the developmental stage of the organism, among other likely factors. For instance, when pyrethroid insecticides are directly applied to isolated tissues, cell functions can be can interrupted within milliseconds; when applied directly to insects immobilization or irritation/avoidance may occur within seconds of exposure, and when applied to water bodies by dosing or overspray, massive changes to aquatic communities can be expected within a few hours (Clark and Brooks 1989; Coats et al. 1989; Davies and Cooke 1993; Friberg-Jensen et al. 2003). Similarly, tricaine methanesulfonate (MS 222) takes effect within seconds when applied to immobilize or euthanize fish, and low pH or low DO conditions can be lethal within minutes. In contrast, the time course of waterborne toxicity of metals can vary greatly, on the scale of hours to days (e.g., Marr et al. (1998), this review).

A compilation of toxic responses resulting from short-term exposures to various chemicals is given in Appendix 1. The compilation suggests some generalizations. First, mode of action or chemical class was not always an obvious, simple screen for fast or slow-acting toxicants, in part because the same substances can have more than one mode of toxic action. Certainly, some insecticides, cyanide, and ammonia stand out as being fast acting toxicants under some combinations. Speed of responses of fish to metals and other inorganics show a wide range.

Second, sublethal behavioral effects such as avoidance by fishes or drift of insects in stream are typically fast responses, initiated within minutes to hours. This stands to reason, as presumably such responses have evolved as escape mechanisms to evade unsuitable conditions. For mortality responses, fast speed of action and high sensitivity (low EC50 values) appear to co-occur. Tests with high acute toxicity (that is low effects concentrations) were often also the faster responses with mortalities beginning within hours of exposure.

Third, relatively few tests were designed to allow for observation of latent or delayed mortality. Failure to account for delayed mortality can lead to misleading interpretations of the time course of acute mortality, as explored more in a later section. Delayed mortality from short-term exposures of about 24 hours or less appears to be more of an issue with tests with metals than for organic contaminants, major ions, or ammonia. Among the metals, by far the most work was found with Cu and fathead minnow. Some of the studies also suggested that delayed mortalities were more of a risk following acute Zn exposures than from Cu exposures (Brent and Herricks 1998; Zhao and Newman 2004; Diamond et al. 2006). The relatively limited information showing short-term Zn exposures, such as those that might results from urban stormwater, might be more likely than copper to produce delayed mortality is more suggestive than definitive. The vast majority of the studies reviewed relevant to potential toxicity of brief contaminant exposures were laboratory toxicity studies. A few real-world field or quasi-field studies are illustrated in more detail.

A.

B.

Figure 7. In the Pacific Northwest, returning coho salmon wait in saltwater before entering their small spawning streams until winter rains increase the streamflow. In urban areas of the Puget Sound, WA these winter rains also bring pulses of contaminated highway runoff. Particularly during the first flush (early season rains), this causes a Pre-spawn Salmon Mortality Syndrome (PSMS), starting with disorientation and gaping, progressing to loss of equilibrium and death within a few hours (Scholz et al. (2011). **A.** This otherwise healthy pre-spawn female died within hours after entering freshwater, still carrying her eggs; **B.** Toxicity testing with adult salmon in highway runoff mimicked the PSMS observed in field surveys (Spromberg et al. (2016). Photos courtesy of Jenifer McIntyre, Washington State University

Urban stormwater is a commonly encountered setting for episodic pollution events. The flashy nature of small urban streams means that streams can be runoff-dominated for hours or days after a rainstorm, with minimal dilution. Several studies have captured storm runoff and tested its toxicity. High profile settings for episodic stormwater toxicity are the deaths of salmon in urban streams in the Puget Sound region of the Pacific Northwest. There, adult coho salmon stage in salt water near the mouths of their spawning streams until the fall/winter rains arrive and raise the streamflow to allow them to swim upstream to spawn. The coho get nicknamed the “backyard salmon” because of their highly visible and celebrated spawning in very small streams in city parks and suburban settings. In the 1990s, following otherwise successful habitat restoration projects, fish did recolonize the reconnected and restored reaches. However, in surveys of restored reaches following rains, before they could spawn, fish displayed anomalous behavior with erratic surface swimming, gaping, fin splaying, and loss of orientation and equilibrium. Death followed within hours (Figure 7). This “Pre-spawn Salmon Mortality Syndrome” was reproduced in controlled experiments with field collected stormwater. Polycyclic aromatic hydrocarbons (PAHs) from combustion products, paved surfaces, and vehicle leaks were the leading suspects as the causative agents, although published attempts to reproduce the effects with laboratory cocktail mixtures of PAHs and metals were unsuccessful (Scholz et al. 2011; McIntyre et al. 2015; McIntyre et al. 2016; Spromberg et al. 2016).

The toxic nature of urban stormwater runoff often changes greatly over the course of an event, emphasizing the ease with which incorrect conclusions could be reached from routine (i.e., infrequent) sampling. In this storm (Figure 2C and Figure 8), samples collected during the first hour of runoff were highly toxic, but 3 hours into the storm, toxins appeared to have been largely washed away, and the runoff was nontoxic. Toxicity was largely attributed to Zn and Cu in highway runoff. Concentrations likely reached 10-20 cumulative acute criteria units (CCUs) during the first hour of the storms and then rapidly dissipated (Kayhanian et al. 2008).

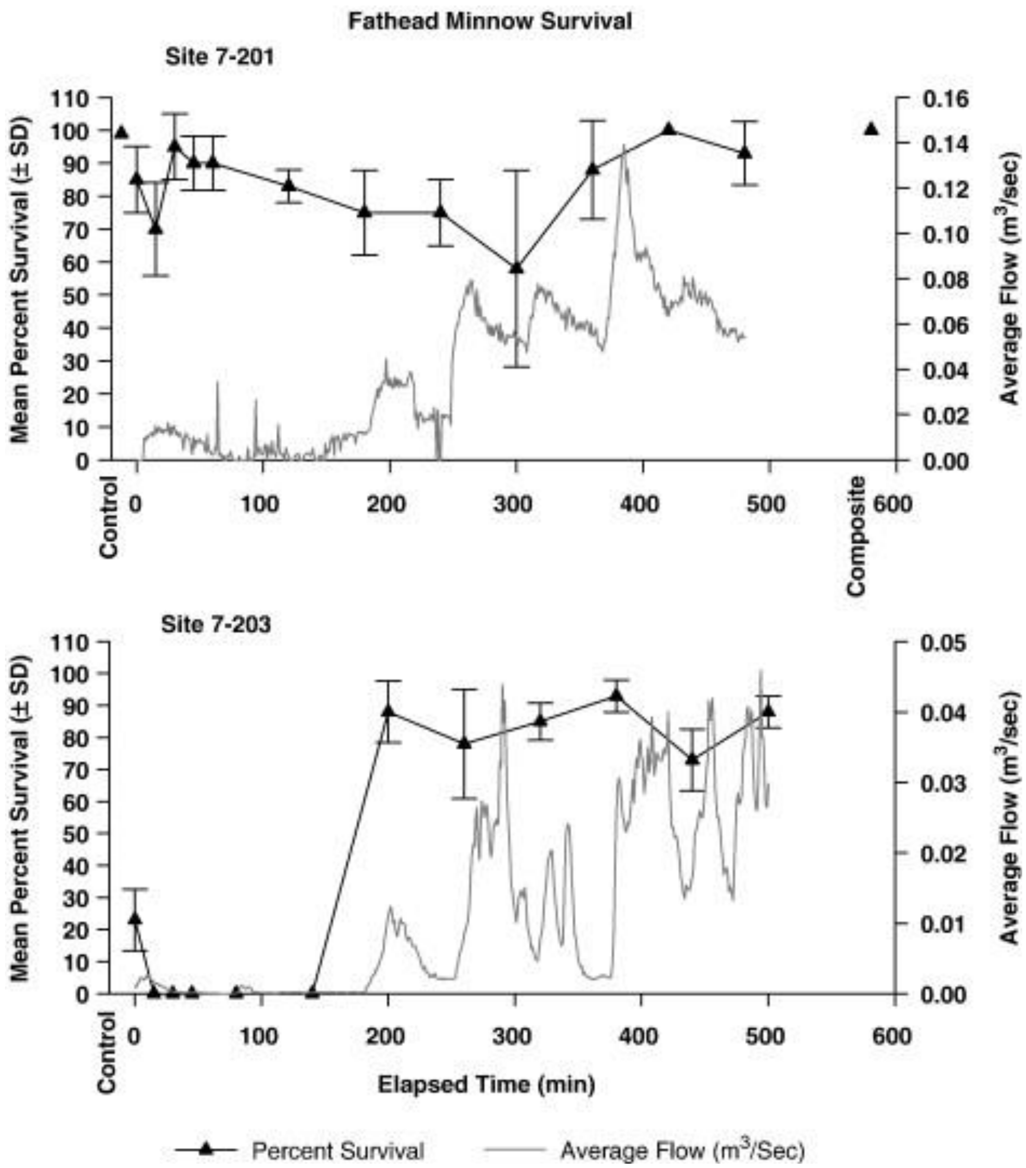


Figure 8. Rapidly changing toxicity of highway stormwater runoff to fathead minnow in samples collected at different times during a single storm event at two sites in the Los Angeles, California basin. Elapsed time is time samples were collected after the onset of runoff; survival is after 7-days. In these tests, samples collected in the first hour of runoff were highly toxic, yet samples collected late in the runoff (3+ hours) had low toxicity (Kayhanian et al. 2008, their Figure 5)

Finally, it should be mentioned that compiling reports of toxicity of short-term exposures was difficult because this information is typically omitted from publications reporting acute toxicity of chemicals. The vast majority of published acute toxicity tests report results for the end of the exposure duration, such as 48 or 96 hours. While effects at shorter periods are presumably recorded in most toxicity tests, most publications only give details on the ending effects concentrations. In some cases, graphs of toxicity over time were digitized in order to extract data for model calculations. Initial search strategies using a curated databases (Scopus) with search terms such as “pulse toxicity”, “delayed mortality” “latent mortality” were of limited direct value, and most relevant studies were located by inspections of references cited or forward citation (cited by) lists from publications. Such a search strategy is assuredly not exhaustive, but it further is not clear that an exhaustive compilation of all important and relevant material is even feasible. Nevertheless, the reviews summarized in Appendix 1 and additional analyses made in the following sections are broad enough to be informative.

Approaches for evaluating averaging periods for criteria

Two general and related approaches have been taken by investigators attempting to factor episodic disturbances as alternative test endpoints, risk assessment or into regulatory programs: experimental and predictive modeling approaches. In experimental approaches for example, if mortality is assumed to result from contamination of body tissues, accumulation at different points in time can be related to mortality (Gordon et al. 2012). Experimental approaches have explored contaminants accumulation and effects to biodynamic processes mediated by the life history and exposure routes of the organism, abiotic factors affecting bioavailability, and internal processes of uptake, regulation, detoxification, and excretion processes in time. These experimental approaches logically led to various predictive modeling approaches.

Predictive modeling approaches

As might be expected for efforts to build predictive models of complex and incompletely understood biodynamics and toxicokinetics, some modeling approaches can also become very complex. Models that seek to comprehensively mimic actual mechanisms become quite complex in order to reflect complex interactions (Jager et al. 2011). In contrast, other models seek to greatly simplify real processes in order to derive mathematical approximations that may be of practical use in environmental management situations. The latter approach is conceptually consistent with the 1985 Guidelines.

The simplest predictive models for evaluating time varying toxicity use traditional constant exposure data to predict toxicity under episodic exposure conditions. Mancini (1983) created a simple model using data obtained from conventional, constant exposure, lethal bioassays along with using toxicokinetic equations to predict toxicant concentration at the site of action within the organism as a function of exposure concentration and uptake/clearance rates. Under varying ambient concentrations, body tissue residue levels of toxicants (controlled by the rates of accumulation and depuration or repair by exposed organisms) predict the levels of biological response and vice versa. Because in reality, organisms do not simply accumulate toxins until they die, others expanded on this approach by adding damage and repair terms (Breck 1988; Landrum et al. 2004; Butcher et al. 2006). Erickson (2007) gives a detailed treatment and testing of different modeling approaches.

With metals, recent criteria development made considerable efforts to incorporate bioavailability adjustments with greater rigor than the long-standing hardness equations (USEPA 2007, 2018). Because biotic ligand models (BLMs) are based on chemical equilibria, their conceptual applicability and performance under nonequilibrium conditions such as episodic pollution events has been questioned (Hassler et al. 2004; Slaveykova and Wilkinson 2005; Meyer et al. 2007). While these issues have been explored to expand BLMs to include uptake/depuration/damage, the potential for models to become unwieldy or overly specific may limit their application. Meyer et al. (2007) modified the of BLM approach by altering the LA50 (lethal accumulation) sensitivity parameter in the Cu BLM to fit observed responses from pulsed exposures. They concluded that a single one-compartment uptake-depuration equation linked to a re-parameterized Cu BLM can be used to predict the acute toxicity of continuous and pulse exposures of Cu to fathead minnow larvae across a range of water quality conditions; but delayed deaths occurring after the pulse limited the accuracy of those predictions (Meyer et al. 2007). EPRI (2008) developed a Biotic Ligand Episodic Exposure Model, or BLEEM, for copper exposure to Daphnia magna. This effort combined a semi-empirical survival model developed by Butcher et al. (2006) with BLM software to produce a model capable of integrating the effects of water chemistry parameters, exposure concentration, exposure duration, and recovery duration on organism survival over time. The investigators had some success predicting the results of 21-day mortality testing with D. magna after short-term episodic exposures to copper.

Predictive models, however, have been criticized for not showing good enough fits with observed biological effects to allow for use in a regulatory setting (Gordon et al. 2012). For example, one of the more successful models, the kinetic model developed by Diamond et al. (2006) described only 50 to 60% of the variability observed in the survival of fathead minnow exposed separately to Cu and Zn, and *Daphnia magna* exposed to Cu (Gordon et al. 2012).

Speed of action calculations

Because of differing interpretations of the USEPA (1995) "*speed of action of metals acute toxicity to aquatic life*" materials, released without contextual explanations, the calculations were reconstructed and the results were explored here with several datasets. The approach builds from the Mancini model for determining the time period for episodic pollution effects to be considered independent (non-additive) between events.

Mancini (1983) proposed that if mortality of aquatic organisms was presumed to result from contaminant accumulation in the bodies reaching a critical threshold, then data from classical bioassay tests obtained using constant toxicant exposure concentrations could be used to infer uptake and depuration rates. Data obtained from lethal bioassays usually consists of tabulated values of the percent mortality observed at various times of exposure for several concentration levels of a contaminant. These data can be used to infer rate curves for uptake and depuration to reach steady state. The rate parameter, k, defines how steep or shallow the curves are. This rate, considered to be a detoxification rate, controls the time required for the organisms to reduce the internal level of the chemical to non-toxic conditions. Mancini (1983) considered an application of this approach to evaluate how long the period between exposures must be so that the effect of the previous exposure is reduced to nearly non-toxic conditions. From this he concluded that long times between exposures are needed to make two independent exposures, so that the toxicity responses of the first exposure do not add to the responses of the second exposure. This in turn, suggests effects of exposures from urban runoff, combined sewer discharges, frequent spills or continuous discharges may not be independent (Mancini 1983).

Erickson (2007) further developed these and related concepts for potentially incorporating time-varying toxicity into criteria formulations. The relationships of LC50s to exposure duration for aquatic organisms often follows an exponential decline from high values at short durations to a steady value ("asymptotic" or "threshold" LC50, or incipient lethal concentration) at long durations. Assuming death occurs when toxicant accumulation at critical sites reaches a lethal threshold, at short durations, high concentrations are needed to accumulate enough chemical fast enough to reach the lethal accumulation threshold quickly, in a manner that exceeds detoxification. As duration increases, lower water concentrations will cause mortality because there is more time for a chemical to accumulate. With even greater duration, accumulation approaches steady state, and the lethal water concentration will approach an asymptotic value. The rate at which this asymptotic lethal water concentration is approached will be equivalent to the rate at which steady-state accumulation is approached (Erickson 2007).

The "speed of action" equation (Equation 2) is a rearrangement of Mancini-type equations (Erickson 2007). Here, 1/k is used as a rule of thumb for limiting a pulse duration to not exceed the effects of an asymptotic LC50. It can be shown that if the LC50(∞) is the average LC50 over the period of a toxicity test, then an LC50 for a shorter pulse with duration "t" averages out to the LC50(∞)[2]. This is simply the ratio of LC50(t) to LC50(∞) times the pulse duration t (Equation 3). Then, a conservative P should be set based on a short, intense pulse, which would be its limit as t approaches 0.

Equation 2. $$LC50(t) = \frac{LC50(\infty)}{1-e^{-k_r t}}$$

Where:

LC50(t) = the LC50 calculated at a point in time during a toxicity test;
LC50(∞) = the asymptotic LC50, or the ultimate LC50 that could be attained if an acute toxicity test could be continued indefinitely. Synonymous with the Incipient Lethal Level (ILL) term used in toxicity testing;
k_r = Rate of detoxification over the (1/hours), and
t = time in hours

[2] Russ Erickson, USEPA Mid-Continent Ecology Division, written communication, March 22, 2018

Equation 3. $$P(t) = \frac{LC50(t)}{LC50(\infty)} \cdot t = \frac{t}{1 - e^{-k \cdot t}}$$

Thus a 1/k averaging period, when used with LC50(∞) will limit the worst-case exposure within that averaging period to no more than 50% mortality (assuming the model is a good approximation).

Some of these fundamental concepts are illustrated in Figure 9. In a conventional acute toxicity test, LC50s can be calculated for any time interval when counts of responding organisms. In this example, mortalities have occurred by 24 hours, but the responses at 48, 72, and 96 hours are nearly identical, indicating that the asymptotic LC50 was approached by 48 hours. The Mancini-type curve fits the observed data well (where the observed asymptotic LC50 of 0.97 µg/L is close to the calculated LC50(∞) of 0.87 µg/L). The 1/k "averaging period" of 32 hours is not a property that can be compared to observed data, but can roughly be thought of as a half-life to the asymptotic LC50(∞).

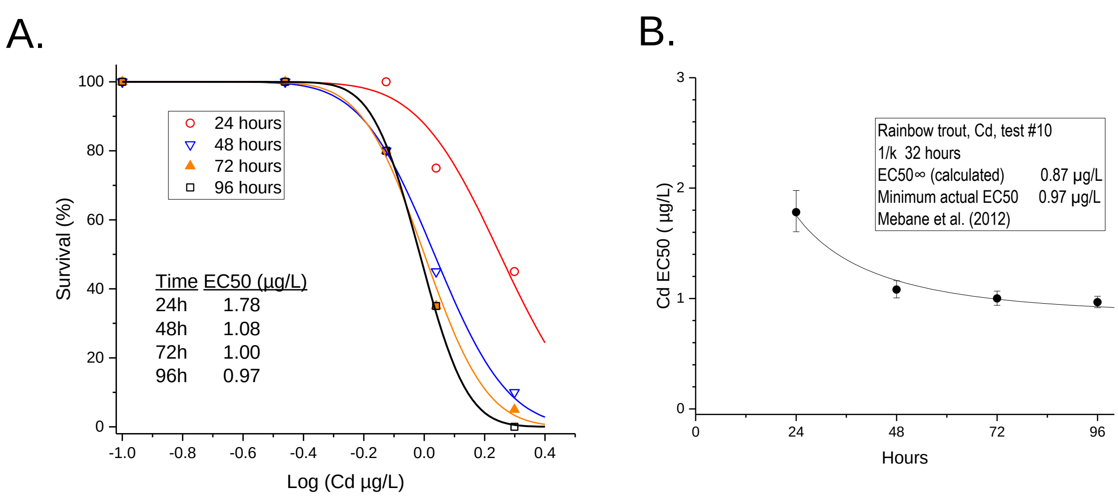


Figure 9. The acute toxicity of chemicals to aquatic organism is influenced by the magnitude and duration of exposure. In this example from a conventional acute toxicity test with cadmium and rainbow trout, mortalities had only begun to accrue by 24 hours, but by 48 hours most of the deaths that are going to occur by 96-hours have already occurred. Thus, the concentration response curves calculated at 48, 72, and 96 hours collapse on each other (A). When EC50s are plotted as a function of time, they approach an asymptotic, time-independent EC50, variously called the ILL (incipient lethal level) or EC50∞. This Mancini-type curve of time varying toxicity will be shallow for relatively fast-acting toxicants which reach their ultimate acute toxicity quickly. Steep curves reflect slower acting toxicants as deaths continue to accrue over time and EC50s continue to decline over the course of the test.

A variety of toxicity test results that reported varying toxicity over time were fit to the Mancini equation. The fits were achieved using the Frontline Solver optimization tool within Excel to optimize the two unknown variables k and LC50, using the GRG nonlinear solver engine to minimize the sum of squares for the differences between the fit and observed values. 1/k was constrained to be ≥1;however, solving this nonlinear equation using an iterative sum of squares minimization of residuals using actual data will bias the curve fit to better fit the higher LC50s calculated early in test periods such as the 24-h LC50s. Later LC50s such as the 96-h LC50s that are closer to the curve asymptote will produce inherently smaller residuals because there is less arithmetic space around a point along a curve as it approaches the asymptote. As a result, smaller LC50s at higher time periods will carry minimal influence in the curve fit relative to higher LC50s at lower time points. A log transformation may reduce this influence. Rearranging Eq. 2 with a logarithmic transformation results in Eq. 4.

Equation 4. $\log\left(LC50(t)\right) = \log(LC50\infty) - \log\left(1 - e^{-k_r t}\right)$

A variety of Mancini-type curve fits to Zn and sodium cyanide toxicity datasets are shown in Figure 10. In some cases, the relationships of EC50s to exposure duration did follow an exponential decline from high values at short durations to a steady value by the end of the test (sodium cyanide, figure 10.H). In other cases, the EC50 values over time did not stabilize by the end of the test, indicating that the 96-hour test was not long enough to obtain a time independent, asymptotic response (Figure 10.B, test #146; Figure 10D). In others, the responses over time were nearly flat, with most of the toxicity that ultimately resulted occurring within the first 24 hours of the test (Figure 10C). In that case, the 1/k calculated averaging periods of 25 – 33 hours were unexpected since most of the toxicity had occurred within the first 24-hour observation period. Finally, as several of these examples were from within the same study using rainbow trout or the closely related cutthroat trout, the remarkable variation in speed of action cautions that simple classifications of chemicals and organisms as sensitive or tolerant, or fast or slow to respond can be mistaken. The same factors that may affect toxicity/bioavailability (pH, temperature, hardness) also modify speed of action. This is illustrated well in Figure 10G where two tests of Zn with rainbow trout at different temperatures produced almost identical asymptotic LC50(∞) values, but the effects in the cooler test were slower.

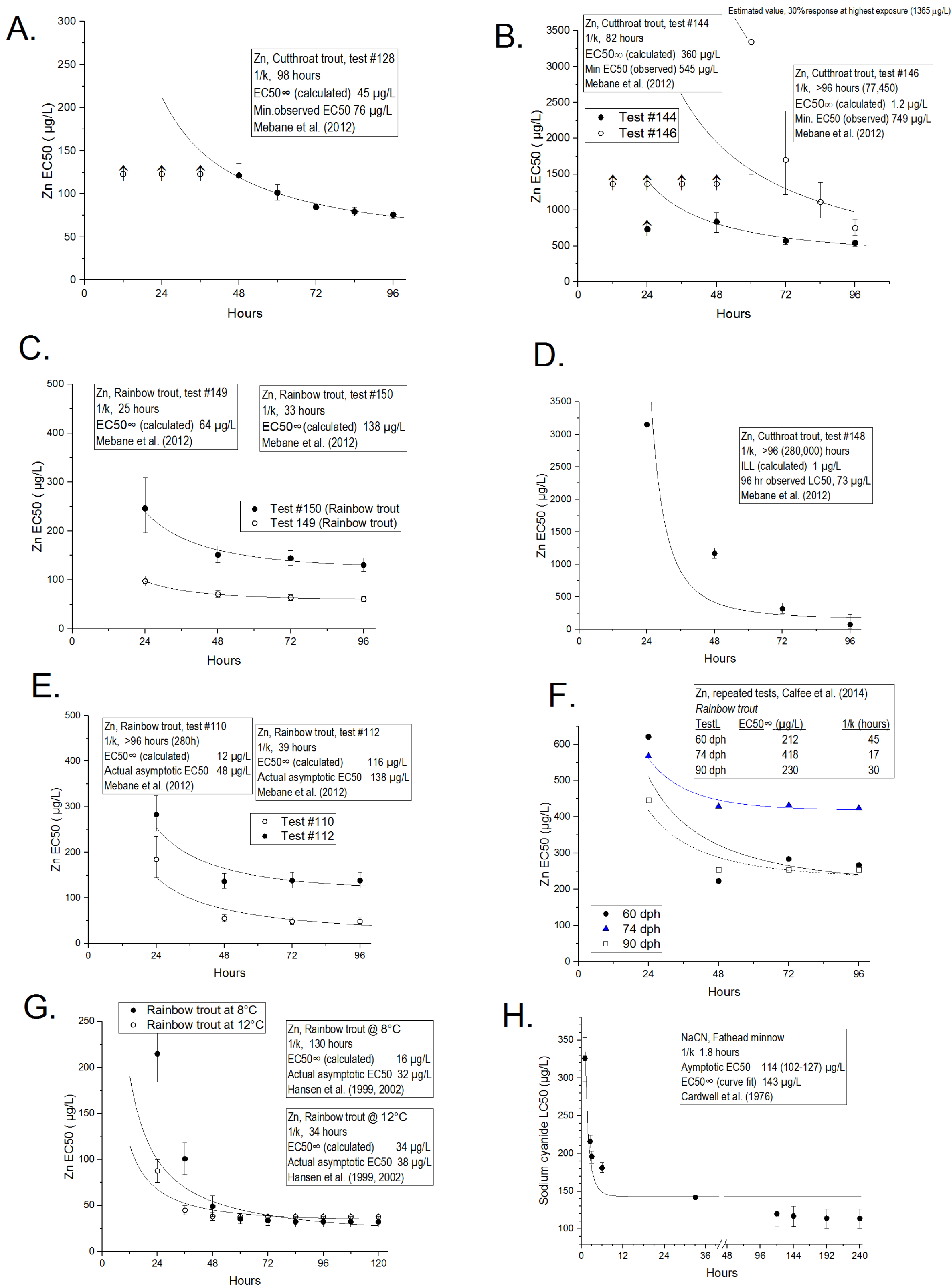


Figure 10. Selected Mancini-type speed of action curves of time varying toxicity: Zinc and trout (A-G) and sodium cyanide and fathead minnow (H). Upward arrow symbols indicate insufficient effects to calculate EC50s.

Similar plots and calculation results are shown for Cd, Cu, and pentachlorophenol (PCP) in Figure 11. The exponential decline in LC50s with increasing exposure time expected by the Mancini model was seen in the tests with PCP and Cu, which included more frequent observations than just every 24 hours (plots 11A,B,D). The cadmium tests with rainbow trout had mixed speed of action, with toxicity occurring rapidly in two tests (plots 12E,F in which most mortalities happened in the first 24 hours), but noticeably slower action in two others, with all the toxicity happening by 72 hours. As with Zn, the calculated 1/k averaging periods >24 hours for tests in which most mortality occurred in the first 24 hours were unexpected. These examples suggest that conventional toxicity testing results with fish which provide responses at 4 points in time (every 24 hours) are ill suited for Mancini- type modeling. If so, this could be a limitation to the approach, as these are the most commonly available time course of mortality data, and these were also likely the most common data types used in the previous analyses (USEPA 1995; Santore et al. 2016).

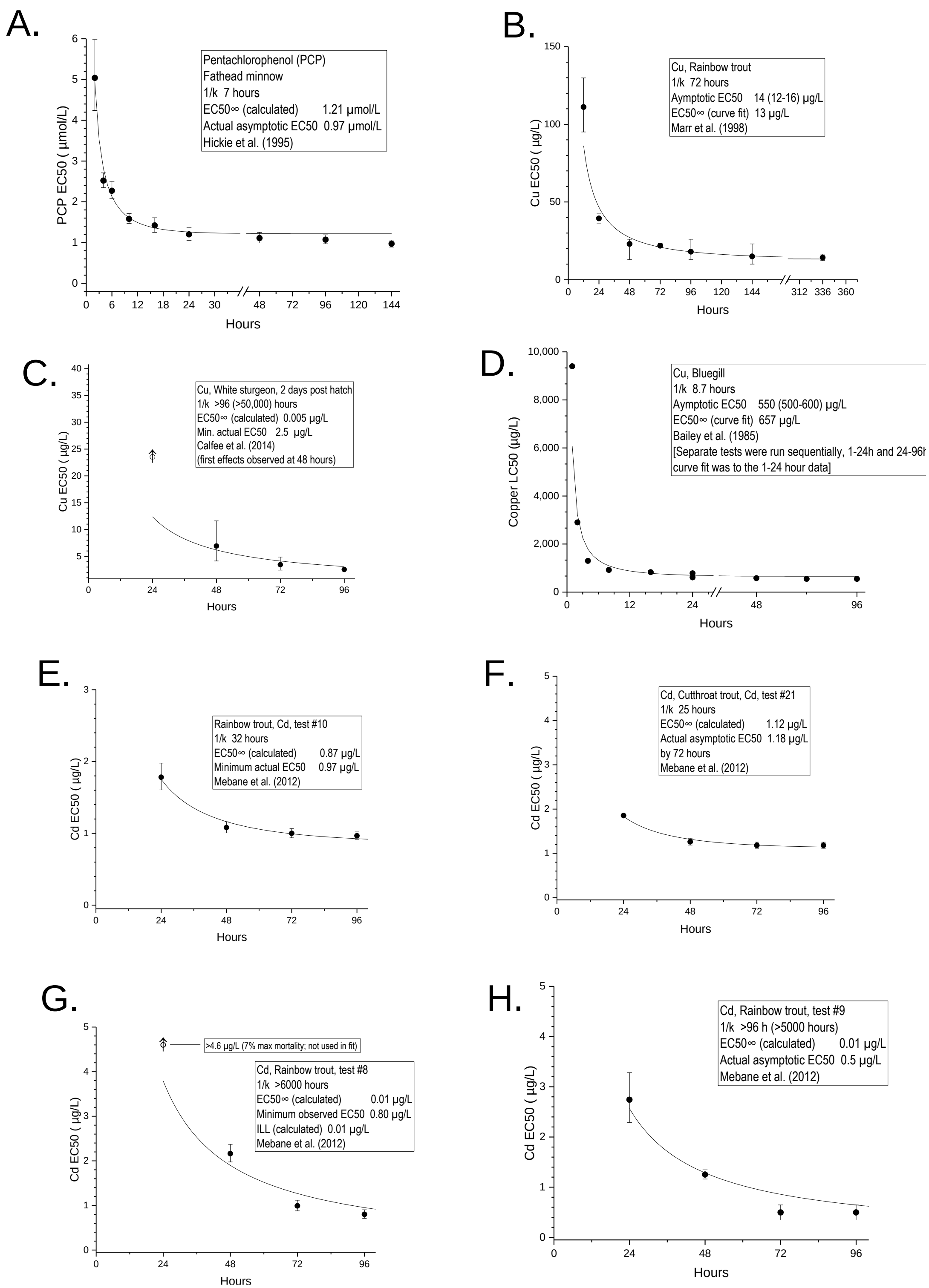


Figure 11. Selected Mancini-type speed of action curves of time varying toxicity: Cadmium, copper, and an organic toxicant (pentachlorophenol)

The final set of data reanalyzed through the Mancini speed of action model were pulse exposure tests by Brent and Herricks (1998) with Zn, *Ceriodaphnia dubia*, and *Hyalella azteca*. The organisms were briefly exposed to high concentrations of Zn, Cd, or phenol for up to 4 hours and then observed out to 144 hours in clean test waters (only Zn results were worked up and shown in Figure 12). These results also show the near exponential decline in LC50 values expected with increasing exposure time. However, almost all of the "exposure" time is actually post-exposure observation time when no metals were present and delayed mortality accounted for almost all of the responses. The 1/k averaging periods of about 3 hours for two of the *Ceriodaphnia* exposures were among the shortest calculated.

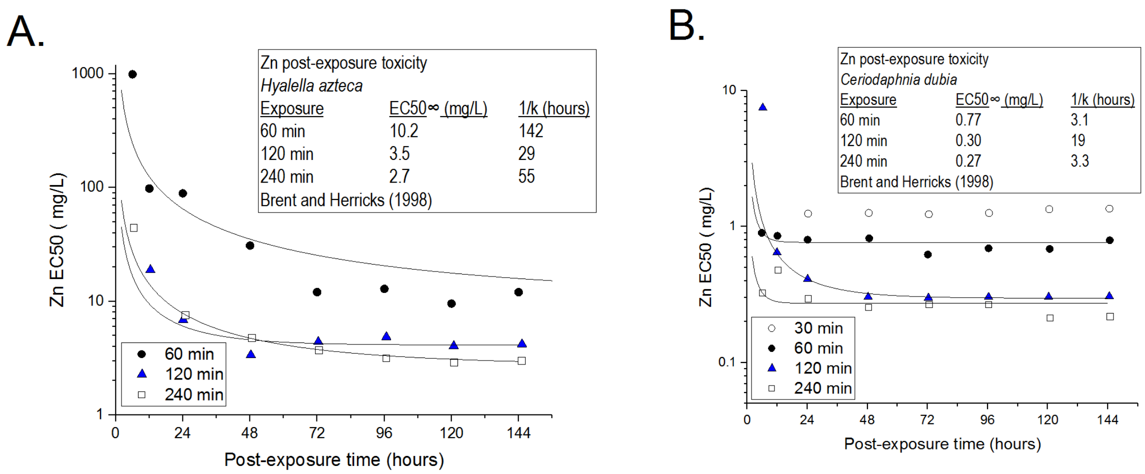


Figure 12. Mancini-type speed of action curves of time varying toxicity applied to latent toxicity following short pulsed exposures. Almost all effects are from delayed mortality, occurring after the animals were transferred to clean water. The strong latent toxicity cautions against sole reliance on Mancini-type modeling to evaluate averaging periods, as results can be misleadingly reassuring.

This apparent rapid speed of action resulting from post-pulse exposures suggests that further thought needs to be given to the potential of the use of 1/k statistic as a quantitative guide when considering appropriate averaging periods for criteria. The initial intent of USEPA's explorations of the use of the 1/k value from Mancini modeling as a rule-of-thumb/approximation for the acute averaging period for aquatic life criteria was restricted to situations with moderate-to-fast kinetics for lethality. In such situations, the LC50 used for criterion development is not that much greater than the threshold LC50, that is, the 48- or 96-hour concentrations are close to the asymptotic LC50. This limitation was justified, because the intent at the time was to determine if the averaging period could be more than 1 hour, but not more than 1 day. So slower kinetics with 1/k>1 day, where the 1/k approximation breaks down, were irrelevant. This constraint has not been recognized over the years.[3] The present analyses also violate that constraint. Were the present analysis to have been constrained to only those data found that had multiple observations over the first 24 hours of testing to get good curve fits in that

[3] Russ Erickson, USEPA Mid-Continent Ecology Division, written communication, March 22, 2018

part of the test responses, it would have been a much shorter analysis of speed of action calculations. Overall, the 1/k statistic can roughly be thought of as a half-life for the test populations to reach their time-independent LC50∞. The 1/k "characteristic time of effects" metric ranged from 1.8 to > 96 hours. Most of the examples produced 1/k values of >24 hours.

The delayed mortality problem

Erickson (2007) cautioned that one simplistic assumption in the single-compartment, lethal-accumulation-threshold model is that an organism will die immediately upon reaching a lethal accumulation threshold, but survive indefinite exposures just below the threshold. More realistically, once chemical is accumulated, any overt expression of toxicity involves a series of biochemical reactions with kinetic constraints that might affect time-to-death as much as, if not more than, accumulation kinetics. Some organisms would be expected die after damage occurs over time, or else recover.

The rationale for extending Mancini's (1983) 1/k statistic to inform criteria averaging times assumes that toxicity, at a given point in time, depends on both the current and the past exposure concentrations. Mancini's single-compartment, lethal-accumulation-threshold model assumes that an organism will die immediately upon reaching a lethal accumulation threshold. Delayed mortalities are ignored, or more precisely, are included in the tally for the next time period. This mis-tallying is unimportant in conventional uses of end-of-test LC50s where ending LC values are calculated using the cumulative mortalities. Such mis-tallying is particularly unimportant late in toxicity tests after mortalities have slowed or stopped. For example, most of the test results shown in Figures 7 to 10 indicate the asymptotic LC50s have been achieved by 72 hours, that is, effects are unchanged between 72 and 96 hours. However, the shape of the curves with declining LC50s over time and the calculated 1/k averaging periods is sensitive to responses in the early stages of the test when organisms begin to die. Curve-shape sensitivity to early time periods is an important implication for tests with moderate to fast kinetics such that the 1/k averaging period is < 1 day, as it is these shorter, pulse exposures when delayed mortality is likely to be an important part of test interpretation and can confound the 1/k averaging period concept and calculations.

Examples of delayed mortality following brief toxicity testing follow. In Figure 13, fathead minnow mortalities are contrasted following either conventional, continuous exposures of Cu for up to 144 hours or pulse exposures of Cu for 3 to 18 hours, and then observed for delayed mortality for at least 200 hours (Zahner 2009). Two important features of these data are, first, the eventual effects from brief pulse exposures were about as severe as were the continuous exposures. For example, a continuous exposure to Cu at 50 µg/L at 96 hours resulted in about 40% mortality. Pulse exposures of fish to 50 µg/L Cu for 9, 12, or 18 hours, and then transferred to clean water and had nearly the same effects, with 40-50% mortality accruing by 96 hours, and a 3-hour pulse resulted in about 30% mortality by 96 hours. Second, could these data been analyzed through the speed of action equation, the 1/k value would have been >48 hours and probably >96 hours since the mortalities had not fully leveled out by then. Yet in the parallel pulse exposures, a 3-hour pulse was sufficient to produce substantial mortalities.

In marked contrast, other toxicity tests with Cu and fathead minnow exposed to brief exposures of 2 to 24 hours and then transferred to clean water and observed for at least 96 hours produced little delayed mortality following exposures of ≥12 hours (Erickson 2007). In the shortest exposure (2 hours), LC50s calculated from mortalities tallied both during and after the exposures were less than half the LC50s calculated from mortalities that occurred only during the 2-hour exposures. These differences declined with increasing pulse length, such that there were almost no differences in the 24-hour pulses (Figure 14A).

In a final example with Cu (Zhao and Newman 2004), *Hyalella azteca* amphipods were exposed to Cu for 48 hours, and then observed to about 120 hours. In 2 of the 3 repeated experiments, mortalities continued to increase after the exposure had ended (Figure14B, top and middle). In the third experiment (bottom), however, there was some delayed mortality, but it was not nearly as pronounced as in the top and middle plots (Figure 14B). In related work, amphipods exposed to Cu for 20 hours and then followed out to 60 hours, had nearly as much mortality at 60 hours as did 60-hour continuous exposures (Zhao and Newman 2006).

In very recent work, a series of tests was designed in part to see if the work of Brent and Herricks (1998) would replicate. That work had found that brief exposures of as little as 30 minutes could result in mortalities days later, even though no mortalities occurred during the actual exposures. In the recent work (Figure 15), *Ceriodaphnia dubia* were exposed to Zn for different periods ranging from 1 hour to 48 hours, and then were transferred to clean water and observed for post-exposure mortality for up to 48 hours. For example, for the 1-hour exposure, animals would be exposed for 1 hour and then transferred for a 47-hour recovery, and 3-hour exposures would be transferred and followed for a 45-hour recovery period, and so on. Continuously exposed animals served as positive controls. Rainbow trout were similarly tested, although the full exposure and recovery test duration was generally 96-hours for the trout. The results were consistent with those of Brent and Herricks (1998), who found that brief (1 and 3-hour) exposures killed most or all *Ceriodaphnia* across ranges of tested concentrations (Figure 12). Rainbow trout were much more resilient to short-term Zn exposures than *Ceriodaphnia*, surviving up to 3-hour exposures to >10 mg/L Zn with low ultimate mortalities. However, fish exposed to 2000 µg/L for 8-hours began to die by 16 hours and suffered high (>50%) mortality by 36 hours (Ivey and Mebane 2019).

Taken together, the examples of delayed mortality from metals exposures show that delayed mortality can be an important factor in interpreting toxicity from pulse or episodic exposures. If delayed mortality is not accounted for in simplistic speed of action calculations, such as those shown in Figures 9-11, the calculations can produce highly misleading results. Figure 14A gives an example of accounting for delayed mortality in the presentation and interpretation of short-duration toxicity testing. Finally, the disparate examples caution against sole reliance on any one dataset in evaluating the importance of delayed mortality.

A summary and list of overall conclusions drawn from this review were listed at the front of this document.

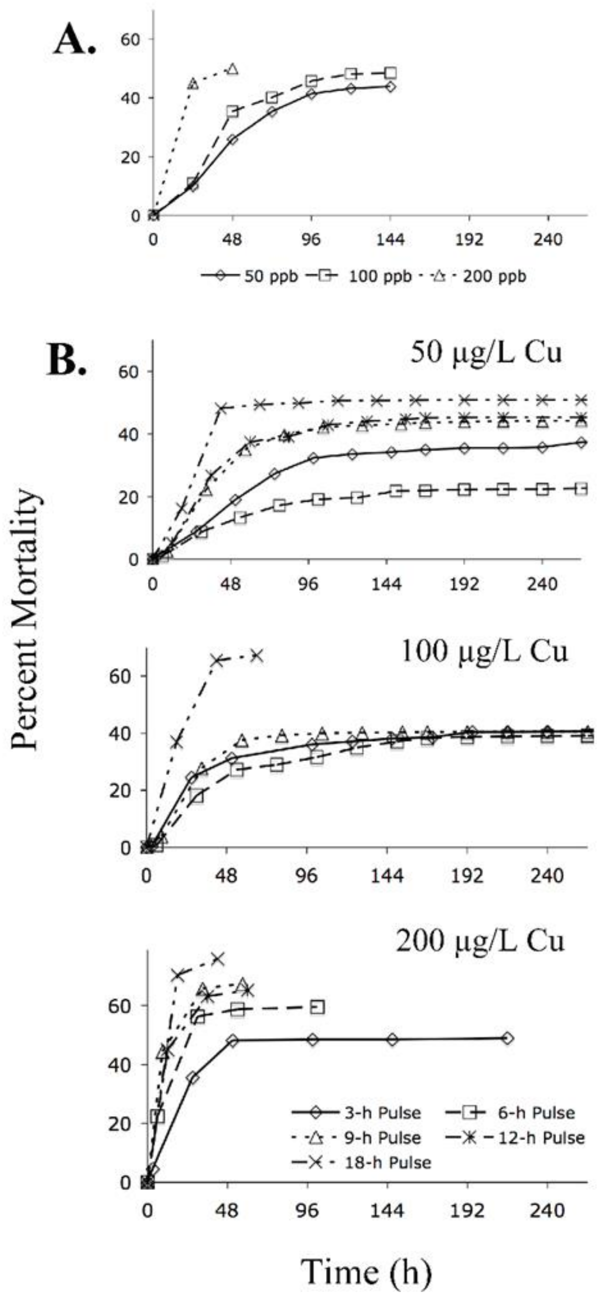


Figure 13. Examples of delayed mortality following exposure of fathead minnows to copper in (A) conventional, continuous exposures; (B) single pulse exposures of 3 to 18 hours 50, 100, or 200 µg/L. For example, mortalities following 3-h pulses at 200 and 50 µg/L continued to accrue until about 48 or 96 hours respectively. Thus, the onset of effects lags the exposures. These sorts of delayed or latent effects are seldom addressed in acute toxicity testing and were largely ignored in speed of action illustrations in Figures 11 and 12. The figure is taken from Zahner (2009).

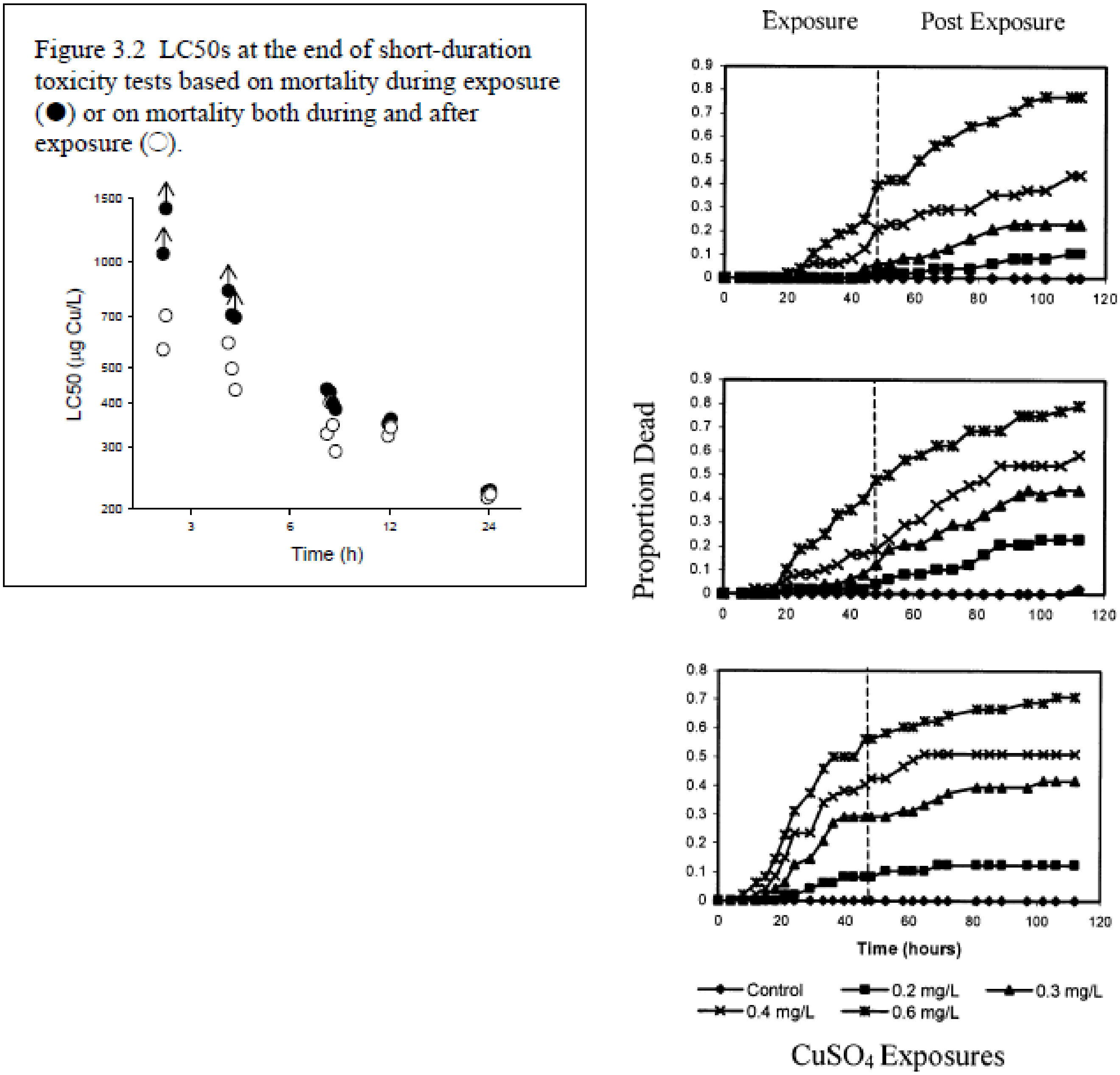


Figure 14. Examples of delayed mortality following exposure of (**A**) fathead minnows to copper and (B) *Hyalella azteca* to copper. With this fathead minnow testing, shorter copper pulses of up to 12h produced mortalities that continued to accrue after the exposures ended, but with 24h exposures, all mortalities that were going to occur did occur within the first 24h; there was no delayed mortality. The figure is taken from Erickson (2007).

With the *Hyalella* testing (**B**), in 2 of 3 repeated experiments, cumulative mortalities continued to accrue after the 48h copper exposures ended (dashed lines). Similar testing with an organic toxicant (PCP) showed almost no delayed mortality. The figure is taken from Zhao and Newman (2004).

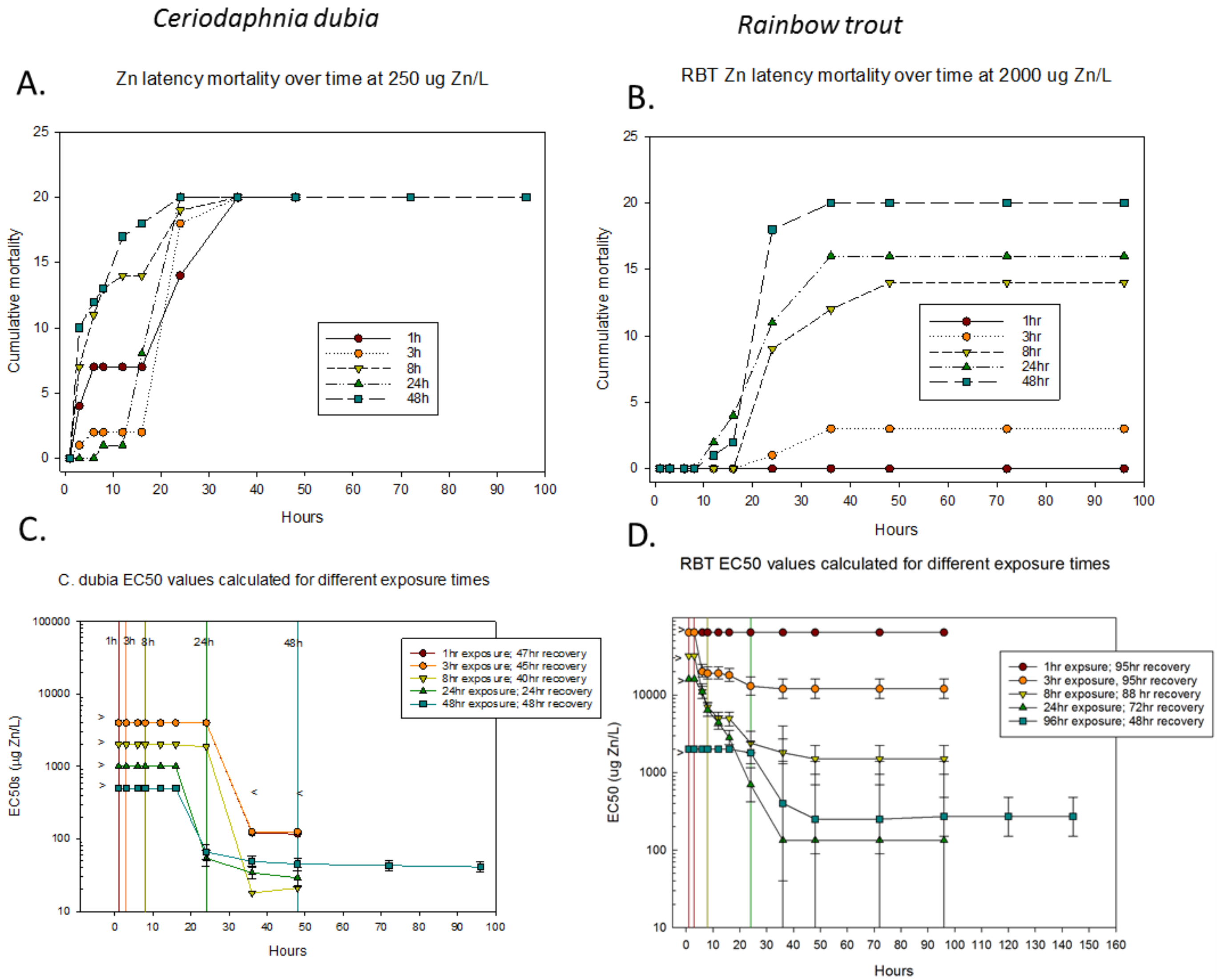


Figure 15. Delayed mortality of *Ceriodaphnia dubia* and rainbow trout to zinc with various exposure durations and recovery periods. Organisms were transferred to clean water after exposures and further monitored. For example, *C. dubia* were exposed to Zn for 1-hour and then survival in clean water was followed for 47 hours, and so on. Rainbow trout were similarly exposed to Zn for 1 hour, transferred to clean water and followed for 95 hours, and so on. *C. dubia* were highly susceptible to delayed mortality. For example, 1-hour exposures to 250 µg/L Zn killed no *C. dubia* during the 1-hour actual exposure period, but after 35 hours of "recovery time" in clean water, 20 of 20 had died. Rainbow trout were much more resilient to short-term Zn exposures, surviving up to 3-hour exposures at very high concentrations with low mortalities. However, fish exposed to 2000 µg/L for 8-hours began to die by 16 hours and suffered high (>50%) mortality by 36 hours. Data from Ivey and Mebane (2019)

References

Alonso, Á. and J.A. Camargo. 2009. Effects of pulse duration and post-exposure period on the nitrite toxicity to a freshwater amphipod. *Ecotoxicology and Environmental Safety*. 72(7): 2005-2008. https://doi.org/10.1016/j.ecoenv.2009.06.008

Angel, B.M., S.L. Simpson, E. Granger, K. Goodwyn, and D.F. Jolley. 2017. Time-averaged concentrations are effective for predicting chronic toxicity of varying copper pulse exposures for two freshwater green algae species. *Environmental Pollution*. 230(Supplement C): 787-797. https://doi.org/10.1016/j.envpol.2017.07.013

Angel, B.M., S.L. Simpson, and D.F. Jolley. 2010. Toxicity to *Melita plumulosa* from intermittent and continuous exposures to dissolved copper. *Environmental Toxicology and Chemistry*. 29(12): 2823-2830. https://doi.org/10.1002/etc.347

Baldwin, D.H., J.F. Sandahl, J.S. Labenia, and N.L. Scholz. 2003. Sublethal effects of copper on coho salmon: impacts on nonoverlapping receptor pathways in the peripheral olfactory nervous system. *Environmental Toxicology and Chemistry*. 22(10): 2266-2274. https://doi.org/10.1897/02-428

Balistrieri, L.S. and R.G. Blank. 2008. Dissolved and labile concentrations of Cd, Cu, Pb, and Zn in the South Fork Coeur d'Alene River, Idaho: Comparisons among chemical equilibrium models and implications for biotic ligand models. *Applied Geochemistry*. 23(12): 3355-3371. https://doi.org/10.1016/j.apgeochem.2008.06.031

Balistrieri, L.S., D.A. Nimick, and C.A. Mebane. 2012. Assessing time-integrated dissolved concentrations and predicting toxicity of metals during diel cycling in streams. *Science of the Total Environment*. 425: 155–168. https://doi.org/10.1016/j.scitotenv.2012.03.008

Bearr, J.S., J.M. Diamond, H. Latimer, and M. Bowersox. 2006. Effects of pulsed copper exposures on early life-stage *Pimephales promelas*. *Environmental Toxicology and Chemistry*. 25(5): 1376-1382. https://doi.org/10.1897/05-259R.1

Blaen, P.J., K. Khamis, C.E.M. Lloyd, C. Bradley, D. Hannah, and S. Krause. 2016. Real-time monitoring of nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future directions. *Science of the Total Environment*. 569-570: 647-660. https://doi.org/10.1016/j.scitotenv.2016.06.116

Boyer, E.W., G.M. Hornberger, K.E. Bencala, and D.M. McKnight. 2000. Effects of asynchronous snowmelt on flushing of dissolved organic carbon: a mixing model approach. *Hydrological Processes*. 14(18): 3291 - 3308. https://doi.org/10.1002/1099-1085(20001230)14:18<3291::AID-HYP202>3.0.CO;2-2

Breck, J.E. 1988. Relationships among models for acute toxic effects: Applications to fluctuating concentrations. *Environmental Toxicology and Chemistry*. 7(9): 775-778. https://doi.org/10.1002/etc.5620070911

Brent, R.N. and E.E. Herricks. 1998. Postexposure effects of brief cadmium, zinc, and phenol exposures on freshwater organisms. *Environmental Toxicology and Chemistry*. 17(10): 2091–2099. https://doi.org/10.1002/etc.5620171027

Brette, F., H.A. Shiels, G.L.J. Galli, C. Cros, J.P. Incardona, N.L. Scholz, and B.A. Block. 2017. A Novel Cardiotoxic Mechanism for a Pervasive Global Pollutant. *Scientific Reports*. 7: 41476. https://doi.org/10.1038/srep41476

Brooks, M.L., C.J. Boese, and J.S. Meyer. 2006. Complexation and time-dependent accumulation of copper by larval fathead minnows (*Pimephales promelas*): Implications for modeling toxicity. *Aquatic Toxicology*. 78(1): 42-49. https://doi.org/10.1016/j.aquatox.2006.02.002

Burton, G.A., R. Pitt, and S. Clark. 2000. The Role of Traditional and Novel Toxicity Test Methods in Assessing Stormwater and Sediment Contamination. *Critical Reviews in Environmental Science and Technology*. 30(4): 413-447. https://doi.org/10.1080/10643380091184228

Butcher, J.B., J.M. Diamond, J. Bearr, H. Latimer, S.J. Klaine, T.C. Hoang, and M. Bowersox. 2006. Toxicity models of pulsed copper exposure to *Pimephales promelas* and *Daphnia magna*. *Environmental Toxicology and Chemistry*. 25(9): 2541-2550. https://doi.org/10.1897/05-630R.1

Chapin, T.P. 2015. High-frequency, long-duration water sampling in acid mine drainage studies: A short review of current methods and recent advances in automated water samplers. *Applied Geochemistry*. 59: 118–124. https://doi.org/10.1016/j.apgeochem.2015.04.004

Chapman, G.A. 1994. Data on effects of chronic copper exposures with steelhead - acclimation, life stage differences, and behavioral effects. Letter of July 5, 1994 to Chris Mebane, [NOAA liaison to] EPA Region X, Seattle, Wash. U.S. EPA Coastal Ecosystems Team, Newport, Oregon.

Characklis, G.W. and M.R. Wiesner. 1997. Particles, Metals, and Water Quality in Runoff from Large Urban Watershed. *Journal of Environmental Engineering*. 123(8): 753-759. https://doi.org/10.1061/(ASCE)0733-9372(1997)123:8(753)

Cherry, D.S., S.R. Larrick, J.D. Giattina, K.L. Dickson, and J. Cairns. 1979. Avoidance and toxicity responses of fish to intermittent chlorination. *Environment International*. 2(2): 85-90. https://doi.org/https://doi.org/10.1016/0160-4120(79)90046-1

Clark, G.M. and C.A. Mebane. 2014. Sources, Transport, and Trends for Selected Trace Metals and Nutrients in the Coeur d'Alene and Spokane River Basins, Northern Idaho, 1990–2013. U.S. Geological Survey Scientific Investigations Report 2014-5204. http://dx.doi.org/10.3133/sir20145204.

Clark, M.J. and M.W. Brooks. 1989. Aquatic toxicology of the pyrethroid insecticides: Neurotoxicology of pyrethroids: Single or multiple mechanisms of action? *Environmental Toxicology and Chemistry*. 8(5): 361-372. https://doi.org/doi:10.1002/etc.5620080502

Clearwater, S.J., P.G. Jellyman, B.J.F. Biggs, C.W. Hickey, N. Blair, and J.S. Clayton. 2011. Pulse-dose application of chelated copper to a river for *Didymosphenia geminata* control: Effects on macroinvertebrates and fish. *Environmental Toxicology and Chemistry*. 30(1): 181-195. https://doi.org/10.1002/etc.369

Coats, J.R., D.M. Symonik, S.P. Bradbury, S.D. Dyer, L.K. Timson, and G.J. Atchison. 1989. Toxicology of synthetic pyrethroids in aquatic organisms: An overview. *Environmental Toxicology and Chemistry*. 8(8): 671-679. https://doi.org/doi:10.1002/etc.5620080805

Corsi, S.R., D.J. Graczyk, S.W. Geis, N.L. Booth, and K.D. Richards. 2010. A fresh look at road salt: Aquatic toxicity and water-quality impacts on local, regional, and national scales. *Environmental Science & Technology*. 44(19): 7376-7382. https://doi.org/10.1021/es101333u

Crumpton, W.G. and C.M. Hersh. 1987. Diurnal patterns of ammonium loading and loss in a stream receiving secondary treatment effluent. *Journal (Water Pollution Control Federation)*. 59(1): 60-62

Davies, P.E. and L.S.J. Cooke. 1993. Catastrophic macroinvertebrate drift and sublethal effects on brown trout, *Salmo trutta*, caused by cypermethrin spraying on a Tasmanian stream. *Aquatic Toxicology*. 27(3-4): 201-224. https://doi.org/10.1016/0166-445X(93)90055-6

Diamond, J.M., M. Bowersox, H. Latimer, C. Barbour, J. Bearr, and J.B. Butcher. 2005. Effects of pulsed contaminant exposures on early life stages of the fathead minnow. *Archives of Environmental Contamination and Toxicology*. 49(4): 511-519. https://doi.org/10.1007/s00244-005-7023-8

Diamond, J.M., S.J. Klaine, and J.B. Butcher. 2006. Implications of pulsed chemical exposures for aquatic life criteria and wastewater permit limits. *Environmental Science and Technology*. 40(16): 5132-5138. https://doi.org/10.1021/es0604358

Dittman, J.A., J.B. Shanley, C.T. Driscoll, G.R. Aiken, A.T. Chalmers, and J.E. Towse. 2009. Ultraviolet absorbance as a proxy for total dissolved mercury in streams. *Environmental Pollution*. 157(6): 1953-1956 https://doi.org/10.1016/j.envpol.2009.01.031

Duncan, D.A. and J.F. Klaverkamp. 1983. Tolerance and resistance to cadmium in White Suckers (*Catostomus commersoni*) previously exposed to cadmium, mercury, zinc, or selenium. *Canadian Journal of Fisheries and Aquatic Sciences*. 40(2): 128-138. https://doi.org/10.1139/f83-022

EPRI. 2008. Development of an Integrated Episodic Exposure and Bioavailability Model. Electric Power Research Institute 1014065, Palo Alto, CA USA. 50 pp. https://www.epri.com/research/products/1014065 [Accessed March 1, 2018].

Erickson, R.J. 2007. Quantification of Toxic Effects for Water Concentration-Based Aquatic Life Criteria, Part A., EPA/600/R-07/065, U.S. Environmental Protection Agency, Duluth, Minnesota. 78 pp. https://archive.epa.gov/med/med_archive_02/web/pdf/600r07065.pdf [Accessed 23 April 2020].

Etheridge, A.B. 2015. Occurrence and Transport of Selected Constituents in Streams near the Stibnite Mining Area, Central Idaho, 2012–14. U.S. Geological Survey Scientific Investigations Report 2015-5203. 47 pp. https://pubs.er.usgs.gov/publication/sir20155166.

Fisher, D.J., D.T. Burton, L.T. Yonkos, S.D. Turley, B.S. Turley, G.P. Ziegler, and E.J. Zillioux. 1994. Acute and short-term chronic effects of continuous and intermittent chlorination on *Mysidopsis bahia* and *Menidia beryllina*. *Environmental Toxicology and Chemistry*. 13(9): 1525-1534. https://doi.org/10.1002/etc.5620130915

Friberg-Jensen, U., L. Wendt-Rasch, P. Woin, and K. Christoffersen. 2003. Effects of the pyrethroid insecticide, cypermethrin, on a freshwater community studied under field conditions. I. Direct and indirect effects on abundance measures of organisms at different trophic levels. *Aquatic Toxicology*. 63(4): 357-371. https://doi.org/10.1016/S0166-445X(02)00201-1

Gama-Flores, J.L., S.S.S. Sarma, and S. Nandini. 2007. Exposure time-dependent cadmium toxicity to *Moina macrocopa* (Cladocera): a life table demographic study. *Aquatic Ecology*. 41(4): 639-648. https://doi.org/10.1007/s10452-007-9121-3

Gammons, C.H., J.N. Babcock, S.R. Parker, and S.R. Poulson. 2011. Diel cycling and stable isotopes of dissolved oxygen, dissolved inorganic carbon, and nitrogenous species in a stream receiving treated municipal sewage. *Chemical Geology*. 283(1): 44-55. https://doi.org/10.1016/j.chemgeo.2010.07.006

Goldstone, M.E., P.W.W. Kirk, and J.N. Lester. 1990a. The behaviour of heavy metals during wastewater treatment I. Cadmium, chromium and copper. *Science of the Total Environment*. 95: 233-252. https://doi.org/10.1016/0048-9697(90)90068-6

Goldstone, M.E., P.W.W. Kirk, and J.N. Lester. 1990b. The behaviour of heavy metals during wastewater treatment II. Lead, nickel and zinc. *Science of the Total Environment*. 95: 253-270. https://doi.org/10.1016/0048-9697(90)90069-7

Gordon, A.K., S.K. Mantel, and N.W.J. Muller. 2012. Review of toxicological effects caused by episodic stressor exposure. *Environmental Toxicology and Chemistry*. 31(5): 1169-1174. https://doi.org/10.1002/etc.1781

Hansen, J.A., J.C.A. Marr, J. Lipton, and H.L. Bergman. 1999a. Differences in neurobehavioral responses of chinook salmon *(Oncorhynchus tshawytscha)* and rainbow trout *(Oncorhynchus* *mykiss)* exposed to copper and cobalt: behavioral avoidance. *Environmental Toxicology and Chemistry*. 18(9): 1972-1978. https://doi.org/10.1002/etc.5620180916

Hansen, J.A., J.D. Rose, R.A. Jenkins, K.G. Gerow, and H.L. Bergman. 1999b. Chinook salmon (*Oncorhynchus tshawytscha*) and rainbow trout (*Oncorhynchus mykiss*) exposed to copper: neurophysiological and histological effects on the olfactory system. *Environmental Toxicology and Chemistry*. 18(9): 1979-1991. https://doi.org/10.1002/etc.5620180917

Hansen, J.A., P.G. Welsh, J. Lipton, D. Cacela, and A.D. Dailey. 2002. Relative sensitivity of bull trout *(Salvelinus confluentus)* and rainbow trout *(Oncorhynchus mykiss)* to acute exposures of cadmium and zinc. *Environmental Toxicology and Chemistry*. 21(1): 67–75. https://doi.org/10.1002/etc.5620210110

Hassler, C.S., V.I. Slaveykova, and K.J. Wilkinson. 2004. Some fundamental (and often overlooked) considerations underlying the free ion activity and biotic ligand models. *Environmental Toxicology and Chemistry*. 23: 283–291. https://doi.org/10.1897/03-149

Hoang, T.C. and S.J. Klaine. 2007. Influence of organism age on metal toxicity to *Daphnia magna*. *Environmental Toxicology and Chemistry*. 26(6): 1198-1204. https://doi.org/10.1897/06-515r.1

Hogan, A.C., M.A. Trenfield, A.J. Harford, and R.A. van Dam. 2013. Toxicity of magnesium pulses to tropical freshwater species and the development of a duration-based water quality guideline. *Environmental Toxicology and Chemistry*. 32(9): 1969-1980. https://doi.org/10.1002/etc.2251

Holdway, D.A., K. Lok, and M. Semaan. 2001. The acute and chronic toxicity of cadmium and zinc to two hydra species. *Environmental Toxicology*. 16(6): 557-565. https://doi.org/10.1002/tox.10017

Incardona, J.P. and N.L. Scholz. 2017. Environmental Pollution and the Fish Heart. Pages 373-433 *in* A. K. Gamperl, T. E. Gillis, A. P. Farrell, and C. J. Brauner, editors. *Fish Physiology*, volume 36. Academic Press, http://www.sciencedirect.com/science/article/pii/S1546509817300316.

Ivey, C.D. and C.A. Mebane. 2019. Acute and latent effects of zinc on two commonly tested species (*Ceriodaphnia dubia* and *Oncorhynchus mykiss*). *U.S. Geological Survey Data Release*. https://doi.org/10.5066/P9X4HH4T.

Jager, T., C. Albert, T.G. Preuss, and R. Ashauer. 2011. General Unified Threshold model for Survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science & Technology*. 45(7): 2529-2540. https://doi.org/10.1021/es103092a

Kayhanian, M., C. Stransky, S. Bay, S.L. Lau, and M.K. Stenstrom. 2008. Toxicity of urban highway runoff with respect to storm duration. *Science of the Total Environment*. 389(2): 386-406. https://doi.org/10.1016/j.scitotenv.2007.08.052

Kuivila, K.M. and C.G. Foe. 1995. Concentrations, transport and biological effects of dormant spray pesticides in the San Francisco Estuary, California. *Environmental Toxicology and Chemistry*. 14(7): 1141-1150. https://doi.org/doi:10.1002/etc.5620140704

Laetz, C.A., D.H. Baldwin, T.K. Collier, V. Hebert, J.D. Stark, and N.L. Scholz. 2009. The synergistic toxicity of pesticide mixtures: implications for risk assessment and the conservation of endangered Pacific salmon. *Environmental Health Perspectives*. 117(3): 348-353. https://doi.org/10.1289/ehp.0800096

Landrum, P.F., J.A. Steevens, D.C. Gossiaux, M. McElroy, S. Robinson, L. Begnoche, S. Chernyak, and J. Hickey. 2004. Time-dependent lethal body residues for the toxicity of pentachlorobenzene to *Hyalella azteca*. *Environmental Toxicology and Chemistry*. 23(5): 1335-1343. https://doi.org/10.1897/03-164

Lawler, D.M., G.E. Petts, I.D.L. Foster, and S. Harper. 2006. Turbidity dynamics during spring storm events in an urban headwater river system: The Upper Tame, West Midlands, UK. *Science of the*

Total Environment. 360(1): 109-126. https://doi.org/https://doi.org/10.1016/j.scitotenv.2005.08.032
Lee, H., S.-L. Lau, M. Kayhanian, and M.K. Stenstrom. 2004. Seasonal first flush phenomenon of urban stormwater discharges. *Water Research*. 38(19): 4153-4163. https://doi.org/10.1016/j.watres.2004.07.012
Lewis, W.M., Jr. and M.C. Grant. 1979. Relationships between stream discharge and yield of dissolved substances from a Colorado mountain watershed. *Soil Science*. 128(6): 353-363
Ma, H., S.D. Kim, D.K. Cha, and H.E. Allen. 1999. Effects of kinetics of complexation by humic acid on toxicity of copper to *Ceriodaphnia dubia*. *Environmental Toxicology and Chemistry*. 18(5): 828-837. https://doi.org/10.1002/etc.5620180504
Makepeace, D.K., D.W. Smith, and S.J. Stanley. 1995. Urban stormwater quality: Summary of contaminant data. *Critical Reviews in Environmental Science and Technology*. 25(2): 93-139. https://doi.org/10.1080/10643389509388476
Maltby, L. 1995. Sensitivity of the crustaceans *Gammarus pulex* (L.) and *Asellus aquaticus* (L.) to short-term exposure to hypoxia and unionized ammonia: Observations and possible mechanisms. *Water Research*. 29(3): 781-787. https://doi.org/10.1016/0043-1354(94)00231-U
Mancini, J.L. 1983. A method for calculating effects, on aquatic organisms, of time varying concentrations. *Water Research*. 17(10): 1355-1362. https://doi.org/10.1016/0043-1354(83)90264-6
Marr, J.C.A., H.L. Bergman, M. Parker, J. Lipton, D. Cacela, W. Erickson, and G.R. Phillips. 1995. Relative sensitivity of brown and rainbow trout to pulsed exposures of an acutely lethal mixture of metals typical of the Clark Fork River, Montana. *Canadian Journal of Fisheries and Aquatic Sciences*. 52(9): 2005-2015. https://doi.org/10.1139/f95-792
Marr, J.C.A., J.A. Hansen, J.S. Meyer, D. Cacela, T.L. Podrabsky, J. Lipton, and H.L. Bergman. 1998. Toxicity of cobalt and copper to rainbow trout: application of a mechanistic model for predicting survival. *Aquatic Toxicology*. 43: 225-237. https://doi.org/10.1016/S0166-445X(98)00061-7
McIntyre, J.K., J.W. Davis, C. Hinman, K.H. Macneale, B.F. Anulacion, N.L. Scholz, and J.D. Stark. 2015. Soil bioretention protects juvenile salmon and their prey from the toxic impacts of urban stormwater runoff. *Chemosphere*. (0). https://doi.org/10.1016/j.chemosphere.2014.12.052
McIntyre, J.K., R.C. Edmunds, B.F. Anulacion, J.W. Davis, J.P. Incardona, J.D. Stark, and N.L. Scholz. 2016. Severe Coal Tar Sealcoat Runoff Toxicity to Fish Is Prevented by Bioretention Filtration. *Environmental Science & Technology*. 50(3): 1570-1578. https://doi.org/10.1021/acs.est.5b04928
Mebane, C.A. 2006. Cadmium risks to freshwater life: derivation and validation of low-effect criteria values using laboratory and field studies. U.S. Geological Survey Scientific Investigation Report 2006-5245 (2010 rev.). 130 pp. https://doi.org/10.3133/sir20065245.
Mebane, C.A., F.S. Dillon, and D.P. Hennessy. 2012. Acute toxicity of cadmium, lead, zinc, and their mixtures to stream-resident fish and invertebrates. *Environmental Toxicology and Chemistry*. 31(6): 1334–1348. https://doi.org/10.1002/etc.1820
Mebane, C.A., R.J. Eakins, B.G. Fraser, and W.J. Adams. 2015. Recovery of a mining-damaged stream ecosystem. *Elementa: Science of the Anthropocene*. 3(1): 000042. https://doi.org/10.12952/journal.elementa.000042
Mebane, C.A., D.P. Hennessy, and F.S. Dillon. 2008. Developing acute-to-chronic toxicity ratios for lead, cadmium, and zinc using rainbow trout, a mayfly, and a midge. *Water, Air, and Soil Pollution*. 188(1-4): 41-66. https://doi.org/10.1007/s11270-007-9524-8

Mebane, C.A., D.P. Hennessy, and F.S. Dillon. 2010. Incubating rainbow trout in soft water increased their later sensitivity to cadmium and zinc. *Water, Air, and Soil Pollution*. 205(1-4): 245-250. https://doi.org/10.1007/s11270-009-0070-4

Mebane, C.A., G. Mladenka, L. Van Every, M.L. Williams, M.A. Hardy, and J.R. Garbarino. 2014. Selenium in the upper Blackfoot River watershed, southeastern Idaho, 2001-2012. U.S. Geological Survey Scientific Investigations Report 2014-5203. 34 pp. http://dx.doi.org/10.3133/sir20145203.

Melcer, H., H. Monteith, and S.G. Nutt. 1988. Variability of Toxic Trace Contaminants in Municipal Sewage Treatment Plants. *Water Science and Technology*. 20(4-5): 275-284

Meyer, J.S. and W.J. Adams. 2010. Relationship between biotic ligand model-based water quality criteria and avoidance and olfactory responses to copper by fish. *Environmental Toxicology and Chemistry*. 29(9): 2096–2103. https://doi.org/10.1002/etc.254

Meyer, J.S., C.J. Boese, and J.M. Morris. 2007. Use of the biotic ligand model to predict pulse-exposure toxicity of copper to fathead minnows (*Pimephales promelas*). *Aquatic Toxicology*. 84(2): 268-278. https://doi.org/10.1016/j.aquatox.2006.12.022 |

Meyer, J.S. and D.K. DeForest. 2018. Protectiveness of copper water quality criteria against impairment of behavior and chemo/mechanosensory responses: An update. *Environmental Toxicology and Chemistry*. 37(5): 1260-1279. https://doi.org/10.1002/etc.4096

Meyer, J.S., D.D. Gulley, M.S. Goodrich, D.C. Szmania, and A.S. Brooks. 1995. Modeling toxicity due to intermittent exposure of rainbow trout and common shiners to monochloramine. *Environmental Toxicology and Chemistry*. 14(1): 165-175. https://doi.org/10.1002/etc.5620140119

Miller, W.R. and J.I. Dreaver. 1977. Water chemistry of a stream following a storm, Absaroka Mountains, Wyoming. *GSA Bulletin*. 88(2): 286-290. https://doi.org/10.1130/0016-7606(1977)88<286:WCOASF>2.0.CO;2

Milne, I., J. Seager, M. Mallett, and I. Sims. 2000. Effects of short-term pulsed ammonia exposure on fish. *Environmental Toxicology and Chemistry*. 19(12): 2929-2936. https://doi.org/10.1002/etc.5620191213

Muyssen, B.T.A. and C.R. Janssen. 2002. Accumulation and regulation of zinc in *Daphnia magna*: links with homeostasis and toxicity. *Archives of Environmental Contamination and Toxicology*. 43(4): 0492-0496. https://doi.org/10.1007/s00244-002-1245-9

Nagorski, S.A., J.N. Moore, T.E. McKinnon, and D.B. Smith. 2003. Scale-dependent temporal variations in stream water geochemistry. *Environmental Science and Technology*. 37(5): 859–864. https://doi.org/10.1021/es025983+

Nimick, D.A., J.R. Gammon, and S.R. Parker. 2011. Diel biogeochemical processes and their effect on the aqueous chemistry of streams: A review. *Chemical Geology*. 283(1-2): 3-17. https://doi.org/10.1016/j.chemgeo.2010.08.017

Nimick, D.A., C.H. Gammons, T.E. Cleasby, J.P. Madison, D. Skaar, and C.M. Brick. 2003. Diel cycles in dissolved metal concentrations in streams: Occurrence and possible causes. *Water Resources Research*. 39(9): 1247. https://doi.org/10.1029/2002WR001571

Nordstrom, D.K. 2009. Acid rock drainage and climate change. *Journal of Geochemical Exploration*. 100(2): 97-104. https://doi.org/10.1016/j.gexplo.2008.08.002

Parker, S.R., S.R. Poulson, M.G. Smith, C.L. Weyer, and K.M. Bates. 2010. Temporal variability in the concentration and stable carbon isotope composition of dissolved inorganic and organic carbon in two Montana, USA rivers. *Aquatic Geochemistry*. 16(1): 61-84. https://doi.org/10.1007/s10498-009-9068-1

Pascoe, D. and N.A.M. Shazili. 1986. Episodic pollution: a comparison of brief and continuous exposures of rainbow trout to cadmium. *Ecotoxicology and Environmental Safety*. 12(3): 189-198. https://doi.org/10.1016/0147-6513(86)90010-2

Poléo, A.B.S. and F. Bjerkely. 2000. Effect of unstable aluminium chemistry on Arctic char (*Salvelinus alpinus*). *Canadian Journal of Fisheries and Aquatic Sciences*. 57(7): 1423-1433. https://doi.org/10.1139/f00-084

Rosseland, B.O., I.A. Blakar, A. Bulger, F. Kroglund, A. Kvellstad, E. Lydersen, D.H. Oughton, B. Salbu, M. Staurnes, and R. Vogt. 1992. The mixing zone between limed and acidic river waters: complex aluminium chemistry and extreme toxicity for salmonids. *Environmental Pollution*. 78(1): 3-8. https://doi.org/10.1016/0269-7491(92)90003-S

Santore, R.C., P.R. Paquin, D.M. Di Toro, H.E. Allen, and J.S. Meyer. 2001. Biotic ligand model of the acute toxicity of metals. 2. Application to acute copper toxicity in freshwater fish and *Daphnia*. *Environmental Toxicology and Chemistry*. 20(10): 2397-2402. https://doi.org/10.1002/etc.5620201035

Santore, R.C., A.C. Ryan, K. Croteau, and D.K. DeForest. 2016. A review of metal toxicity and exposure duration for determining the suitability of a 24-hour averaging period for comparison with acute water quality criteria. Attachment to NCASI (National Council on Air and Stream Improvement), September 15, 2016 comments to Oregon Department of Environmental Quality proposed aquatic life standards for copper. Windward Environmental, LLC. 12 pp.

Scholz, N.L., M.S. Myers, S.G. McCarthy, J.S. Labenia, J.K. McIntyre, G.M. Ylitalo, L.D. Rhodes, C.A. Laetz, C.M. Stehr, B.L. French, B. McMillan, D. Wilson, L. Reed, K.D. Lynch, S. Damm, J.W. Davis, and T.K. Collier. 2011. Recurrent Die-Offs of Adult Coho Salmon Returning to Spawn in Puget Sound Lowland Urban Streams. *PLoS ONE*. 6(12): e28013. https://doi.org/10.1371/journal.pone.0028013

Scholz, N.L., N.K. Truelove, B.L. French, B.A. Berejikian, T.P. Quinn, E. Casillas, and T.K. Collier. 2000. Diazinon disrupts antipredator and homing behaviors in chinook salmon (*Oncorhynchus tshawytscha*). *Canadian Journal of Fisheries and Aquatic Sciences*. 57: 1911-1918. https://doi.org/10.1139/f00-147

Seegert, G.L. and A.S. Brooks. 1978. The Effects of intermittent chlorination on coho salmon, alewife, spottail shiner, and rainbow smelt. *Transactions of the American Fisheries Society*. 107(2): 346-353. https://doi.org/10.1577/1548-8659(1978)107<346:TEOICO>2.0.CO;2

Seim, W.K., L.R. Curtis, S.W. Glenn, and G.A. Chapman. 1984. Growth and survival of developing steelhead trout (*Salmo gairdneri*) continuously or intermittently exposed to copper. *Canadian Journal of Fisheries and Aquatic Sciences*. 41(3): 433-438. https://doi.org/10.1139/f84-051

Slaveykova, V.I. and K.J. Wilkinson. 2005. Predicting the bioavailability of metals and metal complexes: critical review of the biotic ligand model. *Environmental Chemistry*. 2(1): 9-24. https://doi.org/10.1071/EN04076

Son, J.-H. and K.H. Carlson. 2015. Real-time surrogate analysis for potential oil and gas contamination of drinking water resources. *Applied Water Science*. 5(3): 283-289. https://doi.org/10.1007/s13201-014-0190-x

Spromberg, J.A., D.H. Baldwin, S.E. Damm, J.K. McIntyre, M. Huff, C.A. Sloan, B.F. Anulacion, J.W. Davis, and N.L. Scholz. 2016. Coho salmon spawner mortality in western US urban watersheds: bioinfiltration prevents lethal storm water impacts. *Journal of Applied Ecology*. 53(2): 398-407. https://doi.org/10.1111/1365-2664.12534

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman, and W.A. Brungs. 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic

Organisms and their Uses. U.S. Environmental Protection Agency, EPA 822-R-85-100, NTIS PB85 227049, Duluth, Narragansett, and Corvallis. 98 pp.
Thurston, R.V., C. Chakoumakos, and R.C. Russo. 1981. Effect of fluctuating exposures on the acute toxicity of ammonia to rainbow trout (*Salmo gairdneri*) and cutthroat trout (*S. clarki*). *Water Research*. 15(7): 911-917. https://doi.org/10.1016/0043-1354(81)90147-0
USEPA. 1978. Water Quality Criteria. *Federal Register*. 43(97, May 18, 1978): 21506-21518
USEPA. 1979. Water Quality Criteria. *Federal Register*. 44(52, March 15, 1979): 15926-15981
USEPA. 1991. Technical support document for water quality-based toxics control. Office of Water, U.S. Environmental Protection Agency, EPA 505/2-90-001, Washington, D.C. 143 pp. http://www.epa.gov/npdes/pubs/owm0264.pdf.
USEPA. 1995. Speed of action of metals acute toxicity to aquatic life. U.S. Environmental Protection Agency, EPA 822/R/95-002, Washington, D.C.
USEPA. 1999. 1999 update of ambient water quality criteria for ammonia. U.S. Environmental Protection Agency, EPA/822/R-99-014, Washington, D.C. 147 pp.
USEPA. 2001. 2001 update of the ambient water quality criteria for cadmium. U.S. Environmental Protection Agency, EPA/822/R-01-001, Washington, D.C. 266 pp.
USEPA. 2007. Aquatic life ambient freshwater quality criteria - copper, 2007 revision. U.S. Environmental Protection Agency, EPA-822-R-07-001 (March 2, 2007), Washington, DC. 208 pp. [Accessed 30 March 2008].
USEPA. 2013. Aquatic Life Ambient Water Quality Criteria For Ammonia – Freshwater 2013. EPA 822-R-13-001. 255 pp.
USEPA. 2016a. Aquatic Life Ambient Water Quality Criterion for Cadmium – 2016. EPA-820-R-16-002. 721 pp.
USEPA. 2016b. Aquatic Life Ambient Water Quality Criterion for Selenium – Freshwater 2016. 822-R-16-006, U.S. Environmental Protection Agency. 807 pp. https://www.epa.gov/wqc/aquatic-life-criterion-selenium [Accessed February 21, 2018].
USEPA. 2018. Aquatic Life Ambient Water Quality Criteria for Aluminum. U.S. Environmental Protection Agency, EPA-822-R-18-001. 278 pp. https://www.epa.gov/wqc/aquatic-life-criteria-aluminum.
UWAG. 2016. Comments of the Utility Water Act Group on EPA's request for scientific views: draft recommended aquatic life ambient water quality criteria for cadmium—2015; EPA-HQ-OW-2015-0753, FRL-9939-57-OW, 80 Fed. Reg. 75,097 (Dec. 1, 2015). Transmitted by Samuel L. Brown, Hunton and Williams LLP, San Francisco, CA. 10 pp.
Van Genderen, E.J., J.R. Tomasso, and S.J. Klaine. 2008. Influence of copper exposure on whole-body sodium levels in larval fathead minnows (*Pimephales promelas*). *Environmental Toxicology and Chemistry*. 27(3): 1442–1449. https://doi.org/10.1897/07-467.1
Zahner, H.M. 2009. Recovery of fathead minnows (*Pimephales promelas*) following episodic copper exposure: a biochemical, physiological, individual, and population perspective. PhD Dissertation. 247 pp. Environmental Toxicology, Clemson University, http://tigerprints.clemson.edu/all_dissertations/360.
Zahner, H.M., E.J. Van Genderen, J.R. Tomasso, and S.J. Klaine. 2006. Whole-body sodium concentration in larval fathead minnows (*Pimephales promelas*) during and following copper exposure. *Environmental Toxicology and Chemistry*. 26(6): 1635-1639. https://doi.org/10.1897/05-543R.1

Zhao, Y. and M.C. Newman. 2004. Shortcomings of the laboratory-derived median lethal concentration for predicting mortality in field populations: exposure duration and latent mortality. *Environmental Toxicology and Chemistry*. 23(9): 2147-2153. https://doi.org/10.1897/03-557

Zhao, Y. and M.C. Newman. 2006. Effects of exposure duration and recovery time during pulsed exposures. *Environmental Toxicology and Chemistry*. 25(5): 1298–1304. https://doi.org/10.1897/05-341R.1

Appendix 1: A compilation of time-to-effects in pulse or short-term toxicity testing

.

Appendix 1. Some time-to-effects in short-term toxicity tests or following pulse exposures.

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Al	Arctic char	36 hrs	LT50 ~65 hours, longer time to effects and lower effects to Al than in less tolerant species, such as brown trout and Atlantic salmon.	No	(Poléo and Bjerkely 2000)
Al	Atlantic salmon	<7 hrs	LT50 in mixing zone was 7 hrs where fish were exposed to Al precipitates; LT50 in high dissolved Al in acidic waters was 22 hours	Yes,	(Rosseland et al. 1992)
Al	Brown trout	<7 hrs	LT50 in mixing zone was 7 hrs where fish were exposed to Al precipitates; LT50 in high dissolved Al in acidic waters was 40 hours	Yes,	(Rosseland et al. 1992)
Ammonia	Freshwater crustaceans Gammarus and Asellus	24 hours	Full range of mortalities (0-100%) in tests. Adults more sensitive than juveniles.	No	(Maltby 1995)
Ammonia	Rainbow trout	1 hours	[Mortality of fish following 1, 6 or 24 hour exposures were similar with EC50s roughly 2X that of 14-day exposures; ~200 minutes was the critical exposure time for whether fish could recover from the pulses	Yes, to 6-days	(Milne et al. 2000)
Ammonia	Cutthroat and rainbow trout	2.5 hours	First deaths during a single high pulse occurred a 2.5 h, by which time all fish were stressed. Fish with moderate stress (erratic swimming) all recovered after transfer to clean water; fish with severe stress (immobilized but alive) all later died	Yes, to 96-hours	(Thurston et al. 1981)
Ammonia	Fathead minnow, Daphnia magna	24 hours	Mortality follow 24 h ammonia pulses ceased shortly after removal of the pulse	Yes to 7 or 21d	(Diamond et al. 2006)
Cd	White sucker	12 hours	12-hr LC50 was 5X higher than 96-hr LC50; (1.1 vs. 5.35 mg/L)	No	(Duncan and Klaverkamp 1983)
Cd	Moina macrocopa	3 hours	Adverse effects noted in all treatments; nominal exposures were very high (≥ 80 µg/L) relative to expected environmental concentrations	Yes, to 21-days	(Gama-Flores et al. 2007)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Cd	Bull trout, Rainbow trout	<24-hours	For a given concentration, time to death shorter at low hardnesses and higher pH. LT50s ranged between 24 and 84 hours	Yes, to 6 days	(Hansen et al. 2002)
Cd	Fathead minnow	6 hours	A 6-hr pulse to 60 µg/L (5X the 7d EC25) killed ~55% within 24h after the pulse; a 12-hr pulse to 40 µg/L (3X the 7d EC25) killed ~60% within 24h after the pulse;	Yes, to 7-days	(Diamond et al. 2005)
Cd	Green hydra	>90 minutes	No effects of a 1.5 µg/L, 90-minute pulse of Cd on hydra numbers 7-days after exposure (highest pulse was 2X the 7-d LOEC concentration)	Yes, to 7-days	(Holdway et al. 2001)
Cd	Rainbow trout	<24 hours	7% mortality by 24h at 4.6 µg/L in conventional 96-h exposure (test #8)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	60% mortality by 24h at 1.1 µg/L in conventional 96-h exposure (test #9)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	55% mortality by 24h at 2 µg/L in conventional 96-h exposure (test #10)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	50% mortality by 24h at 2 µg/L in conventional 96-h exposure (test #11)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	25% mortality by 24h at 3 µg/L in conventional 96-h exposure (test #12)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	20% mortality by 24h at 2.4 µg/L in conventional 96-h exposure (test #13)	No	(Mebane et al. 2012)
Cd	Rainbow trout	<24 hours	27% mortality by 24h at 2.1 µg/L in conventional 96-h exposure (test #14)	No	(Mebane et al. 2012)
Cd	Cutthroat trout	<24 hours	27% mortality by 24h at 1.7 µg/L in conventional 96-h exposure (test #21)	No	(Mebane et al. 2012)
Cd	Rainbow trout	~2 hours	100 min at 1 mg/L Cd produced 50% delayed mortality by 86h Although continuous exposure may reinforce toxicity and accelerate mortality, the initial toxic effects from brief Cd exposures were irreversible and the lethal effect was inevitable. 1 mg/L Cd in continuous exposure required 95 hours for 100% mortality.	Yes, to 96 h	(Pascoe and Shazili 1986)
Cd+Zn	Rainbow trout	<24 hours	65% mortality by 24h in mining polluted ambient water in conventional 96-h exposure (test #142)	No	(Mebane et al. 2012)
Cd+Zn	Rainbow trout	<24 hours	90% mortality by 24h in mining polluted ambient water in conventional 96-h exposure (test #147)	No	(Mebane et al. 2012)
Cd+Zn	Cutthroat trout	<24 hours	43% mortality by 24h in mining polluted ambient water in conventional 96-h exposure (test #148)	No	(Mebane et al. 2012)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Cd+Zn mixtures	Rainbow Trout	<24 hours	Up to 85% mortality by 24h in mining polluted ambient water in conventional 96-h exposure	No	(Mebane et al. 2010)
Chlorine	Six fish species	~ 1 hr	Avoidance of different chlorine forms usually occurred at about 0.2 to 0.3X the continuous 48-hour LC50s, but at <0.1X for the sensitive coho salmon	No	(Cherry et al. 1979)
Chlorine	Mysid shrimp	2 hours	LC50 from 2-hour exposures followed to 96 h were 3.7X higher than LC50 from continuous 96-h exposure	Yes, to 96-hr	(Fisher et al. 1994)
Chlorine	Silverside (Menidia)	2 hours	LC50 from 2-hour exposures followed to 96 h were 1.8X higher than LC50 from continuous 96-h exposure	Yes, to 96-hr	(Fisher et al. 1994)
Chlorine	Mysid shrimp	2 hours	NOEC from 2-hour exposures followed to 144 h were 4X higher than NOEC from continuous 144-h exposure	Yes, to 144-hr	(Fisher et al. 1994)
Chlorine	Silverside (Menidia)	2 hours	NOEC from 2-hour exposures followed to 144 h were 2X higher than NOEC from continuous 144-h exposure	Yes, to 144-hr	(Fisher et al. 1994)
Chlorine	Five fish species	30 minutes	Mortalities usually occurred within 24 h after the 30-min exposures. Fish rarely recovered following loss of equilibrium. Mortalities higher at higher temperatures	Yes, to 48-h	(Seegert and Brooks 1978)
Cu	Amphipod, *Melita plumosa*	12-hour	Negligible deaths occurred during pulse; 96-h was sufficient post exposure duration to capture mortalities.	Yes, to 240-h	(Angel et al. 2010)
Cu	Green algae, Chlorella	1 hour	Effects differed greatly between species. With Chlorella, 1-h at 15 µg/L Cu inhibited growth at 72-hrs similarly as did 24-hr at 20 µg/L to *Pseudokirchneriella subcapitata*	Yes, to 72-h	(Angel et al. 2017)
Cu	Coho salmon	30 – 60 minutes	Olfactory neurotoxicity; Increases in copper impaired the neurophysiological response to all odorants within 10 min of exposure.	No	(Baldwin et al. 2003)
Cu	Chinook salmon, rainbow trout	30 minutes	Behavioral avoidance of Cu concentrations at about 0.1X the concentration required to kill olfactory receptor cells	No	(Hansen et al. 1999a)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Cu	Chinook salmon, rainbow trout	1 hour	Olfactory neurotoxicity ;number of olfactory receptors was significantly reduced in chinook salmon exposed to >=50 µg/L Cu and in rainbow trout exposed to >=200 µg/L Cu for 1 h. By 4h, both species has lost receptors at 25 µg/L	No	(Hansen et al. 1999b)
Cu	Rainbow trout	<12 hours	24-hour LC50 about 2.5X higher than asymptotic, ultimate LC50	No	(Marr et al. 1998)
Cu	Fathead minnow	12-hour	Fish survival in various pulsed exposures (6 h, 12 h, 4 d, and 14 d) was significantly lower than in continuous exposure. Little delayed mortality observed	Yes, to 144-h	(Bearr et al. 2006)
Cu	Fathead minnow	4 hours	Cu bioaccumulation on gills increased rapidly, reaching half saturation by 4 h and approaching an asymptote in exposures longer than 12 h,	No	(Brooks et al. 2006)
Cu	Stream communities	1-hour	Release of high dose (12 mg/L) chelated Cu in efforts to reduce Didymo "rock snot" diatom blooms. Pulse exposure of Cu produced localized fish kills and decreased mayflies but minimal adverse effects were detectable 21-d after the pulse	Yes, to 21-days	(Clearwater et al. 2011)
Cu	Stream communities	~6-hours	Rain storm on reclaimed mine waste produced a pulse of Cu that peaked at about 3X the acute criterion; biological sampling 4-days later showed unusually low densities of the usually common mayfly *Baetis*. No effects were detectable on trout abundance in the affected waters.	In effect, as field monitoring was 4 days later	(Mebane et al. 2015)
Cu	Fathead minnow	12 hours	A 12-hr pulse to 50 µg/L (3X the 7d EC25) killed ~25% within 24h after the pulse; 6-hr pulses had little effects	Yes, to 7-days	(Diamond et al. 2005)
Cu	Fathead minnow	6-hour	Fish survival in various pulsed exposures (6 h, 12 h, 4 d, and 14 d) was significantly greater than in continuous exposures. In that study, the majority of mortality effects were manifested within the first 24 hours following the copper pulse.	Yes, to 28-d	(Erickson 2007)
Cu	Fathead minnow	4-8-hours	A 30% loss of whole body Na from control levels was reached between 4 and 8 h exposures. Death was associated with 20 to 40% loss.	Yes, 48-hours	(Van Genderen et al. 2008)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Cu	Fathead minnow	3 hours	3, 6, or 9 hour pulses of Cu resulted in similar mortality to fathead minnows; 18hr pulses were more toxic. Most mortalities occurring within the first 48 hours after exposures). Mortality lagged behind whole-body sodium loss by hours to days after the end of the exposure; organisms could recover from brief whole-body sodium losses.	Yes, to 240 hours	(Zahner 2009)
Cu	Fathead minnow	12-hours	12-hours was point of no return for loss of Na ions to fatal levels; fish could recover from exposures of 3-9 hours	Yes, 48-hours	(Zahner et al. 2006)
Cu	Fathead minnow	2 hours	Fish exposed to 2- or 8h daily pulses for 3d. Kinetic model considered feasible to predict accumulations and mortalities, but delayed deaths during the recovery phases of the exposures precluded accurate predictions of a pulse-exposure Cu lethal accumulations	Yes, to 72 hours	(Meyer et al. 2007)
Cu	Fathead minnow, Daphnia magna	24 hours	Mortality follow 24 h Cu pulses ceased shortly after removal of the pulse	Yes to 7 or 21d	(Diamond et al. 2006)
Cu	Hyalella azteca	12 hours	Complete LC50s that included latent mortality were 25-50% lower than conventional LC50s when observations ended with end of exposure. Conventional = exposure and observations ended at 48-hours; complete = exposed for 48hr, transferred to clean water and observed to 112 h	Yes, to 112 hours	(Zhao and Newman 2004)
Cu	Hyalella azteca	3 hours	In high Cu concentrations (0.6 and 1.0 mg/L) with amphipods exposed for 20-61 hours, mortalities began by first observation at 3 hrs	Yes, to 7 days	(Zhao and Newman 2006)
Cu, Zn	Daphnia magna	12-hours	Pulse sensitivity depended on organism age. For most sensitive aged Daphnia, mortalities following a single 12-h pulse were similar to those from 96-h exposures. Survivors had no lingering effects on growth or reproduction	Yes, to 21-days	(Hoang and Klaine 2007)
Cu, Zn, Cd, Pb	Rainbow Trout, Brown trout	8-hours	8-hr LC50s were 4X to 10X greater than 96-hr LC50s	Yes, to 96-h	(Marr et al. 1995)
Diazinon	Chinook salmon	2h + 1h recovery	Olfactory-mediated alarm responses inhibited at concentrations as low as 1 µg/L, which is only about 2X the lowest chronic effect concentration for fish	No	(Scholz et al. 2000)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Mg	Plants (Green algae, duckweed)	4 hours	4-hour exposures were similarly toxic to Chlorella as 96-hour exposures in 4,8,24, and 96-hour exposures. Duckweed growth only adversely affected by 8- and 24-hour exposures with 8-hr EC50 ~3X higher than 96-hours	Yes, to 96-hours	(Hogan et al. 2013)
Mg	Invertebrates (daphnid, pulmonate snail)	4 hours	4-hour pulse that hit daphnids at start of reproductive period had more severe effects than similar pulse during non-reproductive periods. Lowest 4-h EC50s ~3X higher than lowest 96-hour EC50s. Snails more resistant than daphnids to pulses; 8-hr snail EC50 30X higher than 96-hour EC50	Yes, to 96-hours	(Hogan et al. 2013)
Mg	Hydra	4 hours	Vulnerable to pulse toxicity with 4-hour EC50 ~3X higher than 96-hour exposures	Yes, to 96-hours	(Hogan et al. 2013)
Mg	Fish (trout gudgeon)	>24 hours	No mortality following 4, 8, or 24-hour pulse exposures	Yes, to 96-hours	(Hogan et al. 2013)
Mono-chloramine	Rainbow trout, fathead minnow	2 hours	Repeated 2-h pulse/22-h recovery cycles; EC50s from pulsed exposures only about 1.1 to 1.3X greater than continuous exposures	Yes, to 96-hours	(Meyer et al. 1995)
NaCl	Fathead minnow	3 hours	A 3-hr pulse to 5X the 7d EC25) killed ~65% with delayed mortalities occurring up 96h after the pulse; a 24-hr pulse to 4X the 7d EC25 killed none during the 24-h exposures, but delayed mortalities killed 75%, with most dying in the next 24 hrs after the pulse ended	Yes, to 7-days	(Diamond et al. 2005)
Nitrite	Amphipod	8 hours	No mortality at end of exposures, but deaths occurred during the post-exposure observation period	Yes, to 96-hours	(Alonso and Camargo 2009)
Pb	Cutthroat trout	<24 hours	35% maximum mortality by 24h in conventional 96-h exposure (test #91). However, most tests with Pb and cutthroat or rainbow trout showed little or no mortality by 24 hours.	No	(Mebane et al. 2012)
PCBZ	Hyalella azteca	29 hrs	Pentachlorobenzene (PCBZ) exhibited a classic time-concentration response with LT50s ranging from 29-388 hours	Yes, variable times up to 28-d	(Landrum et al. 2004)
PCP	Hyalella azteca	NA	Little mortality observed among amphipods exposed for 20-60 hours to high concentrations (0.4 and 0.6 mg/L pentachlorophenol, PCP).	Yes, to 7 days	(Zhao and Newman 2006)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
PCP (penta-chlorophenol	Hyalella azteca	4 hours	In contrast to tests with Cu, complete LC50s including latent mortality were similar to conventional LC50s. Conventional = exposure and observations ended at 48-hours; complete = exposed for 48hr, transferred to clean water and observed to 112 h	Yes, to 112 hours	(Zhao and Newman 2004)
Pesticide mixtures	Coho salmon	24 hours	Mixture of 2 pesticides (diazinon + malathion at their 1.0 and 0.4 96h EC50 concentrations) caused severe effects within first 24h of exposure. Brain biochemistry affected followed by loss of equilibrium, rapid gilling, altered startle response, and increased mucus production, and then 100% death within 24h. Exposures to were near extreme range of measured environmental concentrations	Yes, to 96 hours	(Laetz et al. 2009)
pH (acid pulse)	Fathead minnow	15 minutes	A 15 min drop in pH from 7.2 to 6 caused about 10% delayed mortalities. A 1h drop in pH from 7.2 to 3 killed 100% within the hour, while a 1h drop from 7.2 to 4 killed only 35%, with most deaths occurring within 24 hours after the pulse	Yes, to 7-days	(Diamond et al. 2005)
Phenanthrene	Zebrafish	30 minutes	Cardiotoxicity led to outright deaths or followed by abnormal development which in turn led to latent impacts on physiology at later life stages	Varied	(Brette et al. 2017; Incardona and Scholz 2017)
Urban highway runoff	Coho salmon	<4 hours	100% of adult salmon were dead after 4h exposure to rain runoff collected from urban highway overpass	No, all dead	(Spromberg et al. 2016)
Urban stormwater	Coho salmon	2-4 hours	Majority of the healthy salmon entering urban spawning streams on stormwater freshets suffered pre-spawn mortality syndrome with altered behavior (lethargy, rolling, gaping, loss of equilibrium) followed by death.	Yes	(Scholz et al. 2011)
Urban stormwater	Coho salmon	12 hours	100% of juvenile coho salmon died within the first 12 h of exposure	No, all dead	(McIntyre et al. 2015)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Urban stormwater	Ceriodaphnia dubia, fathead minnow, green algae, sea urchin, bacteria	1 hour	Samples collected at 0 to 60 minutes after storm onset generally showed greatest toxicity in short-term marine (20m to 22h) and longer term (7d) freshwater tests. Up to 100% toxicity of undiluted rain runoff. Toxicity to both Ceriodaphnia and fathead minnows was generally greatest in grab samples collected early during the storm event; with the greatest effects present in samples collected during the first 60 min. of discharge. Most toxicity attributed to Cu and Zn in runoff.	Yes, to 4 or 7 days	(Kayhanian et al. 2008)
Zn	Fathead minnow, Daphnia magna,	24 hours	24 h zinc pulses caused continued effects for several days following removal of the pulse (not so with Cu or ammonia)	Yes to 7 or 21d	(Diamond et al. 2006)
Zn	White sucker	<12 hours	12-hr LC50 5X higher than 96-hr LC50 (2.2 vs. 13.3 mg/L)	No	(Duncan and Klaverkamp 1983)
Zn	Bull trout, Rainbow trout	<24-hours	For a given concentration, time to death shorter at low hardnesses and higher pH. LT50s ranged between 24 and 84 hours	Yes, to 6 days	(Hansen et al. 2002)
Zn	Rainbow trout	<24 hours	Up to 37% mortality by 24h in conventional 96-h exposure (test #109)	No	(Mebane et al. 2012)
Zn	Rainbow trout	<24 hours	Up to 65% mortality by 24h in conventional 96-h exposure (test #110)	No	(Mebane et al. 2012)
Zn	Rainbow trout	<24 hours	Up to 55% mortality by 24h in conventional 96-h exposure (test #110)	No	(Mebane et al. 2012)
Zn	Rainbow trout	<24 hours	Up to 65% mortality by 24h in conventional 96-h exposure (test #111)	No	(Mebane et al. 2012)
Zn	Rainbow trout	<24 hours	Up to 55% mortality by 24h in conventional 96-h exposure (test #112)	No	(Mebane et al. 2012)
Zn	Cutthroat trout	<24 hours	Up to 53% mortality by 24h in conventional 96-h exposure (test #123)	No	(Mebane et al. 2012)
Zn	Daphnia magna	<24-hours	Accumulation. Zinc uptake and elimination are rapid processes; major increases and decreases in body content occurred within 1 day.	Yes, to 12 days	(Muyssen and Janssen 2002)
Zn	Ceriodaphnia dubia	1 hour	*C. dubia* were highly susceptible to delayed mortality. For example 1-hour exposures to 250 µg/L Zn killed none during the 1-hour actual exposure period, but after 35 hours of "recovery time" in clean water, all had died.	Yes, up to 48 hours	(Ivey and Mebane 2019)

Substance	Organisms	Minimum time to onset of effects or pulse duration	Effects	Organisms observed for delayed mortality?	Source
Zn	Rainbow trout	8 hours	Rainbow trout were largely unaffected by short-term (≤ 3-hour) exposures to very high concentrations of Zn. No fish died by 96 hours following a 1-hour exposure to >50 mg/L Zn. In contrast, an 8-hour exposure to 2 mg/L killed >50% of the fish by 36 hours.	Yes, up to 144 hours	(Ivey and Mebane 2019)

Appendix C. The Capacity of Aquatic Ecosystems to Recover from Exceedances of Aquatic Life Criteria

The capacity of aquatic ecosystems to recover from exceedences of aquatic life criteria

Workshop discussion initiation paper no. 2 for the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019*

Contents

Summary 2
Observations and Findings: 2
Background 4
History of the "how often" provision of criteria 5
1985 5
1991 6
Allowable percentage of exceedences approaches 6
Recovery from what? 7
Approach 11
Previous reviews 11
Capacity of individual organisms to recover from chemical stress 14
Tissue accumulation/depuration – slow kinetics. 15
Time to recover from tissue burdens 15
Recurrence interval for tissue burdens in fish 17
Recovery of aquatic communities 20
Modeled population recoveries 25
Frequency versus return period of exceedences 35
References 45
Appendix 1: Freshwater disturbance and recovery case study summaries 59
Appendix 2: Stage-structured population model summaries 72

* Prepared by Christopher A. Mebane on behalf of the Workshop participants. If further referenced, the following form is suggested:
Mebane, C.A. 2019. The capacity of aquatic ecosystems to recover from exceedences of aquatic life criteria. Workshop discussion initiation paper no. 2 from the EPA Invited Experts Workshop on the Frequency and Duration terms in Aquatic Life Criteria, Arlington, VA, September 11-12, 2019. 72 pp.

Summary

In the USA, most chemical water quality criteria for the protection of aquatic life follow a three-part magnitude-duration-frequency form. This form of criteria presumes that aquatic ecosystems have sufficient resiliency to recover from occasional toxic disturbances that might result from criteria exceedences if the exceedences occurred no more frequently than once every three years. This guidance was based on a 1985 review of case studies of aquatic ecosystem recovery from diverse disturbances such as floods, droughts, spills, and deliberate eradication efforts.

Here, aquatic ecosystem recovery from toxic disturbances was analyzed through three major steps:

Step 1. Literature on the ability of individual organisms and communities to recover from a range of disturbances was compiled and summarized. A major limitation of step 1 is that most recovery studies were of unplanned disturbances that resulted in more severe effects than would be expected from variability or minor upsets in regulated discharges.

Step 2. Potential differences in recovery of aquatic populations with different life history or population characteristics were illustrated through population modeling. Specifically, population models were used to contrast the different population-level consequences of the same magnitude of initial disturbance applied at either a 1-in-3 years or 1-in-2 years return periods. To contrast the role of differing life history characteristics on recovery times, populations of one invertebrate (*Hyalella azteca*) and four fish species were modeled (fathead minnow, brook trout, Chinook salmon, and shortnose sturgeon) with common disturbance magnitudes and intervals.

Step 3. The potential biological implications of different short-term allowable exceedence approaches were contrasted using several real-world examples of time-variable pollution. Expressing exceedance as an allowable sample-exceedance percentage has been suggested as an alternative to the 1-in-3 year exceedance frequency to facilitate monitoring and enforcement. Implications of a 5% of samples exceedence frequency were compared to the 1-in-3 year return interval in real monitoring datasets.

Observations and Findings:

1. Reports on the physiological recovery of organisms following sublethal exposures to contaminants in laboratory settings usually involved accumulation/depuration kinetics in organisms with few signs of "clinical" morbidity. Depuration usually took longer to reach baseline conditions than did the original exposures, and except for highly persistent substances, usually took several weeks to months to decline to baseline. In one particularly robust study with realistic selenium exposures in experimental streams, selenium concentrations in fish had not declined to baseline after 1-year of recovery. This lengthy recovery suggests that for criteria expressed as tissue residues in fish, longer intervals between potentially harmful exceedences may be warranted than for more rapidly varying water concentrations. Minimizing the number of tissue criterion exceedence episodes during the life span of most fish is recommended. A goal of limiting exposures to no more than one exceedence per life span is doubtfully feasible, since freshwater fish lifespans

range from about 1 to 100 years, and many species have average life spans of 5 years or more.

2. Time to recovery of aquatic communities after toxic disturbances is highly situational, depending upon factors such as the severity of disturbance, spatial extent of disturbance, availability of refuge habitat, proximity of colonizing sources, population size, the life history of the affected organisms, the disturbance history of the affected habitats, and recovery metric considered. Times to recovery ranged from a few weeks to >10 years for toxic disturbances unrelated to physical habitat alteration. When limited to non-catastrophic, pulse events, recoveries occurred within a year in 75% of the case studies.
3. Modeled recoveries from substantial but non-catastrophic scenarios (50% mortality event to all life stages, either once in 3 years, or every other year) showed no population decline for *Hyalella azteca* or fathead minnow. These species have short life cycles and high reproductive potential. Brook trout were projected to have declines in abundance but had negligible risk of population extirpation. A vulnerable population of Chinook salmon had very high baseline risk of extirpation but applying additional mortality every third generation to juveniles during their freshwater resident stage did not appreciably increase risk. A vulnerable population of shortnose sturgeon, a slow maturing and long-lived species, was projected to be decimated in these same scenarios. However, a chronic low-effects scenario patterned after the fish-tissue based selenium criteria was (mathematically) tolerated by the population.
4. Comparisons of allowable exceedence frequencies with return interval exceedences for some intensely monitored waterbodies showed that a 5% exceedence rate and a 1-in-3 year recurrence of 30-day average concentrations were almost the same thing. However, in datasets from small, flashy streams, a 5% exceedence allowance would allow more exceedences than a 1-in-3-year recurrence of a 4-day average concentration.
5. Some conceptual biological implications of an allowable exceedence percentage vs a 1-in-3 years exceedence return frequency were explored using high-aluminum streams as examples. Applying exceedence magnitudes to the aluminum chronic criterion species-sensitivity distribution (SSD), could potentially cause brief, adverse effects to up about 25% of more sensitive species in the SSD. Similarly, applying the exceedence episodes to a brook trout mortality model projected a single 30% mortality event during the 3-year periods.
6. An allowable exceedence expressed as a percentage (i.e., 5%) of samples cannot be separated from magnitude of exceedences. In the scenarios examined, a 2X criteria exceedence was projected to be considerably more harmful than a 1.1X exceedence, owing to steep response curves. This would also apply to the 1-in-3 year return interval approach.
7. Nothing in any of the analyses refutes EPA's (1985) and Stephan et al.'s (1985) judgements that most aquatic ecosystems can tolerate water-based criteria exceedences once in a three year period, assuming that the magnitude of exceedence is <2X the criterion and the water bodies are not subject to anthropogenic stress other than the exceedence of concern. Nor did the analyses suggest that this 1-in-3 year exceedence allowance was the only appropriate judgement that could be reached.
8. For fish-tissue based criteria such as Hg, Se, or persistent organic compounds, the fish tissue concentrations build up through the foodweb, and to diminish, require elimination of source reservoirs lower in the food web. For these "long wave" pulses, a longer recurrence interval

than that for water column exceedences seems appropriate, such as on the order of 5 to 10 years.

9. Because the assumption that water bodies are not subject to anthropogenic stress other than the exceedence of concern is untrue in most discharge settings, limiting the allowable magnitude of criteria excursions could be more explicitly stated. For instance, a criteria expression could be along the lines of "*freshwater aquatic organisms and their uses should not be affected unacceptably if the four-day average concentrations of [a chemical] do not exceed [the "chronic" criterion continuous concentration] more than once every three years on the average and if no criterion exceedences are greater than a factor of 2, ….*"
10. Similarly, should an allowable frequency of exceedences approach be used as an alternative approach, to maintain a similar level of protection as the present 1-in-3 return interval approach, a lower allowable exceedence magnitude could be applied. For instance, a criteria expression might be worded along the lines of "*freshwater aquatic organisms and their uses should not be affected unacceptably if no more than 5% of the representative sample data exceed [the "chronic" criterion continuous concentration] during [the period of interest] and no single exceedence is greater than a factor of 1.5.*"
11. The most straightforward approach to tying a CCC exceedence magnitude cap to an allowable frequency definition may be to use the CMC, as by design, the CMC protects against short-term exceedences. This would have the advantage of continuing to make use of the large body of acute data and retains the familiar two-number criteria structure. Some adjustment would be needed for substances which have similar acute and chronic criteria, such as Zn.

Background

In the USA, most chemical water quality criteria for the protection of aquatic life follow a three-part intensity-duration-frequency format. This format effectively asks, "how often may an elevated concentration of specified duration occur without causing unacceptable harm to aquatic communities?" Since 1985, the U.S. Environmental Protection Agency (USEPA) has defined aquatic life criteria in this format, in which the "how often" term is defined as a recurrence interval, the "magnitude" term is defined as a concentration expected to fully protect 95% of the genera in a distribution of sensitivity values, and the "duration" term is defined as an averaging period sufficient to meet the 95% protection goals in short and longer-term exposures yet to allow for variability in wastewater engineering controls. The "how often" criteria component is defined as allowing an exceedence once every three years (Stephan et al. 1985).

There is some disparity between (a) the 1 in 3 year exceedence frequency recommended in the guidelines for derivation of aquatic life criteria (Stephan et al. 1985); (b) the 1 in 10 year frequency of low streamflows used in permit design conditions to comply with the 1 in 3 year exceedence (USEPA 1985, 1991), and (c) the 10% frequency used for judging waters to be impaired by "conventionals" in monitoring assessments of impaired waters done under CWA Sections 303(d) and 305(b) (USEPA 1997). For the latter purpose, "conventionals" were listed as dissolved oxygen, pH, and temperature, which is subtly different from the term "conventional pollutants" which has statutory history (Copeland 1993). A brief history of and distinctions between these uses follows.

History of the "how often" provision of criteria

1985

While commonly attributed to Stephan et al. (Stephan et al. 1985), the no more than one criteria exceedence in a three year period expressions used throughout USEPA's aquatic life criteria documents can be traced back to EPA's (1985) "*Technical support document for water quality-based toxics control,*" and no earlier. While this highly influential document was organized and published as an EPA report, it was a collaborative effort with a list of contributors including leading ecotoxicologists, chemists, and engineers of the time from universities, the chemical and petroleum industry, consultancies, state pollution agencies, and EPA staff scientists. This diverse authorship is reflected in the eclectic blend of toxicology, wastewater engineering, ecology, and pragmatic, subjective judgements. In the 1985 "TSD" the contributors recognized that criteria needed to be expressed in terms of magnitude, duration, and frequency, that frequency needed to consider the potential severity of biological impacts. Potential impacts, in turn, considered the characteristics of the disturbance and physical characteristics and biology of the aquatic habitats. They reviewed 8 studies of recovery of freshwater habitats after disturbances, although one, Cairns' (1971) "*The recovery of damaged streams*" was itself a report of 4 case studies.

Based on the recovery case studies, USEPA (1985) concluded that:

> *"an ecosystem capable of reestablishment would be back to its normal function in three to four years after a major stress. Based on logic, one would not want an ecosystem in a constant state of recovery and increased vulnerability from multiple stresses. An ecosystem severely stressed every few years will not reach its full potential.*
>
> *The following are the toxicologically based recommendations for ambient concentrations.*
>
> > *• The one-hour average concentration should not exceed the CMC [acute criterion] more than once every three years on the average.*
> > *• The four-day average concentration should not exceed the CCC [chronic criterion] more than once every three years on the average.*
>
> *If the biological community is under stress because of spills, multiple discharges, etc., or has a low recovery potential, or if a local species is very important, the frequency should be decreased."*

Although there were no listed authors in common between the USEPA (1985) and the Stephan et al. (1985) reports, the precise language and common terminology demonstrates the close coordination between the two efforts. However, while the USEPA (1985) authors recognized that waterbodies receiving industrial or urban dischargers are often affected by upsets or spills, Stephan et al. (1985) subtly modified the rationale for selecting the frequency of allowed exceedences by arguing that the 1-in-3 year exceedence allowance was intended for normal operation and variation in wastewater plant discharges. Stephan et al. (1985) rationalize that the frequency of allowed exceedences should be based on the ability of aquatic ecosystems to recover from the exceedences, which will depend in part on the magnitudes and durations of the exceedences. If spills and accidents are excluded from the usual variation of water quality, "*most of the exceedences will be small and exceedences as large a factor of two will be rare.*" For small magnitude exceedences (<2X the CCC), they considered that "*if the body of water is not subject to anthropogenic stress other than the exceedences of concern and if exceedences as large as a factor of two are rare, it seems reasonable that most bodies of water could tolerate exceedences once every three years on the average.*" Stephan et al. (1985) also introduced the idea that both the frequency and duration aspects of the national criteria expression could be adjusted in a site-specific-criterion, but did not discuss how that might be done.

1991

USEPA (1991) was an update to USEPA (1985), and provided substantially more detail on the duration and frequency aspects of criteria, although the substantial guidance was similar. Some considerations for site-specific criteria adjustments to the national guidelines for allowable exceedences were given. For instance, many lower-order (smaller) streams, particularly those for which refugia are available, may be able to tolerate somewhat higher excursion frequencies than once per 3 years, but recovery periods substantially longer than 3 years may be necessary after multiple minor excursions or after a single major excursion or spill during a low-flow period in medium-to-large rivers, and up to 25 years where long-lived fish species are to be protected.

The USEPA (1991) considerations came on the heels of a 1989 symposium on recovery in aquatic ecosystems, which had been sponsored by multiple USEPA program offices (Yount and Niemi 1990a). The symposium provided a much richer body of work on recoveries of aquatic ecosystems, and controlling factors than had been available to the 1985 efforts. However, this greatly expanded literature review did not lead to substantive differences from the Stephan et al. (1985) and USEPA (1985) allowable frequency recommendations.

Allowable percentage of exceedences approaches

Implementing the 1985 Guidelines (Stephan et al. 1985) approach with its two-number acute and chronic criteria magnitudes for a substance, paired with 1-hour and 4-day exposure durations, respectively, that are not to be exceeded more than once every 3-years on the average, have practical difficulties for compliance monitoring and enforcement (Benson et al. 2003). For instance, while a discrete dip sample from a stream can easily be considered to represent a 1-hour average concentration, sampling sufficient to evaluate compliance with a 4-day average concentration might be interpreted to require 3 sampling trips, spread across 4 days.

The U.S. Clean Water Act sections 305(b) and 303(d) require that states compile inventories of impaired waters that do not meet water quality standards. USEPA (1997) recommended an allowable percentage approach to assessing compliance with certain so-called "conventional pollutants" (pH, dissolved oxygen, and temperature). In this approach, if no more than 10% of the available monitoring data for a waterbody exceeded criteria values, the waterbody was considered in compliance for that criteria. USEPA (1997) gave no explanation why a 10% exceedence percentage was recommended over other percentages, nor was any rationale given why a 10% exceedence percentage of undefined magnitude would be protective. For instance, given a dissolved oxygen (DO) criterion of 6 mg/L, should 10% of the data fall between 6 and 5 mg/L would likely have much lower community effects than would DO concentrations dropping to 1 mg/L. Similarly, various state waterbody assessment programs have adopted or proposed de facto exceedence percentage policies by setting an exceedence percentage combined with a minimum sample size per waterbody combined with a data aging policy (Keller and Cavallaro 2008; IDEQ 2016; NCDEQ 2016). In Colorado for example, for a waterbody to be considered impaired for exceeding dissolved metals criteria, at least 15% of the values must exceed criteria, with a minimum sample size of 10, collected at different times of the year within the previous 5 years (CDPHE 2017). No rationale for the selection of the 85% compliance level was found.

Benson et al. (2003) suggested a 5% exceedence allowance approach, such that a goal would be to protect 95% of the taxa 95% of the time. Again, no explanation of why a 5% exceedence percentage was suggested, although in the absence of explanation it was presumed to just have been a common value, such as using a $P<0.05$ statistical significance threshold or the HC5, hazardous concentration to 5% of taxa threshold commonly used in species-sensitivity distributions for criteria or risk assessment. They further suggested that defining an allowable magnitude of exceedence should also be considered in criteria expression (Benson et al. 2003). Taken literally, protecting taxa for 95% of the time would not be sustainable if that 5% gap prevented them from completing their life cycles. However, just as the goal to protect 95% of the taxa does not necessarily mean that 5% of the taxa in a community are truly sacrificed, such protection goals likely reflect practical cutoffs for calculation rather than true toxic condition fractions.

Recovery from what?

Stephen et al. (1985) presumes effects of small criteria exceedances are small to the point of having no distinguishable effects, ignoring common disturbances such as stormwater runoff, plant upsets, bypasses, sewer overflows, spills, etc. This may be an unrealistic way to attempt to consider a waterbody in isolation from the multiple point and nonpoint pollution sources that affect most waterbodies in populated areas. For example, in EPA's complex effluent toxicity testing project that was undertaken to evaluate the performance of whole-effluent toxicity (WET) tests, they visited 8 urban and industrial waterways each for 7 days. In 25% of their visits, spills into the receiving waters sufficient to cause acute toxicity just happened to occur during these 7-day visits (USEPA 1991). This suggested that while spills upsets or spills causing toxicity might be uncommon for an individual facility, when multiple facilities discharge to the same waterbody, with dense transportation networks and nonpoint sources in a catchment, the impacts caused by spills might be as important as impacts caused by variation in the compositions and flows of the effluent and the receiving water. From an engineering perspective, over the life of a facility, upsets and system failures are predictable if infrequent risks (Sweetapple et al. 2018). However, these risks can become very high when cost concerns lead to deferred maintenance or upgrades, which can lead to failures even in large, sophisticated operations (Willmsen and Mapes 2017).

Fish kills from releases such as wastewater plant upsets or activities as mundane as draining a swimming pool are common occurrences. For example, of the 3501 fish kills investigated by the Maryland Department of the Environment from 1984 through 2016, natural or unknown causes were most common (64%) and only 283 (8%) were attributed to pollution events. Of those, leading causes included municipal sewage and discharges (71), unidentified pollution sources (54) industrial discharges (52), agricultural activities (32), fuel spills (30), and draining swimming pools (19) (MDE 2017). In Missouri, fish kill investigations from 1988-2017 found that municipal discharge events resulted in 3 to 20 investigated fish kills per year, industrial discharge events resulted in 0 to 12 per year, and agricultural events resulted in 0 to 14 investigated kills per year (MDC 2018). The Maryland and Missouri summaries are mentioned because their annual summary reports were conveniently accessible. While no attempt was made to make a comprehensive review of fish kill occurrences, these two state examples do indicate that discharge upsets sufficient to cause noticeable fish kills can be expected to happen with some regularity.

While fish kills are often diligently investigated by resource agencies in response to public outcry and to support compensation claims, quantifying recovery following kills is less exigent and gets less attention. The inability to locate many reports of recovery after major incidents was surprising, as future costs of monitoring recovery was sometimes one of the claimed costs mentioned in settlement news stories. For example, a fish kill in the Ogeechee River, Georgia extended 70 miles downstream from the outfall for an industrial facility. While the specific combination of causes leading to the large kill were unclear, the effluent contained industrial chemicals for which no aquatic life criteria have been developed, although the chemicals may be toxic to aquatic life. According to news reports, after 5-years, resource agencies consider the fish populations in the river to have recovered (Landers 2016). Similarly, a pretreatment failure at an industrial facility that in turned discharged to a municipal wastewater treatment plant, caused a plant upset and a subsequent severe fish kill in the White River, Indiana. The spill resulted in a complete kill for about 30 river miles downstream of the wastewater treatment plant, followed by about 20 river miles with a partial fish kill. 10-years had been considered necessary for long-lived native fish species to reach maturity after the spill, based on journalist's interviews with fisheries managers (Schneider 2010). A wastewater treatment process upset at a pulp and paper plant near Bugalosa, Louisiana resulted in a large fish kill in the Pearl River. This incident did result in rigorous post-spill monitoring to assess recovery of fish and mussel populations (LDWF 2015; Piller and Geheber 2015). Such published assessments were rare in comparison to the numbers of wastewater related fish kills that made the news and could be found in internet searching. For instance, a rendering plant drew the attention of state investigators after 6 discharges with fish kills in two years to Skipjack Creek, Pennsylvania; a power failure led to a municipal wastewater plant upset and a large fish kill in Sager Creek, Arkansas and Oklahoma; miscalculation with caustic soda in a small town sewage plant resulted in fish kill in the Red River, New Mexico; and excessive land application of hog waste killed fish for 20 miles in Beaver Creek, Illinois among many other related incidents (Jackson and Marx 2016). Some of these upset examples show that if the magnitude of criteria exceedence is great enough, a single exceedence can cause biological impairment that persists beyond 3-years (assuming that the substances discharged were regulated by aquatic life criteria).

Photos: (A.) wgno.com; (B.) LDFW 2015; (C.) L. Van Every, Idaho DEQ; Jackson 2018)

Figure 1. Water quality upsets can cause severe effects, at least in the short-term. (A) A large fish kill occurred on the Pearl River near Bogalusa, LA following an upset at an industrial wastewater treatment plant, August 13 – 17, 2011. (B) Recently dead mussels *(Leptodea fragilis)* washed ashore during the same incident (LDFW 2015); (C) a sudden hypoxic episode killed fish including this adult white sturgeon in the Snake River, Idaho, when oxygen levels dropped in a dam outflow, despite a system of oxygen sensors linked to automated reaeration blowers (Jackson 2018).

Figure 2. Factors affecting time to recovery following episodic pollution events include the severity of disturbance, persistence of disturbance, and condition of communities prior to the event. Here for example, a visually horrific spill to the Animas River, Colorado, followed a mishap in a mine remediation project. However, effects to the fish and benthic macroinvertebrate communities were slight. The benthic aquatic community was "protected" from further harm in part because sensitive species were already absent and diversity was low from chronic, low-level metals exposure. Factors protecting fish included the short-duration of the metals pulse (<24 hours for the worst of it) and that elevated metals during the pulse were predominately in particulate form, which have low acute toxicity (Roberts 2016a, b; White 2016). Photos: A. Leading edge of the plume moving down the Animas River; B. Caged hatchery fish were placed ahead of the plume; C. Most fingerling trout were still alive after 4 days. Photos from the Durango Herald (durangoherald.com)

Approach

Previous reviews

The present effort is one more in a series of efforts to synthesize the literature on recovery of aquatic ecosystems from pollution or other disturbance. Some of the previous efforts were substantial, increasing in scope from 4 case studies in 1971 to hundreds, and in some cases appear to represent multi-year efforts by a team of people (Cairns et al. 1971; Resh et al. 1988; Niemi et al. 1990; Detenbeck et al. 1992; Jones and Schmitz 2009; Kattwinkel et al. 2015; Gergs et al. 2016; Stanford et al. 2018). Reviewing these reviews shows a convergence of thought on the ecological factors behind long or slow recovery times, which is generally reflected in Table 1.

The most comprehensive of the previous reviews by far remains the 1990 series of 17 papers following Yount and Niemi (1990a) synthesizing ecological theory and case studies of recoveries of aquatic ecosystems following disturbances. A few of the important findings from subsequent analyses include Kattwinkel et al.'s (2015) observation that for invertebrates recovering through internal production (as in a lake), a maximum of five generation times were needed for full recovery. Thus, populations of animals with longer life spans take longer to recover than those with short life cycles. The length of a generation for lentic invertebrates ranges from days for some zooplankters to a year or longer for some insects. Thus, recovery times expressed in generations can result in highly variable recovery times when expressed in years. Recovery time also needs to be related to the time intervals between exposure events. Repeated unremarkable effects from low-dose pulses of exposure with minor effects may culminate in strong effects when successive pulses of exposure are present (Kattwinkel et al. 2015).

The review by Gergs et al. (2016) is probably the most comprehensive of the post-1990 analyses of recovery times for freshwater ecosystems following disturbances. They (as others) note that ecological recovery depends on complex processes related to species, population, ecosystem and landscape properties. They focused their synthesis on five aspects of ecological recovery in aquatic organisms: (1) The variability in recovery times among different taxonomic groups of freshwater organisms; (2) a comparison of recovery times across taxonomic groups and ecosystem types; (3) variability in recovery times among different types of disturbance for similar taxonomic groups; (4) a comparison of field and semi-field studies; and (5) the relationship between effect magnitude and recovery time.

Gergs et al.'s (2016) discussions on the limitations of their review are important to consider and are relevant to other reviews, including this one. Their main obstacle for the evaluation of field data from the literature on recovery was that studies were designed in different ways with different endpoints, taxonomic classifications, and reporting details. Most studies were not conducted for long enough to determine full recovery. Most micro- or mesocosm experiments conducted with pesticides had observation periods shorter than 6 months and consequently, could often not provide recovery times for taxa with longer life cycles such as Trichoptera and Plecoptera which have annual or longer life cycles. Another uncertainty in the estimation of recovery times relates to the start of the recovery process. They assumed that ecological recovery started from the timepoint at which the maximum effect occurred, although that did not necessarily mean that the stressor was always removed from a system and no longer causing effects. Also, a chemical stressor can quickly disappear from the water column but persist in the sediment. In this case, aquatic populations that predominantly inhabit the water column might show rapid recovery, whereas typical benthic organisms that are in contact with the sediment might show long-term effects from the residual toxicity (Gergs et al. 2016).

In the present effort, rather than attempting an ever more exhaustive review, the approach was to emphasize differences in recovery times in relation to the severity of disturbances. Because the motivation for this review is to consider ecological effects from episodic water pollution efforts which are usually non-catastrophic, an effort was made to contrast ecological recoveries following different disturbance severities, ranging from catastrophic to barely detectable.

Primary literature was searched using keyword and general internet searches on variants of aquatic recolonization and recovery, forward and backward cited reference searches from seminal reviews and analyses (Niemi et al. 1990; Detenbeck et al. 1992; Barnthouse 2004; Gergs et al. 2016), in addition to a body of literature compiled for previous work (Mebane et al. 2015).

The compilations and analyses can only be considered semi-quantitative, as both the disturbance classifications and time to recoveries required subjective judgments. Quantifying the time required for recovery was seldom the objective of the studies found. For example, a common study design would examine disturbance from a pesticide overspray shortly after application, and then upon re-surveying the next year, finding that the biological effects were no longer detectable. Whether in such cases, the recovery took a shorter time, or whether organisms with an annual life cycle simply require a full life cycle to get back in equilibrium cannot be teased out of such studies. In cases when recovery followed pollution control upgrades, the resulting improvements in the water quality conditions were never immediate. Improvements were often progressive and sometimes took many years. In some cases, watershed source controls are difficult and take time to achieve effective controls, and residual contamination pools in sediments or groundwater may take time to dissipate (Hamilton 2012). Deciding when to “start the clock” for biological recovery when the chemical recoveries are progressing is difficult and uncertain.

Reasons why some recovery studies were unusable in this review included: (1) biological conditions appeared to be substantially constrained by ongoing pollution and thus no judgement of recovery times could be made; (2) data were not presented as a time series; (3) no appropriate reference comparison was included; or (4) recovery only addressed chemical endpoints (Rabeni et al. 1985; Chadwick et al. 1986; Hoiland et al. 1994; e.g., Arnekleiv and Størset 1995; Lemly 1997; Smith 2003; Finley and Garrett 2007; Kowalik et al. 2007; Hornberger et al. 2009; Lefcort et al. 2010; Buys et al. 2015).

Table 1. Factors related to rapid or lengthy times to recovery after disturbance in freshwater ecosystems

Physical or chemical factors:	Expectation
Geographic Scale	Wide scale disturbances will have slower recoveries
Severity of disturbance	Less severe disturbances that leave some survivors will have faster recoveries than disturbances causing complete kills and recoveries rely on immigration.
Lingering exposures (repeated pulses or press)	If the disturbances are repeated or ongoing, there will be no recovery, only adaptation
Intactness of habitats and watersheds	Intact and connected habitats would make complete recovery more possible, but the time to recovery may be longer; Otherwise degraded habitats recovered faster because they likely had a "shorter climb" to get back to baseline (Stanford et al. 2018). In frequently disturbed systems, communities might be selected for life-history traits that facilitate rapid re-colonization. Recovery times from these systems might be under-protective for undisturbed communities (Gergs et al. 2016).
Ecological factors	
Baseline stability/disturbance frequency (e.g. intermittent stream community vs. large lake)	Water bodies subject to frequent disturbance, such as intermittent streams, flashy urban streams, or a chronic disturbed state from ongoing pollution or habitat alteration will have resilient communities that recover faster from disturbances than more stable or less disturbed systems (Lake 2000)
Proximity to source populations (metapopulation)	Disturbances to connected stream networks with refugia from disturbance with intact seed populations will recover much more quickly than headwaters or widespread disturbances (Sedell et al. 1990)
Life History traits	
Turnover rate: fast or slow life histories	Taxa with fast life histories, such as zooplankton or some warmwater, small bodied fish such as gambusia, kililish or fathead minnow will recover from population crashes faster than long-lived, less frequently reproducing taxa (Rose and Cowan 2000; Winemiller 2005; Kattwinkel et al. 2015)
Reproductive strategies: late bloomers or inattentive parenting slower to recover	Fish that actively parent their broods through nest building and guarding recover faster than broadcast spawners with no parental controls (Ensign et al. 1997)
Population state prior to disturbance: abundant and widespread or teetering towards extinction	Rare taxa may be less successful finding mates or get crowded out by more abundant taxa
Tough neighborhood? Competition may prevent returning to pre-disturbance state and may result in alternative, stable states	(Beisner et al. 2003; Hobbs et al. 2012; Webster et al. 2013)
Operational and Measurement factors	
How recovery is studied affects apparent recovery times, such as the rigor of the study and whether conducted in field or semi-field (mesocosm) settings.	Weak study designs with inappropriate reference conditions, insensitive measures, or low statistical power may fail to detect altered conditions and yield overly optimistic recovery findings (Mebane et al. 2019). Field studies tended to have longer recovery times than semi-field, mesocosm-type studies in part because of more limited communities in semi-field studies (Gergs et al. 2016).

Capacity of individual organisms to recover from chemical stress

Much of the aquatic disturbance and recovery literature is based upon biosurveys quantifying components of ecological communities. Ecological communities are made up of interacting populations of many different taxa, and the populations are made up of individuals of different sex and ages. This hierarchy of ecological communities and populations logically require that before the structure of a community can be affected by a disturbance, there has to have first been a population-level effect to at least one of the taxa in the community, and a population-level effect cannot occur without first affecting individual organisms. Thus, the concept of recovery of aquatic ecosystems from water quality stress starts with the capacity of individuals to recover from chemical stress.

With "conventional" characteristics of water that sometimes naturally vary greatly in flowing waters, such as dissolved oxygen or temperature, severe stresses can cause lethality within minutes (USEPA 2003; Cook et al. 2015). However, recovery from acute stresses also appear to be rapid. With rainbow trout, a hypoxia intolerant species, fish that survived a 4-hour hypoxic stress without losing equilibrium recovered within 6 hours after return to normal dissolved oxygen levels (Iftikar et al. 2010). Similarly, investigators have attempted to mimic daily temperature cycles in streams by exposing fish to cyclic high temperatures followed by cooler recovery periods. Findings have shown that when daily high temperatures were followed by a daily recovery period, fish can tolerate conditions that would be lethal within a day if maintained constantly (Hokanson et al. 1977; Schrank et al. 2003). This capacity for rapid recovery from near-lethal conditions is presumably an evolutionary adaptation to daily summertime conditions, especially in temperate waters.

Toxicity from chemical stress is presumed to result from accumulation that overwhelms detoxification and depuration mechanisms, until it reaches a critical accumulation that interferes with critical biochemical functions, which initiates cascading effects at the tissue and organ level, leading to death (Jager et al. 2011). In this context, measured environmental chemical concentrations can be thought of as multiple layers of surrogacy for the actual site of toxic action. For instance, concentrations in diet or water "causing" an effect, such as an EC50 concentration, are actually just correlated surrogate measures for organ or whole-body tissue concentrations. The organ concentrations, such as copper accumulated onto fish gill tissues, are in turn just surrogates for concentrations on the "true" sites of action, which are thought to be binding to ionoregulatory cells, disrupting homeostasis. In practice, the true mechanisms of toxicity may not even be known, even for well-studied substances and organisms, such as the chronic toxicity of metals (Wood 2012). Associated with all this surrogacy are lag times between exposure and effects, and lag times between ending exposure and recovery or death. For example, in some case studies with organic chemicals, these depuration and recovery times were in the range of 5 to 30 days (Landrum et al. 2004; Landrum et al. 2013).

With metals, toxicity occurs when the total rate of metal uptake from water and diet into an organism exceeds the combined rates of excretion and physiological detoxification. When toxicant accumulation at the site of action exceeds the organism's tolerance (critical threshold), they will not be able to reverse the damage and will likely die (McDonald and Wood 1993). However, even internal critical thresholds likely have a time aspect. For example, copper interferes with sodium regulation in fishes and when fathead minnow was exposed to copper, mortality was inversely related to whole-body sodium concentration. Larval fish had a significant reduction in whole-body sodium concentration when exposed to copper, reaching a critical threshold for mortality at about 30% sodium loss. However, mortality lagged behind whole-body sodium loss by hours to days after the end of the copper exposure and organisms demonstrated an ability to recover whole-body sodium even when they reached or exceeded the critical threshold. In contrast, by the time they lost 70% of their sodium, they had passed the point of no return and would always die. For the fish that recovered, about 4-days was required to return to previous vigor (Zahner et al. 2006; Zahner 2009).

Tissue accumulation/depuration – slow kinetics.

The slower kinetics of tissue accumulation suggest some differences between the dynamics of exposures of water exposures, in which chemical concentrations can rise and dissipate rapidly and the dynamics in internal organs. For the duration competent of criteria this has been addressed through longer, averaging periods which are less variable than shorter periods (see the comparisons of 30d vs. 4d averaging periods with selenium later in this report). In the case of selenium which has a tissue residue-based expression of the criterion magnitude, and for which water exposures are expressed as a secondary surrogate, the water exposures are defined as 30 day averages (USEPA 2016).

Depending on chemical, organismal, and environmental factors, recovery of individual organisms from harmful tissue accumulations range from days to years, and in the case of some highly persistent organic pollutants or mercury, tissue burdens may be carried for life. Some brief examples follow recovery from tissue accumulations, emphasizing fish studies with mercury and selenium, as these two substances have had tissue-based water quality criteria issued. The selenium criteria were expressly crafted for the protection of aquatic life, whereas mercury criteria are derived from to protect human health via acceptable concentrations for fish consumption.

Time to recover from tissue burdens

Often when fish exposed to elevated chemicals in laboratory exposures are then transferred to clean water for a recovery period, depuration of bioaccumulated tissue residues is slower than was the uptake. For example, tissue residues of Pb in brook trout increased exponentially in the first 2 weeks of exposure, eventually reaching an elevated equilibrium that was maintained for 2 years and caused neurological and developmental effects. When transferred to clean water after 2 years, Pb levels declined over 12 weeks by about 75% in various tissues, close to NOEC levels, but did not return all the way to control levels after 12 weeks (Holcombe et al. 1976). Similar experiments with Cd also suggested that some tissue residues were highly persistent. A 12 week recovery period was sufficient to clear Cd from the gills, but Cd continued to increase in kidneys, apparently from being translocated from other organs (Benoit et al. 1976). In contrast, with Zn, a 12 week recovery was sufficient to bring tissue residue levels in all organs to near control levels (Holcombe et al. 1979).

Recovery time for fish exposed to mercury in diet is much longer than other metals reviewed. For example, no mercury was eliminated by brook trout in a 12 week recovery period after 2-year mercury exposures (McKim et al. 1976). Because Hg is usually excreted at a lower rate than it is ingested, the Hg concentration in fish increases with age and size. The elimination of methylmercury (MeHg) in fish tissues is slower relative to other metals, half-lives range from 20-1,200 days, and tend to be longer in larger fish and in colder climates. Tissue elimination times were longer in field experiments (contaminated fish transferred to low Hg waters) than in lab experiments (Trudel and Rasmussen 1997; Van Walleghem et al. 2013).

Recovery times for fish exposed to elevated dietary selenium were shorter in laboratory feeding experiments than in quasi-natural experimental stream studies. In modeling uptake and depuration in aquatic food webs with fathead minnow or bluegill, DeForest et al. (2016) projected Se concentrations to return to background after about 100 days in several scenarios. Hardy et al. (2010) exposed cutthroat trout to elevated dietary selenium for almost 2 years, followed by a 32-week depuration period. Se concentrations and body burdens decreased in all groups during the depuration period, but the higher the initial Se concentration and body burden, the greater the decrease Half-life (time for tissue burdens to drop by 50%) ranged from about 12 weeks in the fish that had received the highest doses to 74 weeks for lower-dosed fish (Hardy et al. 2010).

Hermanutz et al. (1996) exposed bluegills to 10 µg/L waterborne selenium for two years in experimental streams. The fish accumulated an average of 22 mg/kg dry weight whole body selenium, and 96% of their progeny were deformed with edema and 50% had hemorrhaging, versus 0 to 1% in controls. The selenium inputs to the streams were stopped and the recovery of the streams was monitored for another year. The following year bluegill whole body selenium concentrations had dropped to an average of 10.4 mg/kg dw with 0% deformed progeny (Table 2; Hermanutz et al. 1996). Declines in selenium tissue concentrations were slower than would be expected from laboratory experiments or modeling approaches because after the selenium dosing to the streams stopped, the sediments, detritus, and macrophytes changed from being a sink of selenium loading to the streams, to becoming sources of loading. The selenium concentrations of the aquatic invertebrate prey of the bluegills likewise slowly declined. The dynamics of selenium declines in the streams were described by Swift et al, (2002). The companion Hermanutz et al. portion of the study describing the occurrence and then elimination of adverse reproductive effects to the fish after exposure and recovery periods unfortunately has yet to be formally published. Thus, some of their pertinent data are shown here in Table 2.

Table 2. Recovery of selenium-exposed bluegills in experimental streams after 1-year

Target treatment	Whole body Se (mg/kg ww)	Ovary Se (mg/kg ww)	Whole body Se (mg/kg dw)	Ovary Se (mg/kg dw)	Deformed progeny (%) Hemorrhaging	Deformed progeny (%) Edema
After 1-year continuous exposure to selenium						
Control	0.39	0.8	1.6	3.2	1.4%	0.0%
2.5 µg/L treatment	1.23	2.5	5.1	10.3	14.9%	2.8%
10 µg/L treatment	5.29	10.1	22.0	42.1	50.0%	95.5%
After 1-year of recovery with no added selenium						
Control	0.46	1.1	1.9	4.4	0.4%	0.0%
2.5 µg/L treatment	1.06	1.9	4.4	7.9	0.0%	0.0%
10 µg/L treatment	2.6	3.9	10.8	16.3	0.0%	0.0%

Data from Hermanutz et al. (1996) using data from the lower ("down") spawning pools, dry weight values were estimated assuming a moisture content of 76%.

Recurrence interval for tissue burdens in fish

For substances such as selenium with relatively slow kinetics (slow to accumulate and slow to depurate), it follows that when evaluating criteria expressed in terms of tissue residues that impair reproduction, a longer exceedence return interval than that that applies to water concentration exceedences should be considered. For example, in the bluegill recovery study (Table 2), the highest treatment resulted in 96% deformed offspring, which effectively would have caused an entire year class loss. While after a year's recovery, the tissue concentrations dropped by ~50% and deformities in fry disappeared, a substantial fraction of the fish's reproductive potential was lost. In the 2016 national recommended Se criteria, USEPA (2016) concluded that the usual 3-year return interval for criteria exceedences was too frequent and could lead to sustained ecological impacts, considering the long recovery times of Se contaminated reservoirs. However, no alternative allowable return frequency was indicated, other than beyond stating that the allowed exceedence frequency was that they were "not to be exceeded." With Hg, some states have set de facto exceedence frequencies through data aging policies for issuing fish consumption advisories, in which data have to be less than 5 or 10 years old, for example (Keller and Cavallaro 2008).

One idea for setting an allowable exceedence frequency for fish tissue-based criteria such as Hg and Se that affect reproductive endpoints, is to avoid more than one exceedence during the typical reproductive life span of resident fishes. However, because some fishes can live a very long time with their maximum life spans exceeding those of humans (Figures 3 and 4), linking allowable exceedence frequencies directly to fish life spans would be impractical. Still, long-lived fishes and other vertebrates are generally at higher risk of decline and extirpation than fish with short-life cycles (Birstein 1993; Rowe 2008). Thus, conceptually, repeated sublethal episodic pollution stresses could contribute cumulative stresses and reduced vitality over long lives, and minimizing these would be beneficial. An exceedence frequency approach that only allows infrequent exceedences specifically for long-lived fish might be more practical than expressing long exceedence return intervals >10 years in criteria.

Directly linking the frequency of allowable fish tissue exceedences to fish life spans would be further complicated by the fact that different populations of the same species can have very different life histories. Fish living in nutrient rich, warm waters tend to live fast and die young, compared to those living in nutrient poor and colder waters. For example, walleye in Big Trout Lake, Ontario matured at 8 years and had a maximum lifespan of 20 years, but walleye in Canyon Reservoir, Texas, matured at age 2 and only lived to 3 years (Beverton and Holt 1959). Similarly, most brook trout from Hunt Creek, Michigan reproduce by 3 years and seldom live past 4 years (McFadden et al. 1967), but the large, lake-dwelling coaster brook trout variety regularly reach 8-years (Huckins et al. 2008). At the margins of habitability, brook trout in a slow growing population reached 24 years in Bunny Lake, California, which is a high (10,900 ft), cold, alpine lake (Dill and Cordone 1997). Figure 3 gives further examples of this plasticity in life histories. A related complication is that recent advances in aging animals has shown that some freshwater species have much longer life spans than had been previously estimated. For example, the Bigmouth Buffalo, *Ictiobus cyprinellus*, was recently shown to reach at least 112 years in age, when previous estimates were only about 25 years (Lackmann et al. 2019).

The distribution of a large set of fish longevities, from reports around the world show about 50% of all taxa have maximum life spans of 10 years or longer. In this comparison, both extremes are from marine environments: the Pygmy goby with a maximum lifespan of less than 2 months and the Greenland shark which, at 400 years, is the longest-lived animal known. (Figure 4). A more limited view of some longevities of North American freshwater fish populations show that about half the taxa plotted have longevities of 5 years or more (Figure 3). These comparisons suggest that if the life spans of fishes were relevant to setting the allow return frequency of tissue-based criteria exceedences for substances that primarily impaired reproduction, then it would be desirable for exceedence return intervals to be no shorter than 5 years or as long as 10 years. Return intervals much longer than 10 years might be difficult to apply in criteria, permitting, or assessment programs.

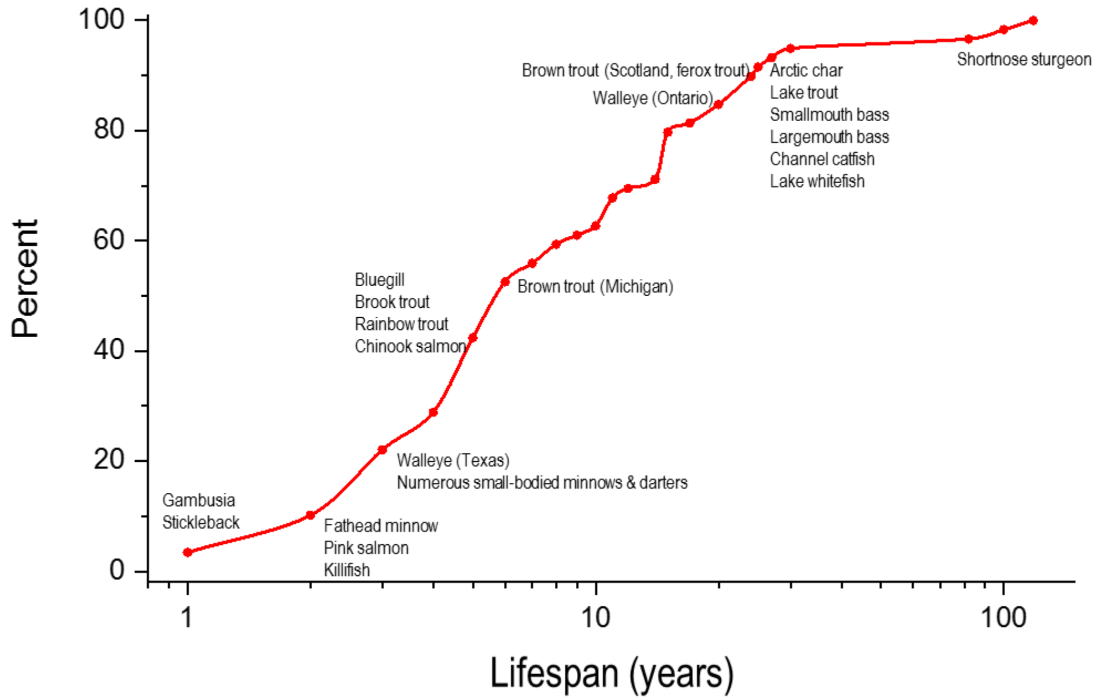


Figure 3. The lifespans of the same species of fish may vary greatly, with those occurring at the colder regions of their range slower growing and longer lived. Primary sources: (Beverton and Holt 1959; Scott and Crossman. 1998; Mangel and Abrahams 2001).

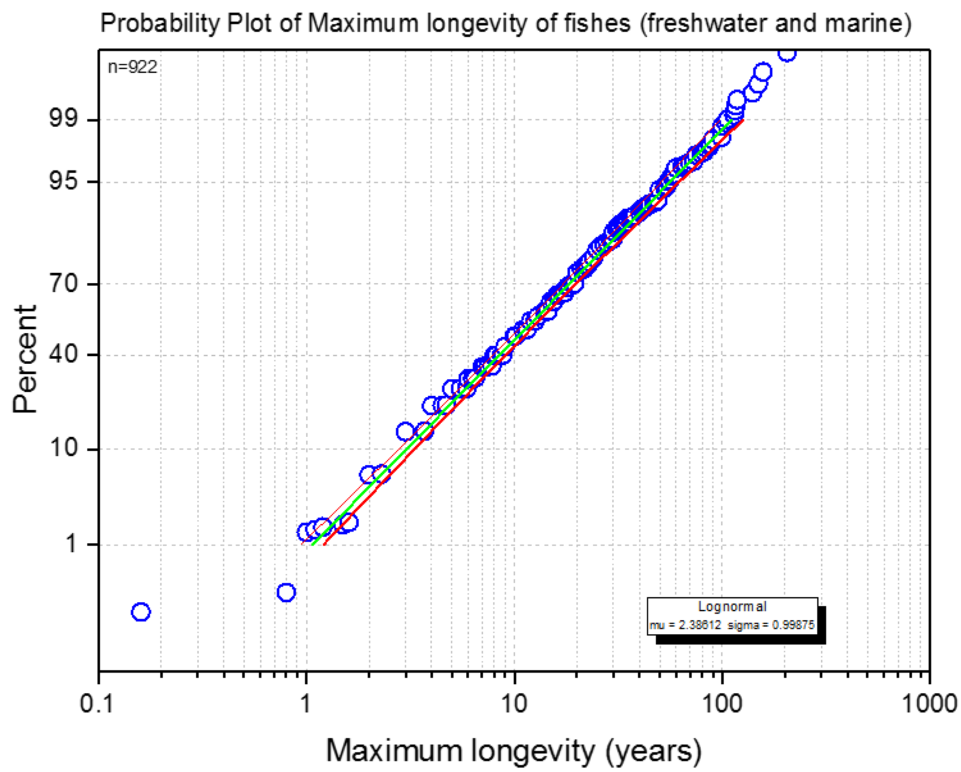


Figure 4. Across all available taxa (922 records), 50% of all fish have maximum longevity >10 years (source: *AnAge, the Animal Ageing and Longevity Database* http://genomics.senescence.info/species/)

Recovery of aquatic communities

About 70 case studies relevant to the recoveries of freshwater aquatic communities from disturbance are summarized in Appendix 1. An objective in searching for, selecting, and summarizing these case studies was to contrast recovery times by disturbance severity. There was no attempt to construct a cumulative list of case studies that compiled all the preceding reviews mentioned earlier. In summarizing these studies, many subjective judgements were required. If authors reported recovery times, those were used. Otherwise, estimates of recovery times could be inferred from data showing population or community metrics over time. In some important case studies, such as recoveries from whole-lake manipulations, graphs were digitized in order to make the estimates (Mills et al. 2000; Blanchfield et al. 2015). Most studies produced multiple endpoints, and only a few endpoints per study were used in plots.

The main object of this review was to evaluate recovery times that might result from frequent, low-level disturbances that are considered most relevant to aquatic life criteria definitions. Thus, relatively minor disturbances, including non-catastrophic deliberate poisonings and spills were distinguished from more severe effects, such as acid rain which eliminates recolonizing refugia areas as well or widespread dewatering from droughts that killed or displaced almost all instream fauna. Each case study was classified as either being a "pulse" or "press" disturbance (Bender et al. 1984; Yount and Niemi 1990b; Lake 2000), where pulse disturbances are more like those from a wastewater plant upset and press disturbances are more representative of a stream with multiple, persistent discharge sources or the long-recoveries expected after long-term watershed disturbances such as some mining disturbances. In an attempt to make these disturbance classifications less subjective, the disturbance severity ratings of Gore and Milner (1990) were simplified and were assigned to each case study. These were:

- Level 1: Disturbance completely destroys communities along the entire stream length leaving no upstream or downstream sources to colonize and may result in a new stream channel.
- Level 2: Disturbance completely destroys communities in a reach of stream but upstream and downstream colonization sources or hyporheic zone refuges remain, leading to succession and faunal organization. Examples: severe chemical spills, reclaimed or diverted river channels, surface mining effects, intermittent streams
- Level 3: Results in reduction of species abundance and diversity from pre-disturbance levels in a section of stream but does not completely eliminate the benthos. Examples: incomplete kills from application of insecticides, chemical spills, wastewater treatments plant malfunctions, floods, chronic nonpoint and point pollution sources including nutrient enrichment.
- Level 4: Results in reduction of species abundance and/or diversity or loss of benthos compared to predisturbance levels in discrete patches within a stream section but such that proximal patches are virtually unaffected. Examples: sediment inputs from highway construction, logging, and introduced substrata, localized fish kills (after Gore and Milner 1990).

The efforts to group case studies in this way were far from perfect. Few case studies reported the factors in Gore and Milner's conceptual scheme, and the "completely destroyed communities" test was not interpreted as requiring the annihilation of all life, since almost no disturbance eliminates all aquatic life. Despite these limitations, several insights can be drawn from the reviews:

Benthic invertebrate communities recovered faster than fish communities in most of the case studies reviewed. In Figure 5, the recovery times are plotted as cumulative distributions without regard to disturbance severity. About 80% of the invertebrate community endpoints and about 50% of the fish endpoints were recovered within 3 years. The faster recovery times are likely linked to shorter life cycles of most invertebrates than fish and the aerial dispersal ability of insects. As major caveat to this generalization of faster recovery times for invertebrates is that few reports of recovery times for long-lived aquatic invertebrates, such as native mussels or some crayfishes, were included in the case studies.

Recovery times for non-catastrophic pulse disturbances, that is, those case studies classified as level 3 or 4 disturbances, did tend to be faster than more severe pulse disturbances or press disturbances (level 1 and 2 disturbances). The median recovery times for the less severe category 3 and 4 disturbances was 1-year, ranging from less than a month to 10-years (Figure 6).

A plot of recovery times was consistent with a lognormal distribution. Assuming recovery times complied in Appendix A are a representative sample of ecosystem recoveries, probability of recovery within a given time predictions can be made (Figure 7). These probability plots suggest that ecosystems subject to non-catastrophic category 3 or 4 disturbances have a high (88%) chance of recovery within 3-years. If the dataset were expanded to include the category 2 disturbances, which were often press disturbances with long-recoveries, then the distribution would shift toward slower recoveries and reduce the overall probability of recovery within 3-years to 71%. The reductions of point and nonpoint pollution sources following Total Maximum Daily Load (TMDL) development or other watershed restoration efforts can take a long time (see for example the Pigeon River, appendix 1) of many polluted waterbodies are subject to more severe, category 2 press disturbances, and the less optimistic center plot in the Figure 7 probability plots including category 2-4 disturbances might not be unrealistic.

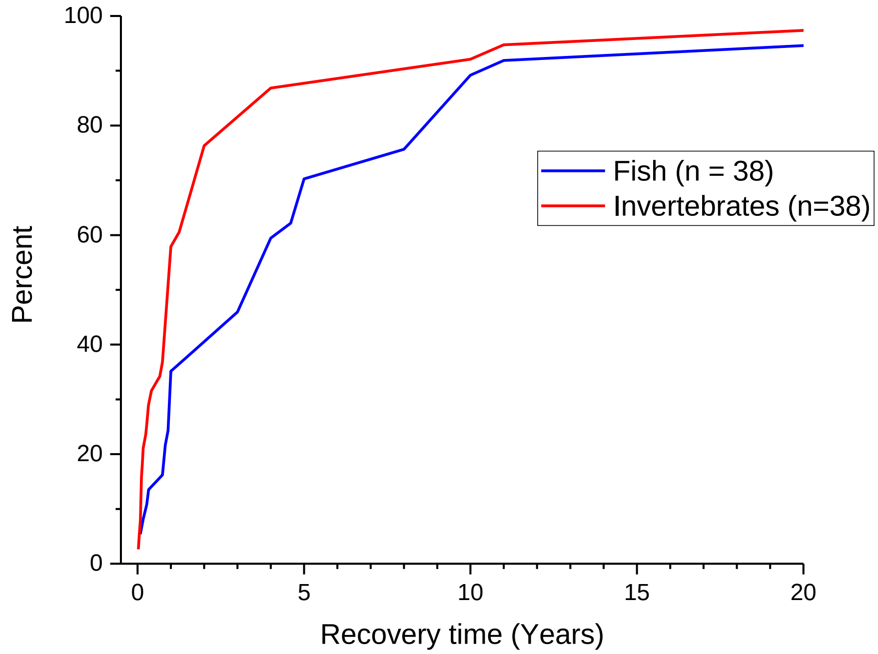


Figure 5. Cumulative distribution frequency of times to recovery, grouped by invertebrate community endpoints or fish population or community endpoints. Most studies reported multiple endpoints per assemblage. Only the longest endpoint per assemblage is plotted.

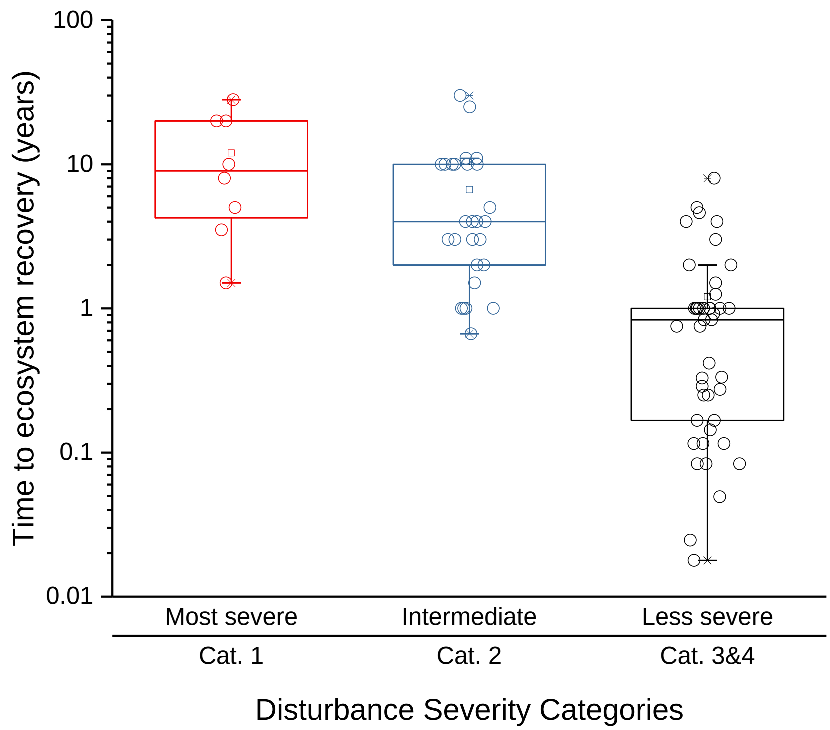


Figure 6. Aquatic ecosystem recovery times, grouped by severity of disturbances. Boxes span the data from indicate the 25th and 75th percentiles, the line through the box shows the median, and the lines above and below the boxes show the 5th and 95th percentiles of the data.

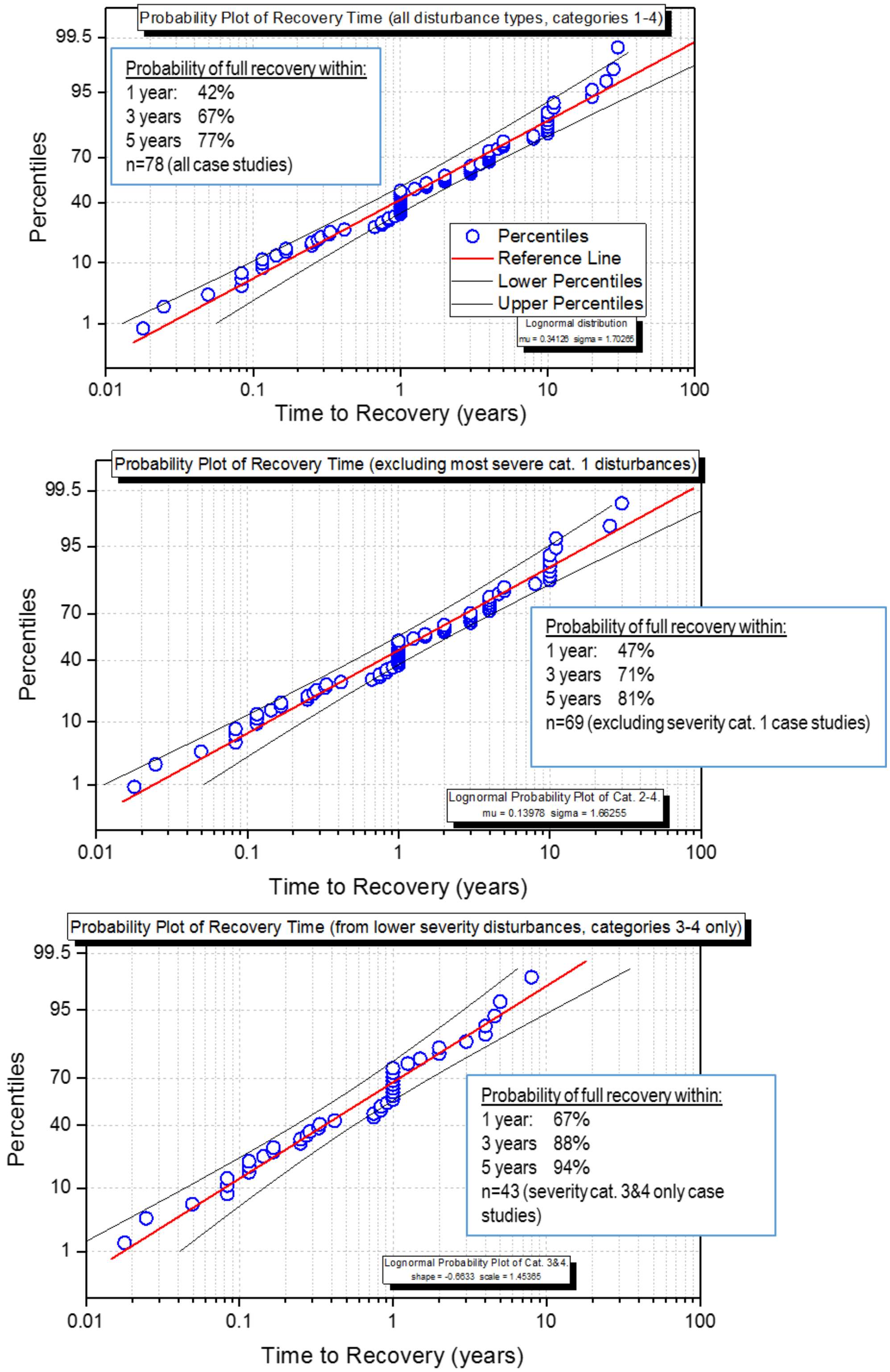


Figure 7. Probabilities of full ecosystem recovery within a given time period, estimated using a lognormal probability distribution. If only the milder category 3 and 4 level disturbances are considered, then there was an 88% probability of full recovery within 3 years. Including long-term "press" disturbances or local catastrophes in the datasets decreased the probability of full recovery within 3-years to 67%.

The fundamental opportunities and constraints of recovery have been well described earlier, and this review located no new case studies or insights that would fundamentally challenge the well-established body of knowledge on factors influencing recovery of aquatic communities from disturbances (e.g., Table 1).

For disturbances of moderate scale and magnitude and nearby recolonization sources, aquatic ecosystems can generally recover quickly, within 3 to five generations. Even ecosystems that were disturbed by severe spills that were visually horrific and can produce high body counts of dead fish recover quickly and completely, if given the chance and not subjected to repeated disturbances. Prominent examples include the spills and fish kills in the Ogeechee, Pearl, Animas, and Rhine Rivers (appendix 1). However, a caveat to this optimistic view is few reports that quantitatively monitored recovery of fish kills from wastewater discharges (usually resulting from "upset" conditions) could be found. The recovery reports often consisted of news articles citing either state fisheries managers or knowledgeable "riverkeeper" activists. The Pearl River incident resulting in a wastewater treatment upset and bypass causing a 70-mile fish kill was a notable exception. Fish and mussel assemblages were rigorously monitored for 3 years post-spill for evidence of recovery and need for intervention (appendix 1)

One impression from review of the case studies is that of the various factors important to recoveries (e.g., Table 1), the particular toxicant does not stand out as being important. For episodic disturbances, it probably does not matter much for a population biology or community ecology perspective whether organisms are killed by heat from forest fires, hypoxia, rotenone, fuel oil, insecticides, or metals for example. Once the episodic stressor is relaxed, recoveries are constrained by reproduction capacity and ecological context more so than whatever the particular acute stressor was. Superficially, this impression might appear to differ from previous reviews. For example, Gergs et al. (2016), in their figure 5, show a compilation of recovery times of lotic macroinvertebrates in which the median recovery time after metals exposures was about twice as long as for pesticides (1 year vs. about 6 months). However, as they do explain, such compilations mix the issue of toxicity characteristics of the substance with the exposure times.

Most of the recovery case studies with pesticides follow the effects of one or more pulse exposures, which are usually intended to mimic the seasonal applications of pesticides in the environment. Pesticide studies provide some of the best evidence for recovery from chemical toxicity from mild to severe pulses and contrasting single disturbances from repeated disturbances that might result from yearly applications. For example, in year 1 of a yearly spray program, reduction in sensitive groups, such as Plecoptera and Ephemeroptera, could lead to reduced recruitment in the following year. Addition of insecticide in the second season could further reduce populations and so on until they are selectively eliminated from the stream. Similar dynamics would be expected from any chemical or disturbance that causes acute toxicity – if repeated during vulnerable stages in the life cycle, persistent effects would be expected.

Recovery studies involving metals mostly observed relatively long recovery times of 1 to >10 years (Nelson and Roline 1996; Clements et al. 2010; Mebane et al. 2015; Herbst et al. 2018). However, these long-time periods reflect incremental source reductions to highly polluted waters in the case studies or "slow bleeds" from metals reservoirs in contaminated sediment, not some inherent characteristic of metals toxicity. For example, in a stream in which copper concentrations were progressively reduced by 3-orders of magnitude from those sufficient to effectively "chemically autoclave" the stream to concentrations that were mostly less than criteria, invertebrate richness steadily stepped up in synchrony with declining copper (Mebane et al. 2015, their figure 6). Recovery times of algal, insect, or fish communities from short-term exposures to metals at acutely toxic conditions were on the order of weeks or less (Effler et al. 1980; Clearwater et al. 2011; Roberts 2016a).

Recovery of ecosystems affected by a food web loading toxicant with slow effect dynamics (such as selenium) may fundamentally differ from faster acting substances. Source pools in sediments may provide ongoing sources after external loadings are reduced. The highly selective nature of effects could obscure the ability to quantify both impairment and recovery. For example, benthic community metrics might indicate a robust invertebrate community following a disturbance, but may not provide the resolution needed to adequality characterize effects to a minor subset of sensitive species that may not have recovered. For instance, amphipods and baetid mayflies were insensitive to selenium and thrived at the expense of sensitive isopods which were severely suppressed in the 10 µg/L treatment and completely extirpated in the 30 µg/L treatment (Swift 2002). Since amphipods such as *Hyalella* are often the "duty invertebrate" serving as a surrogate for all benthic invertebrates in species-sensitivity distribution, judging impacts and recovery on amphipods or other commonly (but not always) sensitive species alone, would obscure conclusions about impairment and subsequent recovery.

The slower recoveries of ecosystems owing to residual contamination bears careful consideration. If periodic exceedences were of a frequency and magnitude to result in sediment and food web contamination, then recovery times would be much longer than in some of the laboratory exposures discussed earlier. Swift (2002) and Hermanutz et al. (Table 2) found that removal of the selenium source resulted in gradual recovery of the ecosystem. Selenium residues decreased slowly in the sediments, plants, macroinvertebrates, and fishes over the two or three years studied after selenium dosing ended. However, macroinvertebrates and plants from all three treatments still contained enough selenium in their tissues to be potentially hazardous to fishes two to three years after selenium dosing ceased. There was also little recovery of sensitive macroinvertebrate population in the three years after Se input stopped. The concept of "recovery" must be used carefully when describing changes in ecosystems (Swift 2002). These slower dynamics suggest fish tissue-based criteria for pollutants with slow kinetics and accumulation/depuration dynamics such as selenium, longer return intervals for allowable exceedences may be appropriate, relative to faster substances such as chlorine, ammonia, copper, or zinc.

Overall, the ecological recovery literature reviewed here does not refute EPA's (1985) and Stephan et al.'s (1985) judgements that most aquatic ecosystems can tolerate water-based criteria exceedences once in a three year period, assuming that the magnitude of exceedence is <2X the criterion and the water bodies are not subject to anthropogenic stress other than the exceedence of concern.

Modeled population recoveries

There is a rich body of literature on recovery of aquatic ecosystems from disturbance, and the present search efforts hopefully made an incremental contribution to this body of knowledge. However, there are obvious limits in the serendipitous discovery of field studies that just capture potential scenarios relevant to evaluating regulatory wastewater management guidelines.

Population modeling has been considered a substitute for observable data in recovery evaluations (Benson et al. 2003; Barnthouse 2004; SAB 2006; Delos 2008; e.g., Baldwin et al. 2009).

Practical difficulties in field-based recovery endpoint measurements include widespread, low level degradation such that even reference sites have some level of stress. Population and community data are often noisy, and the ability to quantitatively measure disturbances in situ is often limited to effect sizes of 20% of reference or larger (Ham and Pearsons 2000; Dauwalter et al. 2009; Munkittrick et al. 2009; Jones and Petreman 2012), although such rules of thumb are endpoint specific and there have been exceptions with smaller detected effect sizes (Mebane et al. 2015). Ironically, the ability to detect changes is stronger for common species of least concern than it is for rare and highly valued fish species (Maxwell and Jennings 2005). Because of these limitations of real data, population modeling can be resorted to compare recoveries of populations under different scenarios. For example, models can be used to contrast the recoveries of different species with different life histories to the same disturbance, or the potential effects of different frequency or magnitude of disturbances can be contrasted.

Possible modeling approaches are many and disturbance scenarios are limited only by the imagination and time required to construct and interpret models. Here, following the illustration by O’Connor (2001), age structured models in Leslie matrices are used, with episodic pollution disturbances were imposed on several species populations representing a range of life histories followed by different recovery periods. The model scenarios are intended to represent episodic, acute toxic pollution events, where the pulse disturbance acutely affects the population and dissipates with no lingering, chronic effects. Thus, the population takes a hit, but then survivors suffer no lingering effects. One exception included a scenario with chronic, low-level effects which is then overlain by periodic exceedences.

The usual paradigm is that populations are controlled by density-dependent factors which act to slow and then halt excessive growth, but can be conversely eased up to cushion any population decline. This prevents population explosions or crashes, allowing the whole system to persist. This is a plausible mechanism and popular with modelers, but hard to prove or mathematically define with confidence. As a result, population modeling exercises often debate whether models should include a density-dependence factor to dampen populations swings, and if so, what is most appropriate? A population that is exactly in balance with every death replaced by one life, has a population growth rate, called lambda (λ) of 1. Given enough iterations, a density independent population will eventually explode if $\lambda>1$ or go extinct if $\lambda<1$. A further disadvantage for exercises like the present one in which the point isn’t to make real population projections, but to compare alternative scenarios, is that any level of effect imposed on individuals, such as a one percent acute effect (i.e.,EC01) will carry through the math to the population level. In practice, this limitation can be sidestepped at least for shorter projections by including stochasticity in the modeling, where for example, the standard deviation of survival or reproduction rate estimates is resampled using a Monte Carlo approach that produces uncertainty bands.

Density-dependent models will tend to dampen effects of acute pollution events or any other action that kills or removes individuals from the populations or reduces the reproductive capacity. For example, if a pollution event, storm, dewatering, fishing or whatever killed half of a year class, the survivors would be better off than they were before as a result of greater resources and less competition. With less competition for food and shelter, the survivors and would likely grow better, survive to graduate to the next year class at higher rates, and have better reproductive success. This principle is fundamental to the management of exploited fish and wildlife populations at sustainable levels, and drives harvest quotas (Myers et al. 1999). Of course, sometimes these efforts get it spectacularly wrong, as with the collapse of fisheries managed for maximum sustainable yield (Peterman and M'Gonigle 1992; Eagle and Thompson 2003). Ginzburg et al. (1990) describes the uncertainly associated with defining density compensation even from rich time series fish population data, and pessimistically (or perhaps cynically) noted that "*by choosing the model of density dependence carefully, one can achieve any quasiextinction risk desired.*" Nevertheless, while some animals are rare and persist far below their habitat carrying capacity limits, for many populations, density compensation is real.

Five species populations were modeled under different exceedence return interval scenarios to contrast potential population-level effects of different criteria exceedance frequencies (Table 3). The species modeled (amphipod *Hyalella azteca* and the fishes fathead minnow, brook trout, Chinook salmon, and shorthead sturgeon) were selected because they represent a range of life spans (1 to >60 years), these or closely related taxa often appear in criteria species-sensitivity distributions, and necessary vital rates needed for population modeling could be readily obtained. Four of the 5 models were modified from previously published models, and one (fathead minnow) was developed anew. To attempt some realism, the models used were all derived from life history studies of natural populations. No life table experiments or other laboratory-based population studies were used. Calculations were made using the software package RAMAS MetaPop v6 (Applied Biomathematics, Setauket, NY). Model parameter information sufficient to reproduce the results is given in Appendix 2.

Density dependence was assumed to fit one of two common approaches to modeling density dependent compensation: scramble or contest competition, depending on the species modeled. Scramble and contest competition represented in the population models are represented by different replacement curve functions. In a scramble competition model, all organisms have equal access to resources, and all suffer alike at high densities. Contest competition has winners and losers, where some animals win contests for food and shelter and do not suffer at all when resources are limiting, whereas losers suffer disproportionately. The amphipods were assumed to follow a scramble density model. The fathead minnow, brook trout, and Chinook salmon were assumed to follow a contest density model since these have territorial behaviors where a dominant fish will occupy and defend prime habitats. The small population of shorthead sturgeon was assumed to be far below its resource limited carrying capacity and no density dependent adjustments to vital rates were included (i.e., it follows a density independent model). While these assumptions seemed logically and mathematically plausible, they are just assumptions and they have strong influence on the modelling results and interpretations. For example, simply changing the fathead minnow density dependence model from contest to scramble and keeping the carrying capacity the same increased baseline extinction risk from 0% to 42%. While a 42% baseline extinction risk seemed highly implausible for a famously resilient species, it points out the uncertainty inherent to these "thought experiments."

Results of the modeled simulations of different episodic pollution return intervals are plotted in Figures 8 -10 and selected results of relative risks of population declines or extinction are tabulated in Table 4. For 3 of the 5 scenarios (amphipod, fathead minnow, and brook trout), the populations recovered quickly after fairly severe episodic “pollution” hits that mathematically killed 50% of all age classes. While population trajectories (topmost graph in each figure) show the reductions after each disturbance, the risks of population extinction or a severe decline (defined as a 90% reduction from initial abundances) remained very low and were not appreciably higher than baseline. The Chinook salmon model and results were more complex because of their anadromous life cycle and overlapping temporal “subpopulations” (Figure 9). The episodic “pollution hits” only affected the freshwater resident juvenile stages, whereas the adults were safely out to sea.

Counterintuitively, the results of imposing a 50% mortality every other year on the juveniles of this threatened Chinook salmon population only moderately increased the risk of year class loss or extinction. This is because of the elevated baseline extinction risks resulting from relatively high and variable mortalities experienced by outmigrating juveniles on the ~700 mile migratory corridor through 8 dams and reservoirs. The influence of high baseline extinction risk overwhelmed the influence of survival in the freshwater nursery stage. The one in 3-year exceedence scenario had little effect on projected risks of extinction or severe declines, assuming exceedance events acutely affected 50% of exposed individuals.

The same episodic 50% mortality scenarios that the short-lived fathead minnow and brook trout populations shrugged off decimated the small population of long-lived shortnose sturgeon (Figure 10).

Table 3. Models used to simulate population-level effects and recoveries from episodic aquatic life criteria exceedences

Species	Species and population characteristics	Setting	Principal sources of model population demographics (1)
Hyalella azteca	Small-bodied, short-lived (~1-year), benthic crustacean with capability for very high population growth; rapid recovery from natural disturbances has been observed in field studies (Strong 1972; Mebane 2006).	Composite life history and demographics compiled from several north-temperate studies. Initial abundances are near carrying capacity. Scramble density dependence assumed (baseline lambda (λ) 1.05; Rmax 2.90)	(Mebane 2006, 2010)
Fathead minnow	Small-bodied, short-lived (~2 year), fish with very high reproductive capability; rapid recovery from catastrophes has been observed in field studies (Mills et al. 2000; Danylchuk and Tonn 2003).	Population demographics represent an exponential growth phase measured during the rebound from a severe winter die off to ~1% of carrying capacity. Contest density dependence assumed (baseline λ 4.96; Rmax 5.0)	(Payer and Scalet 1978; Vandenbos et al. 2006; Divino and Tonn 2007)
Brook trout	Fish with extremely plastic life histories capable of rapid population establishment expansion under favorable conditions, and rapid recovery after local catastrophes. Depending on conditions, may reach sexual maturity between <1 or >16 years of age (Dunham et al. 2002). Often forms high density, stable populations of stunted size adults in lakes and streams.	Stable, abundant population at its carrying capacity. Contest density dependence assumed (baseline λ 1.03; Rmax 1.88)	(McFadden et al. 1967; Ferson and Ginzburg 1996)
Chinook salmon	Migratory, anadromous fish with a rigid 4 or 5-year life cycle that effectively has temporal life stage subpopulations that may buffer against catastrophe: at any given time, part of the population is in freshwater and part is out to sea.	Growing stream population, recovering from near-extinction; strong density dependence owing to limited carrying capacity; highly variable year-to-year survival rates results in high baseline extinction risk (Marsh Creek, Idaho). Population in moderate growth phase. Contest density dependence assumed (baseline λ 1.31; Rmax 1.78)	(Mebane and Arthaud 2010)
Shortnose sturgeon	Long-lived fish that only becomes sexually mature at 10 years and lives to 60-years.	Population is small, stable, and apparently well below its habitat carrying capacity. Density independence assumed (baseline λ 1.01)	(Root 2002)

Lambda (λ) is the finite rate of increase per time step (year except generations for Chinook salmon). A population with λ of 1.0 is stable, neither growing or declining. A population with a λ of 2.0 is doubling in size with every time step; a population with a λ <1 is decline and will eventually go extinct; a population with λ >1 will eventually become resource limited and stop growing.
Rmax is the maximum growth rate per time step for density dependent populations that are not constrained by their habitat carrying capacity.

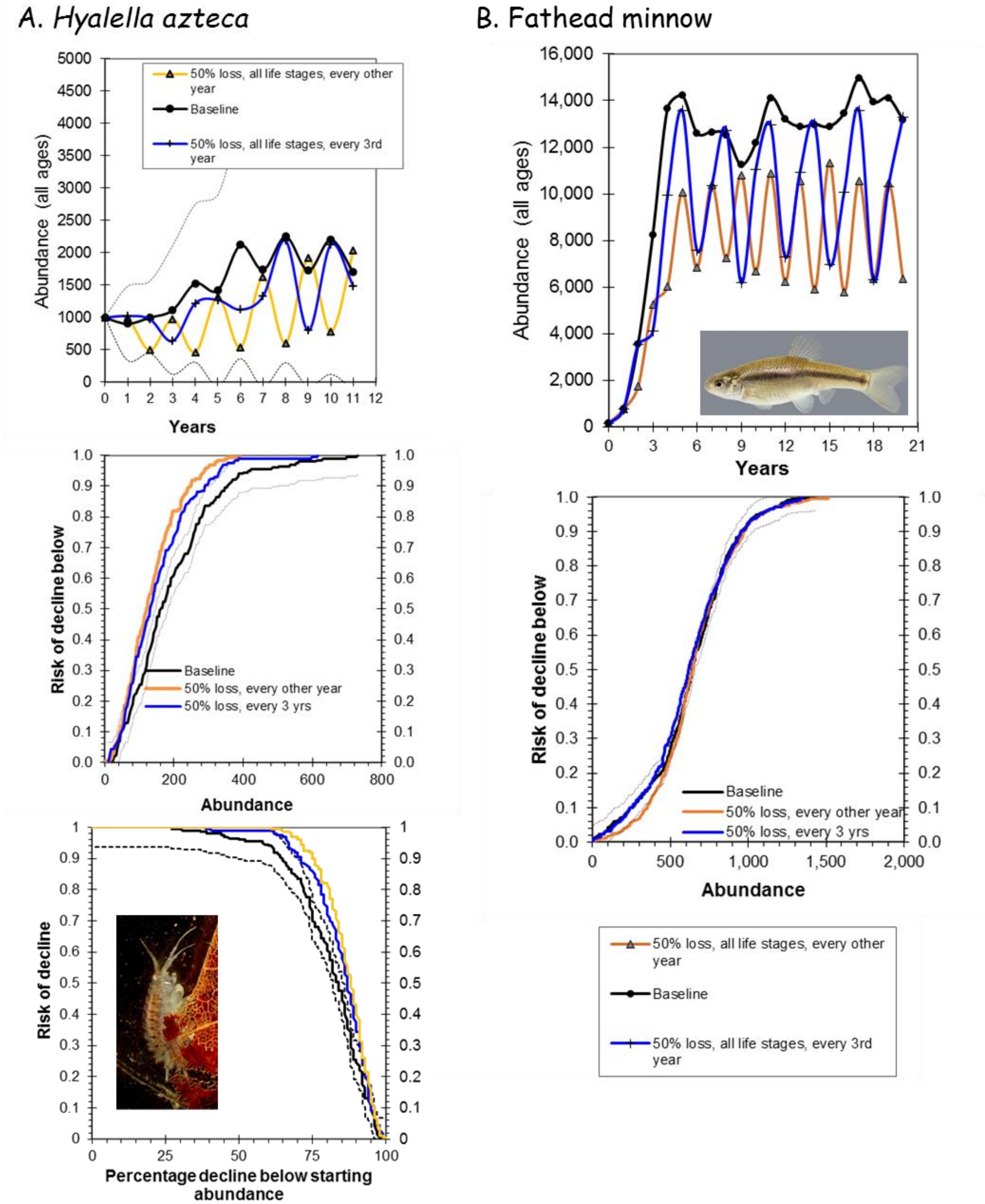


Figure 8. Population modeling projections of scenarios where an episodic pollution event killed 50% of the amphipod *Hyalella azteca* and fathead minnow populations. The scenarios killed 50% of each life stage either every second year, or every third year. In the top row, abundances are projected over time. In the second row, the risks of the population dropping below a given abundance are calculated, and in the bottom row, for *Hyalella*, the risk of the population decline below its initial abundance is plotted. Because the fathead minnow scenario began with a very small population following a near extirpation, risks of decline below starting abundance are not shown. The dotted lines show ± 1 S.D. in the population trajectory plots over time (top) and in the middle and bottom risk of decline plots, the dotted lines show the 95% Kolmogorov-Smirnov confidence limits.

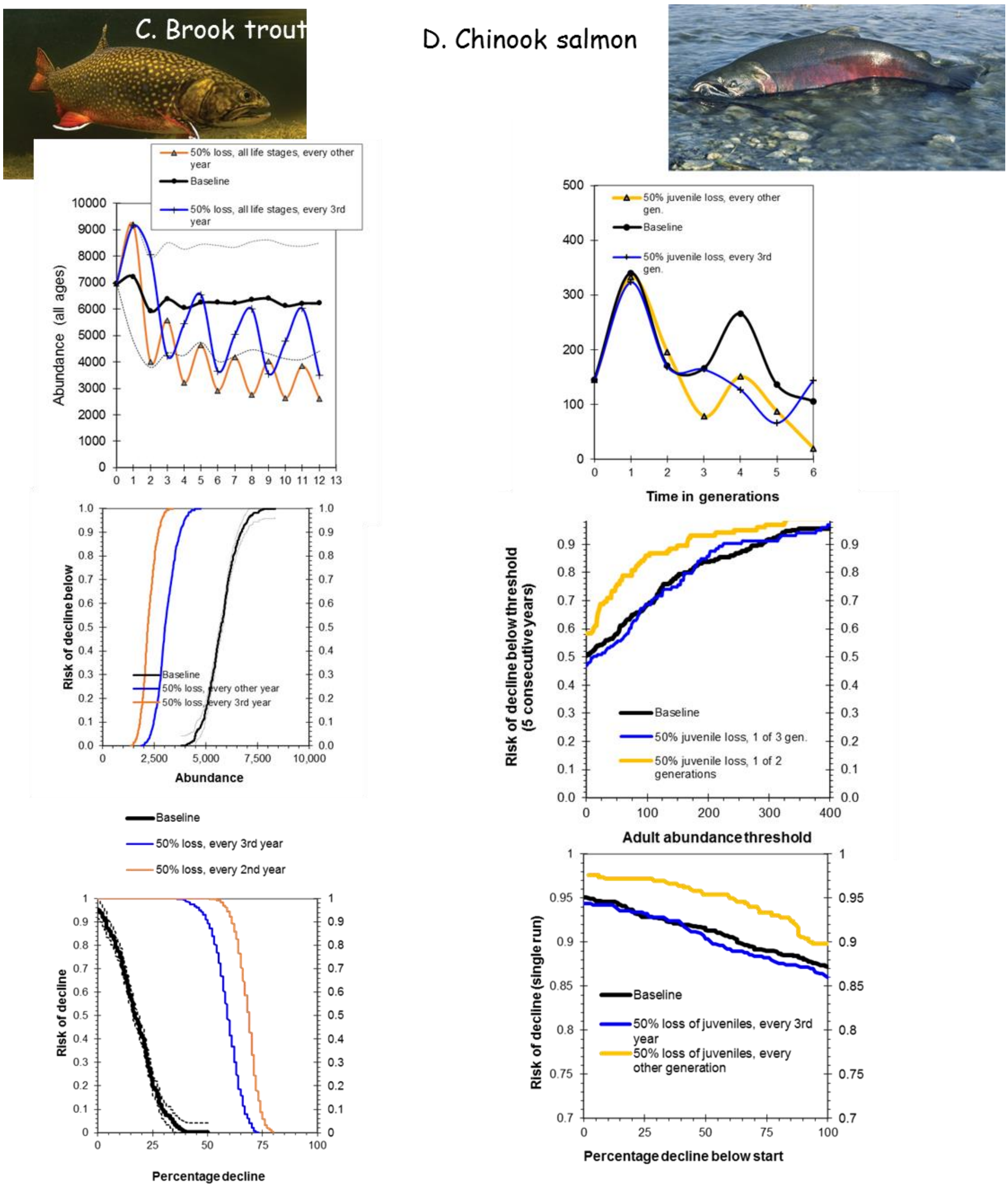


Figure 9. Population modeling projections scenarios (continued): Brook trout and Chinook salmon. With brook trout the periodic mortality scenarios would result in lower overall population and increased risks of a given population decline, but no risk of extinction. The Chinook salmon scenario has high baseline risk of a run failure or complete extirpation during the 30-year (6-generation) projections (87% and 50% respectively), but partial mortality events in their natal streams had little added risk.

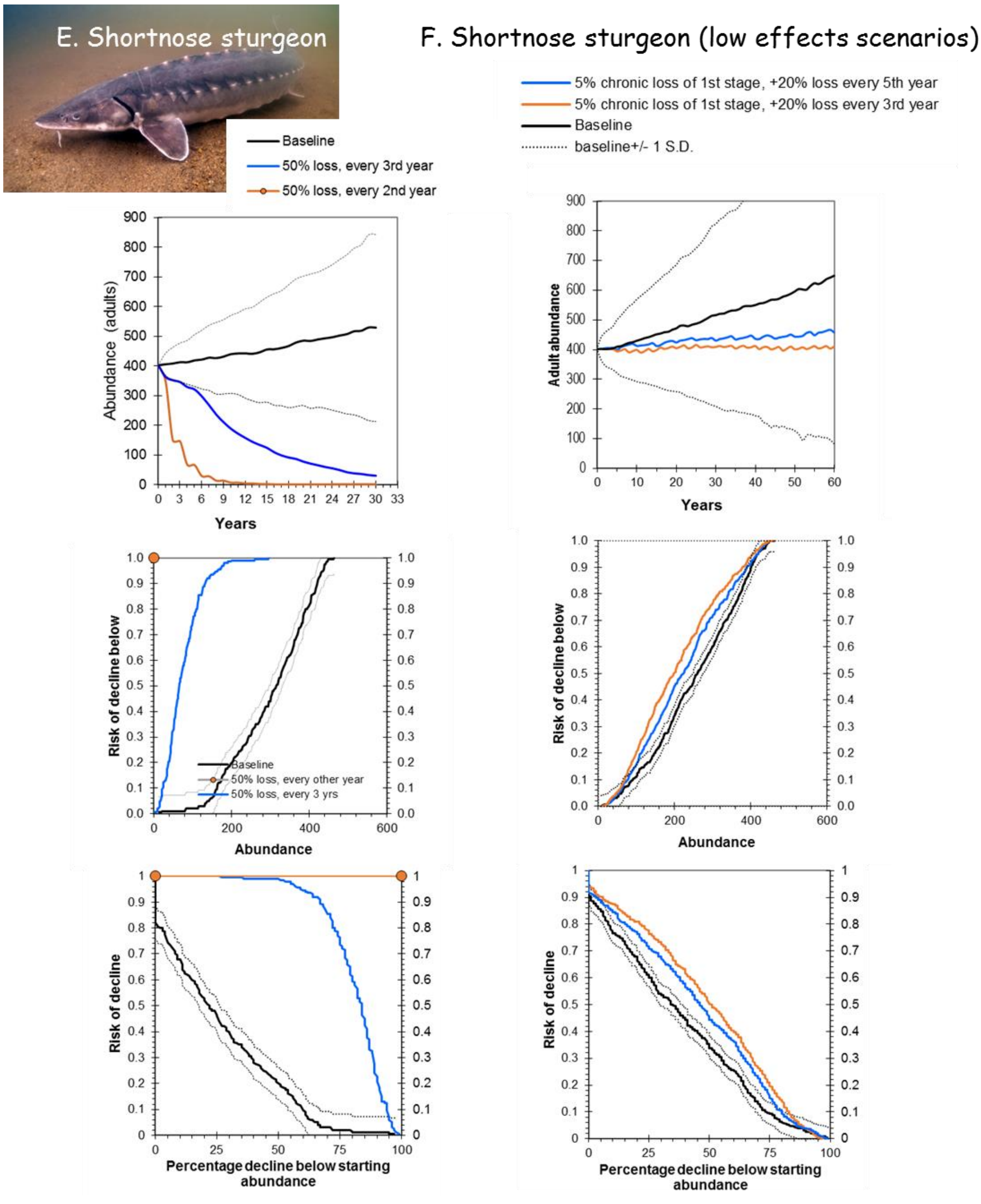


Figure 10. Population modeling projections **E**. Shortnose sturgeon. In these density-independent (no compensation) scenarios, a series of episodic events causing 50% mortality to all life stages (such as hypoxic events) would soon wipe out the entire population. **(F)** A lower-effects scenario presumes that criterion exposure allows up to 10% loss of progeny, such as the design of the Se tissue criterion. The scenario subjects sturgeon to a constant exposure that kills 5% of the progeny each year, in addition to an episodic exposure that hits once every 3-years (orange) or once every 5-years (blue). Results project lower abundances in the "Se criterion" scenarios but risks of extinction or severe decline are only subtly higher.

Table 4. Modeled risks of population declines or local extinctions for species following lethal pollution scenarios

Species	Demographic setting and population characteristics	Scenario	Projected minimum abundance[1]	Risks of severe (90%) population decline[2]
Hyalella azteca	Cold temperate lake, annual life cycle, population near carrying capacity under strong scramble-model density dependence (composite life history)	Baseline, initial abundance 1000 animals/m^2, 11-year projection	192	25% (18-31%)
		Episodic pollution kills 50% of all life stages, every 3rd year	150	37 (31-44%)
		Episodic pollution kills 50% of all life stages, every other year	133	42 (35-47%)
Fathead minnow	Shallow prairie ponds subject to frequent winterkill, Alberta. Strong contest density dependence	Baseline, initial abundance 178 females/m^2, 20-year projection	634	Not meaningful in this scenario because initial abundances were very low following a severe winter die off, ~1% of carrying capacity
		Episodic pollution kills 50% of all life stages, every 3rd year	624	
		Episodic pollution kills 50% of all life stages, every other year	638	
Brook trout	Stunted stream population near carrying capacity (Hunts Creek, MI), under strong contest density dependence	Baseline, initial abundance 7057 fish aged 0-4	5746	<0.2%
		Episodic pollution kills 50% of all life stages, every 3rd year	2897	<0.2%
		Episodic pollution kills 50% of all life stages, every other year	2224	<0.2%
Chinook salmon	Threatened stream population, at low density with highly variable survival rates (Marsh Creek, Idaho)	Baseline, initial abundance 145 spawner	20	88 (85-91)%
		Episodic pollution kills 50% of ELS and juvenile fish (adults safely out to sea), every 3rd gen.	20	87 (83-91)%
		Same, but every 2nd generation affected.	10	91% (87-95)
Shortnose sturgeon	Small, isolated Connecticut River population, with high adult survival and low recruitment rates. Stable and slowly increasing abundance; density independence assumed	Baseline, initial abundance 400 adults	300	1 % (0-7)

Species	Demographic setting and population characteristics	Scenario	Projected minimum abundance[1]	Risks of severe (90%) population decline[2]
		Episodic pollution kills 50% of all life stages, every 3rd year	75	21% (14-27)
		Episodic pollution kills 50% of all life stages, every other year	0	100% (93-100)
	Additional chronic low-effects scenarios:	Baseline Projected for 1 full life span (60 years).	255	2% (0-6)
	Scenario based on Se fish tissue criterion which allows up to 10% mortality to fry of sensitive species	5% of stage 1 (<1yo juveniles) killed annually, with a 20% exceedence event every 3rd year, other ages unaffected. Projected for 1 full life span (60 years).	209	4% (0.4-8)
		5% of stage 1 (<1yo juveniles) killed annually, with a 20% exceedence event every 5th year, other ages unaffected. Projected for 1 full life span (60 years).	224	4% (0.4-8)

Note 1. Projected minimum abundance: the average of the minimum population occurring at any time during each of the 500 population trajectories used in the Monte Carlo simulations **Note 2.** Risks of severe (90%) population decline below initial abundance at any time during simulations

For the shorthead sturgeon, a further scenario of persistent, low-level chronic effects overlain by episodic more severe effects was modeled following the precedent of the Se fish tissue-based aquatic life criteria. The Se criterion is the first example of a tissue-residue based aquatic life criterion and it will presumably inform the approach for any future fish-tissue based criteria. In that example, a 10% effect level (EC10) was used for summarizing effects calculated from tissue residues (USEPA 2016). The assumed effects of chronic selenium exposure are increased rates of deformed fry, which will die before graduating to the next age class. For the modeled scenario, a persistent, background low-level contamination that caused a 5% deformity (death) rate in juvenile sturgeon was imposed, and additionally, once either every 3-years or once every 5-years, higher concentrations causing 20% deformities and death to juveniles were imposed. Survivors suffered no lasting effects, and no other age classes were affected. The projections were run out to 60 years, which is about one full life span. In other words, the exceedences are still fairly low, and the magnitude of effects was lower than in the other scenarios. In this chronic, low-effects scenario (i.e., persistent 5% acute effects to juveniles) which also included a mild episodic event (i.e., 20% acute effect to juveniles), the projected population trajectories were visibly lower than baseline population trajectories, but the population was not projected to crash (Figure 10, Table 4). Risks of decline or extinction were still outside the 95% confidence limits for baseline, and the 1-in-5 year return frequency had lower risks of decline or extinction than did the 1-in-3 year exceedence scenario, but none of the absolute differences in risk were large.

Together, these exercises showed that for the model constructions and scenarios modeled, populations other than sturgeon could withstand a severe effect once in 3 years without going extinct. Longer return intervals would have lower effects, as would if less severe effects were imposed. There is no end of scenarios or species that could be examined, and models could be built for many more species. However, the general applicability to real populations of these or related exercises using generic models should be regarded skeptically.

Frequency versus return period of exceedences

The final evaluations included in this report relate to implications of an exceedence percentage approach that would no longer include the concept of exceedence return intervals. The three main components of ALC expressions (concentration, duration or averaging period, and exceedence frequency or return frequency) are based on either toxicology (concentration and duration) or ecology (the return period). Monitoring expediency was not considered, which can also be important for the facility operator who has to figure out compliance with permits, or state and other regulators who are charged with enforcing compliance. One alternative approach would dispense with defining ALC duration and return frequencies altogether, and instead would use the frequency of criteria exceedences instead. (Benson et al. 2003) suggested that a criteria goal for most waters might be defined as fully protecting 95% of the species 95% of the time, that is, no more than 1 exceedance per 20 monitoring events.

A frequency approach could address a misapplication of the intent of exceedence frequency. For compiling lists of impaired waters, states may determine that, if a waterbody has only been sampled only once during their 2-year assessment and reporting cycle, and if that sample exceeded criteria, then the water body would not be considered to have failed criteria. Failing to meet criteria leads to an "impaired" water listing, and that in turn leads to obligations to develop a TMDL plan to address the impairment and implement measures to restore the waterbody to its designated use. The development and implementation of these plans may be costly, requiring substantial agency and stakeholder investment to develop the plan. ,Plant upgrades or non-point pollution controls needed to implement the plans may take years to implement and be very costly (Keller and Cavallaro 2008). Because of the implications of criteria exceedence, state environmental monitoring officials may have a disincentive to continue monitoring following an exceedence result (NCDEQ 2016).

Using simulated concentrations produced from a lognormal distribution with autocorrelation, Delos (2008) produced theoretical time-varying concentrations, and found a 95% compliance percentage provided a similar exceedance rate as a 30-day average concentration. When applied to an assemblage of population models, the one in 3 year and the 95% compliance rate gave similar, low extinction rates when applied to the ammonia criterion in place at that time. However, it is difficult to judge how synthetic concentration patterns relate to real streams, and here, some real streams with rich monitoring records are examined in this light.

The Blackfoot River, Idaho, is a data-rich stream with elevated Se concentrations that follow a pattern with an annual high concentration pulse during snowmelt runoff, followed by lower concentrations during lower flows (Figure 11). The water-concentration element of the Se criterion is expressed as a 30-day average concentration. For this data-rich stream, it is possible to calculate a rolling 30-day average concentration for each day of the six years shown (at least for the ice free periods when most data were collected). The highest of these rolling 30-day average concentration during a 3-year interval is the statistic that would be compared with the waterborne criterion element. Note that this maximum of the rolling 30-day average concentrations are not the same as maximum monthly concentrations which are consistently lower with this dataset. Consistent with Delos's (2008) prediction, the highest rolling 30-day average concentrations in both of the 3-year return intervals examined were very similar to the simple 95th percentile of the data for the same periods, indicating a similar level of protection. The maximum 4-day rolling average concentrations were consistently higher (Figure 11).

However, having 35-40 samples per year as in the Blackfoot River example is unusual. More common are settings with only a few samples per year, such as the "press disturbance" example of the Coeur d'Alene River, Idaho (Figure 12). At this lower sampling intensity, the 4-day, 30-day, and 95% percentile concentrations are all similar and all metrics would give similar impairment/compliance interpretations.

Let us consider next implications of variability return intervals in episodic, pulse exposure scenarios for a criteria substance which has 4-day duration expression. This example uses aluminum, for which the criteria are a function of 3 parameters that affect toxicity and also vary in time: pH, DOC, and hardness. The examples use the USEPA (2017) draft aluminum criteria. The criteria were finalized after these analyses were prepared. However, since the purpose of the analyses was to examine issues with exceedence frequencies of time variable criteria, the differences in magnitude of criteria exceedences between the draft and final criteria documents are not critical for this purpose.

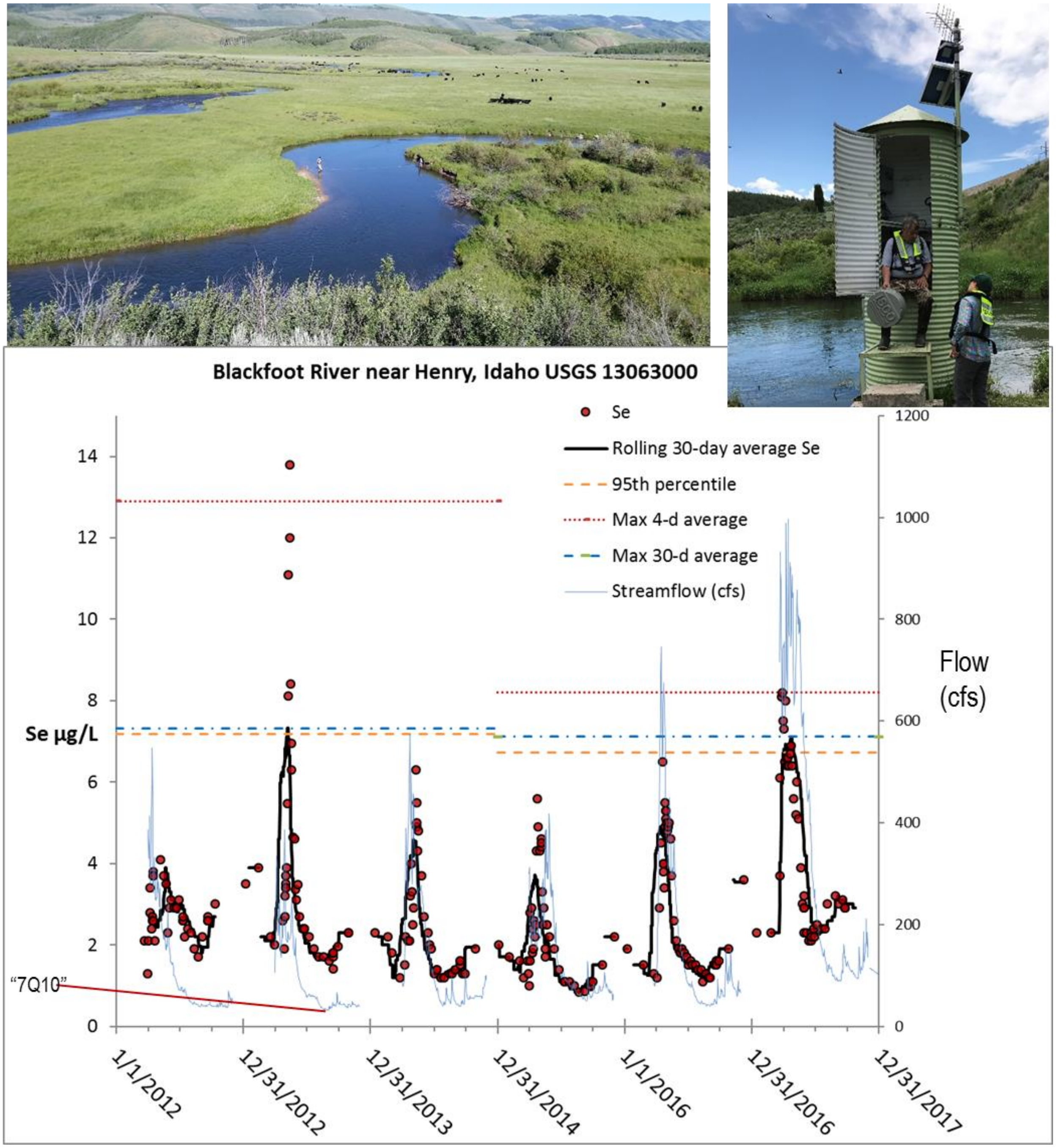


Figure 11. Runoff-pulse disturbance example: Variability in selenium and exceedence statistics for an intensively monitored stream with elevated selenium. Selenium is highest during spring snowmelt runoff, with pronounced, relatively brief annual pulses. Daily rolling 30- and 4-day average concentrations were calculated for each increment with data (i.e., average concentration from Jan 1 to Jan 30, Jan 2 to Jan 31, Jan 3, to Feb 1, 2012, and so on through Oct. 26, 2017. The highest 30-day average concentrations in both 3-year periods were similar to the 95th percentile of the data. The maximum 4-day average concentrations were higher than the 95th percentiles (13 vs 7 µg/L during the 2012-2014 period, and 8 vs 7 µg/L in the 2015-2017 period. The lowest 1-day and lowest 7-day flows are in sync with the low selenium levels 31 vs 32 cubic feet per second, occurring during the week of August 30, 2013 with Se of 1.6-1.7 µg/L.

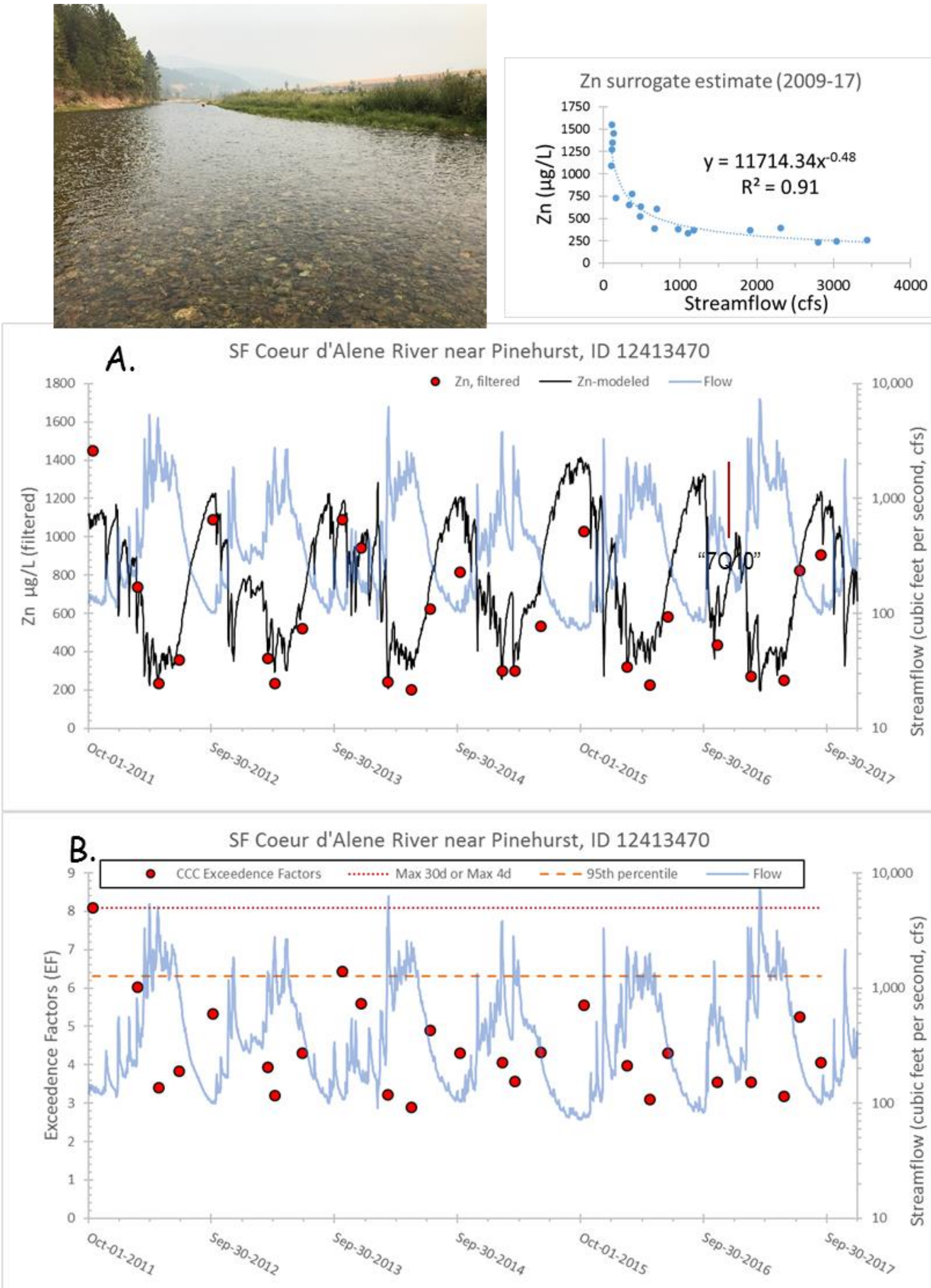


Figure 12. **(A)** Exceedence patterns in a "press disturbance" stream in which the most severe conditions occur during low, stable flows. Concentrations were highest during low flow conditions, owing to contamination from groundwater. Runoff usually dilutes the river. Zinc concentrations can be estimated simply from streamflows (inset). This surrogate model suggests unmonitored peak concentrations would have occurred in summers of 2014, 2015, and 2016.
(B) Criteria exceedence frequencies are the same for 4-day and 30-day averages because each averaging period has n=1. Frequency of "allowable exceedences" at the 4-d and the 95%-tile are similar with this level of monitoring (Because the 4X/year monitoring strategy targets hydrologic periods (baseflow, storm event, rising limb, peak runoff), sample intervals ranged from 36 to 162 days, with no 4-day or 30-day period with >1 sample, other than same day quality control samples.) The 7Q10 corresponds with the predicted highest EF.

The next case study uses data from a pair of intensively monitored streams affected by acid rain and elevated aluminum (Al) in the Adirondack Mountains of New York. Aluminum concentrations are generally low during winter and late summer, with periodic abrupt spikes in spring and summer, which are presumably related to snowmelt and rain storms (Figure 13). To more easily visualize criteria exceedence frequencies and return intervals, the Al concentrations are expressed as criteria exceedence factors (EFs), which are the ratios of Al concentrations divided by the matching Al criterion concentrations (Figure 14). Very few samples were collected within 4-days of each other, so each sample represents a 4-day average. In this case, a single high individual sample can make the 4-day "average" concentration much higher than the 2nd highest concentration or the 95th percentile. 13% and 16% of the EFs were greater than 1 for Vanderwhacker Brook and Durgin Brook, respectively (Figure 14).

To consider the conceptual effects of a 95% exceedence percentage, first the Al samples from the Adirondack datasets were scaled down proportionally so that there were exactly 5% exceedences in both streams (Figure 15). Then, to conceptualize the potentially affected fraction (PAF) of genera, the chronic SSD for aluminum was fit to a logarithmic EF regression (calculated assuming consistent pH [7], hardness [100 mg/L], and DOC [1 mg/L]). Then, the potential effect of EFs of different magnitude on an assemblage of taxa can visualized by how far up the SSD they run. The maximum EFs of about 1.8 and 2.1 for the two streams would predict up to 20% and 30% of the genera being potentially affected during the 3-year period. Greater than 10% of the genera in the SSD would be affected in 4 and 3 episodes for each stream, respectively (Figure 16).

An alternative to using the PAF of an SSD to visualize potential effects is to relate EFs to toxicity of a locally important species. Brook trout are the iconic fish of the Adirondacks and a study with sufficient details to calculate toxicity as a function of criteria exceedence factors was located (Cleveland et al. 1989). In this fashion, scenario A ("Vanderwacker Brook") would be expected to experience a single 10% mortality event and scenario B ("Durgin Brook") would be expected to experience a single 30% mortality event during the period (Figure 17).

These examples are intended illustrate approaches that could be used to visualize potential effects associated with exceedences of differing magnitude under a 95% exceedence percentage approach or with the return interval approach. These examples have limitations, most obviously that the episodic exposure pulses are probably longer than 4 days but shorter than the 20-day brook trout toxicity test, or the exposures of the chronic SSD. The draft 2017 aluminum criteria have an acute criterion to chronic criterion ratio of about 1.5. A similar visualization could be done using a PAF of the acute criterion, although even though the Vanderwhacker and Durgin Brook examples have exceptionally rich datasets, they are still insufficient to capture variability at the time scale relevant to acute criterion.

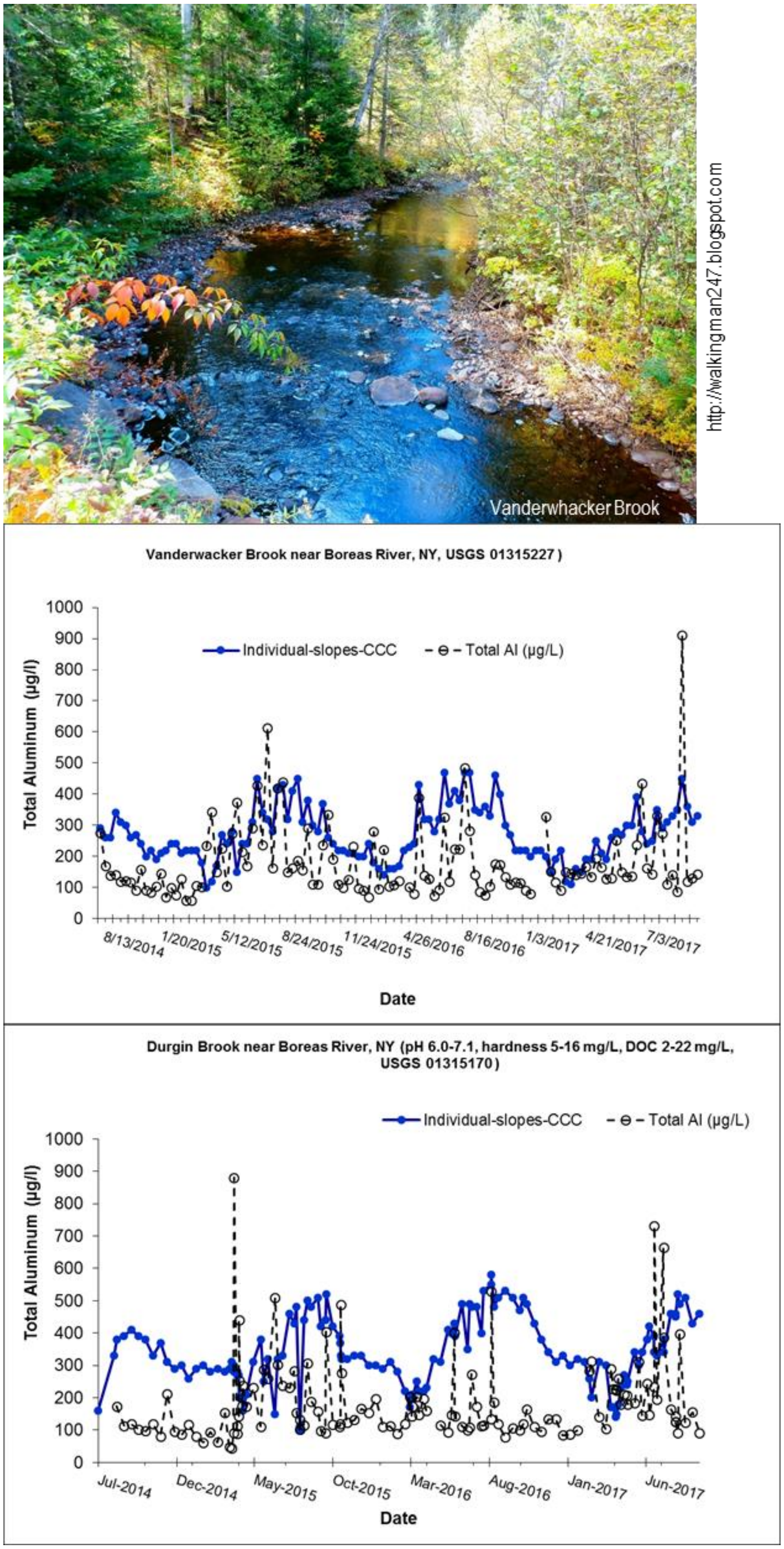
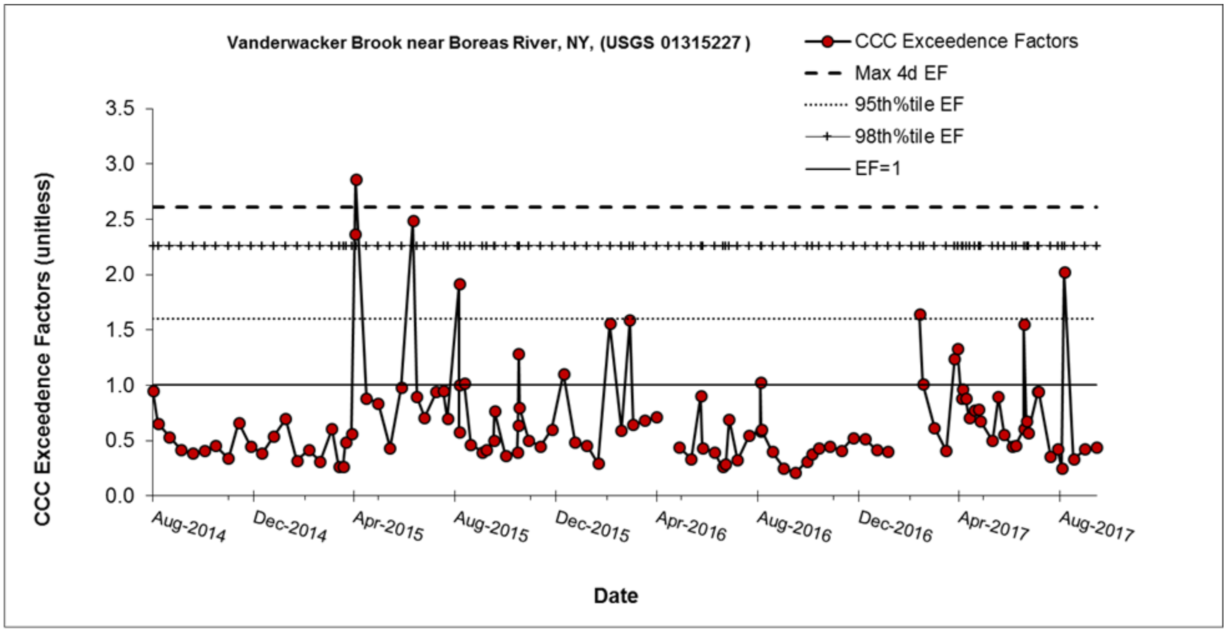


Figure 13. Variability during a 3-year period of record for total aluminum concentrations and calculated chronic criteria concentrations (CCC) in two intensely monitored streams affected by acid rain in the Adirondack Mountains, New York. The "CCC" are non-regulatory draft criteria which are a function of dissolved organic carbon, pH, and hardness.



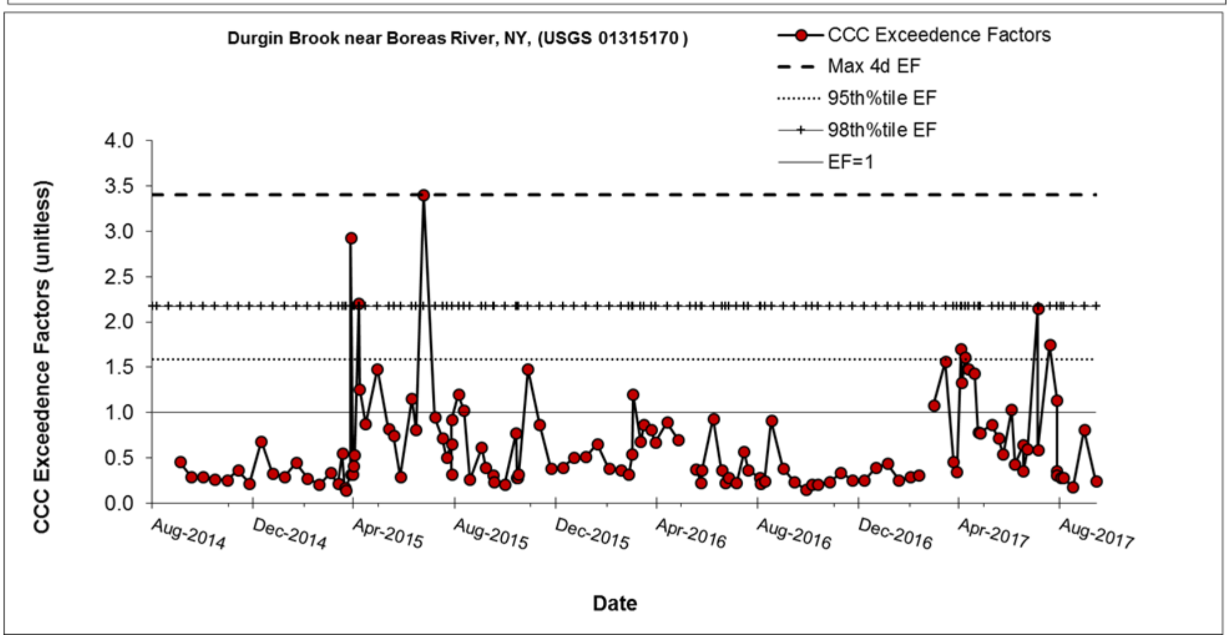
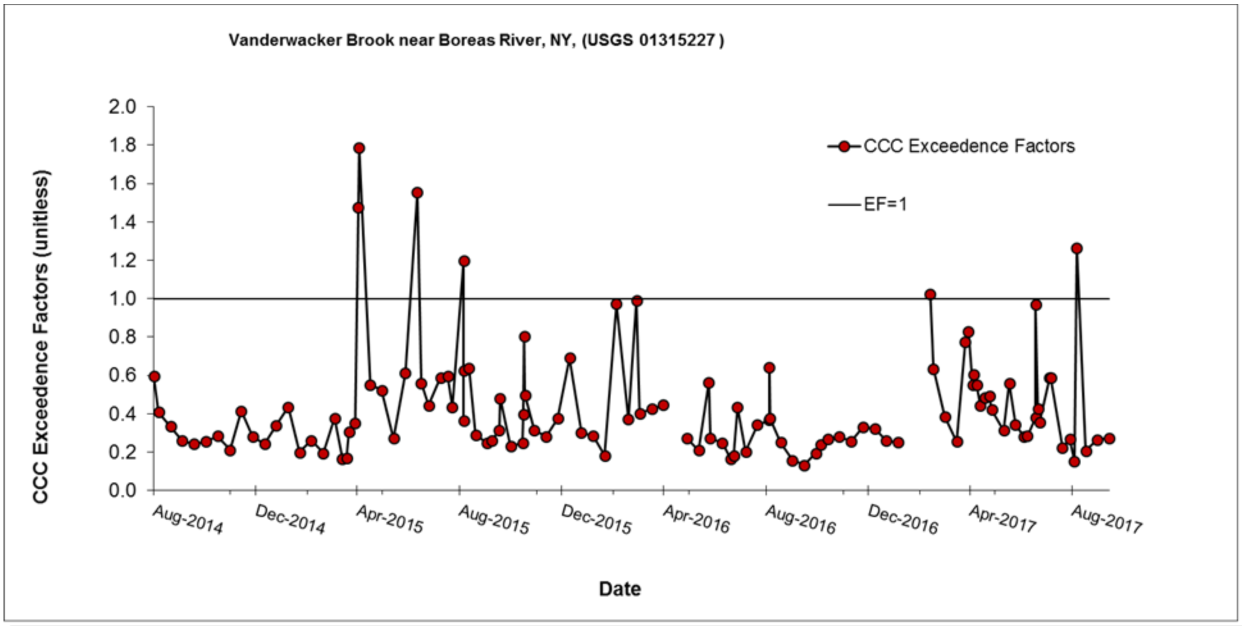


Figure 14. Criteria exceedence factors of total aluminum concentrations relative to some different exceedence factor benchmarks (EF = Ambient concentration/chronic criteria value). The maximum 4-day exceedence factors for both streams were considerably higher than the 95th percentile exceedence factors (2.6 vs 1.6 for Vanderwacker Brook and 3.4 vs. 1.6 for Durgin Brook.



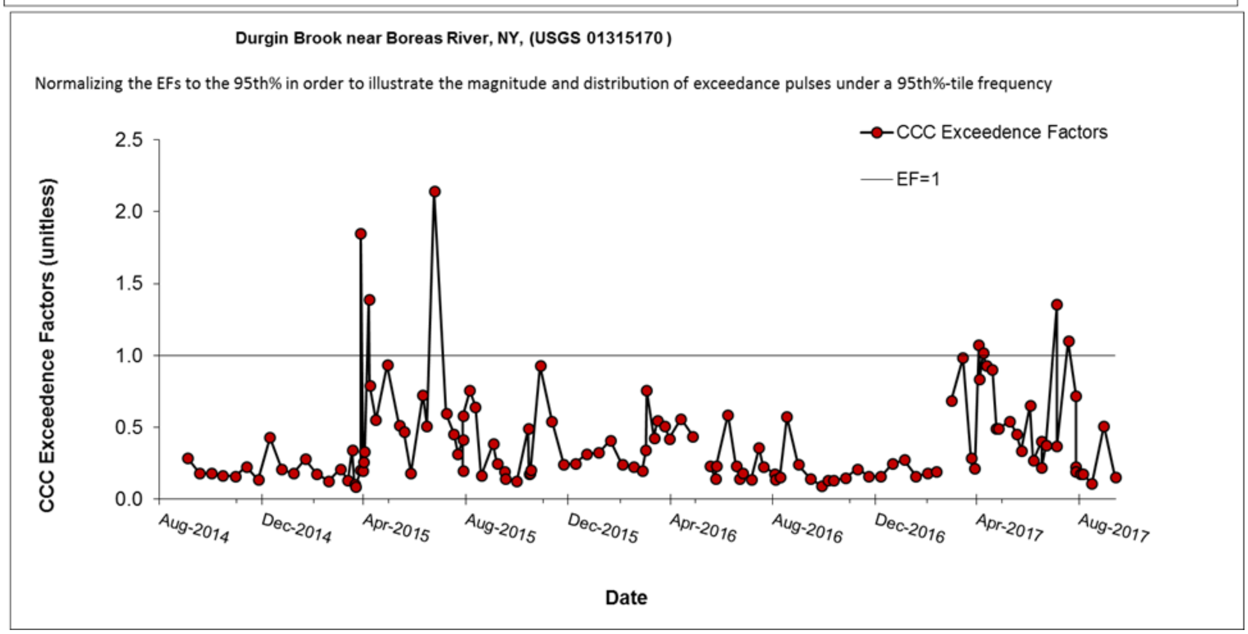
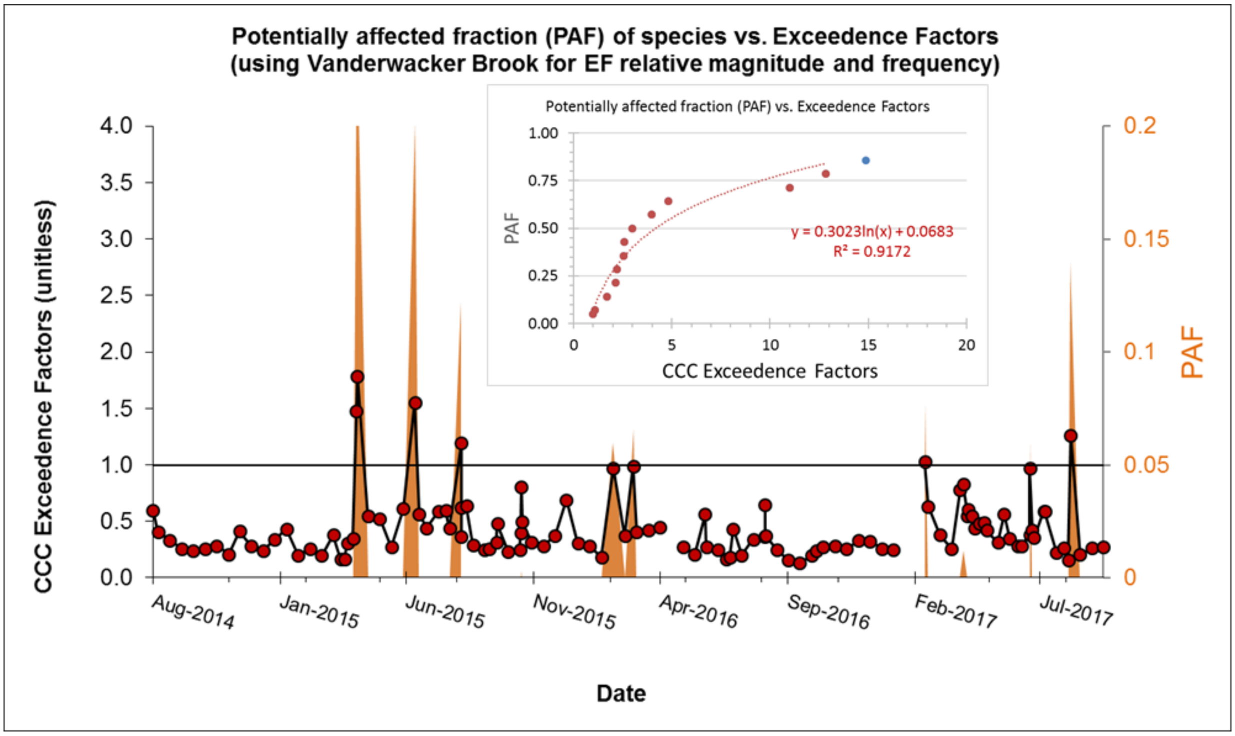


Figure 15. In a situation with an intensely sampled fixed-monitoring benchmark site, a 95th percentile exceedence (EF) factor would allow more exceedences than would the 1-in-3-year exceedence. Here the previous aluminum exceedence plots are rescaled to illustrate that point. By dividing the EF's by the 95th percentile EFs, only those "violations" of a hypothetical 95th percentile EFs exceedence are above the 1.0 line. If the clusters of repeated exceedences were spaced such that organisms that were still in recovery phase were hit with a second exceedence, the organisms could suffer greater effects than from a single exceedence. In practice, very few water bodies would have this intensity of monitoring to detect pulses such as these. This suggests that an exceedence rate should also include an allowable magnitude term, such that in addition to an allowable frequency of exceedences, no single exceedence be greater than, for example, 2X the CCC.



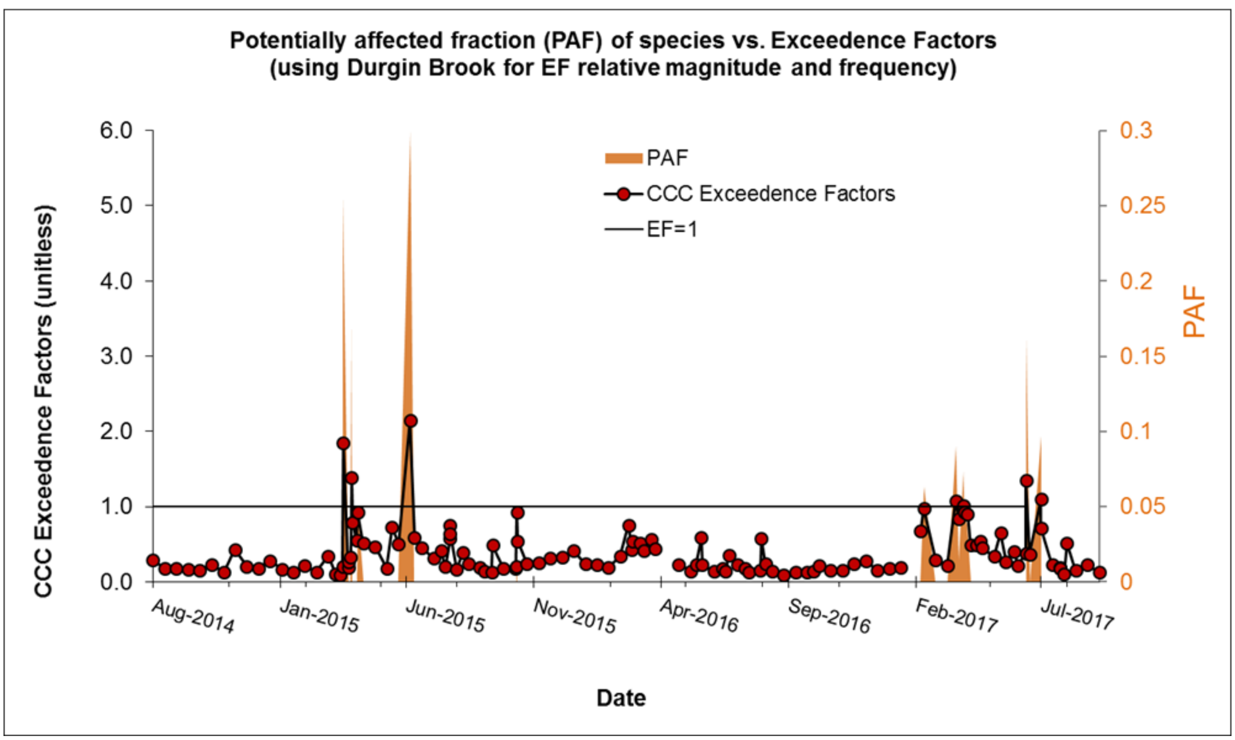


Figure 16. Potential biological implications of an allowable exceedence scheme can be visualized by expressing the chronic species sensitivity distribution (SSD) for aluminum in terms of exceedence factors (EF) versus the potentially affected fraction of species (PAF), and then overlaying the EFs and PAFs. The maximum EFs of about 1.8 and 2.1 for the two streams would predict up to 20% and 30% of the species being potentially affected during the 3-year period and 4 and 3 episodes, respectively, when >10% of the species in the SSD would be affected. PAF estimated from the 2017 draft aluminum criteria document (inset).

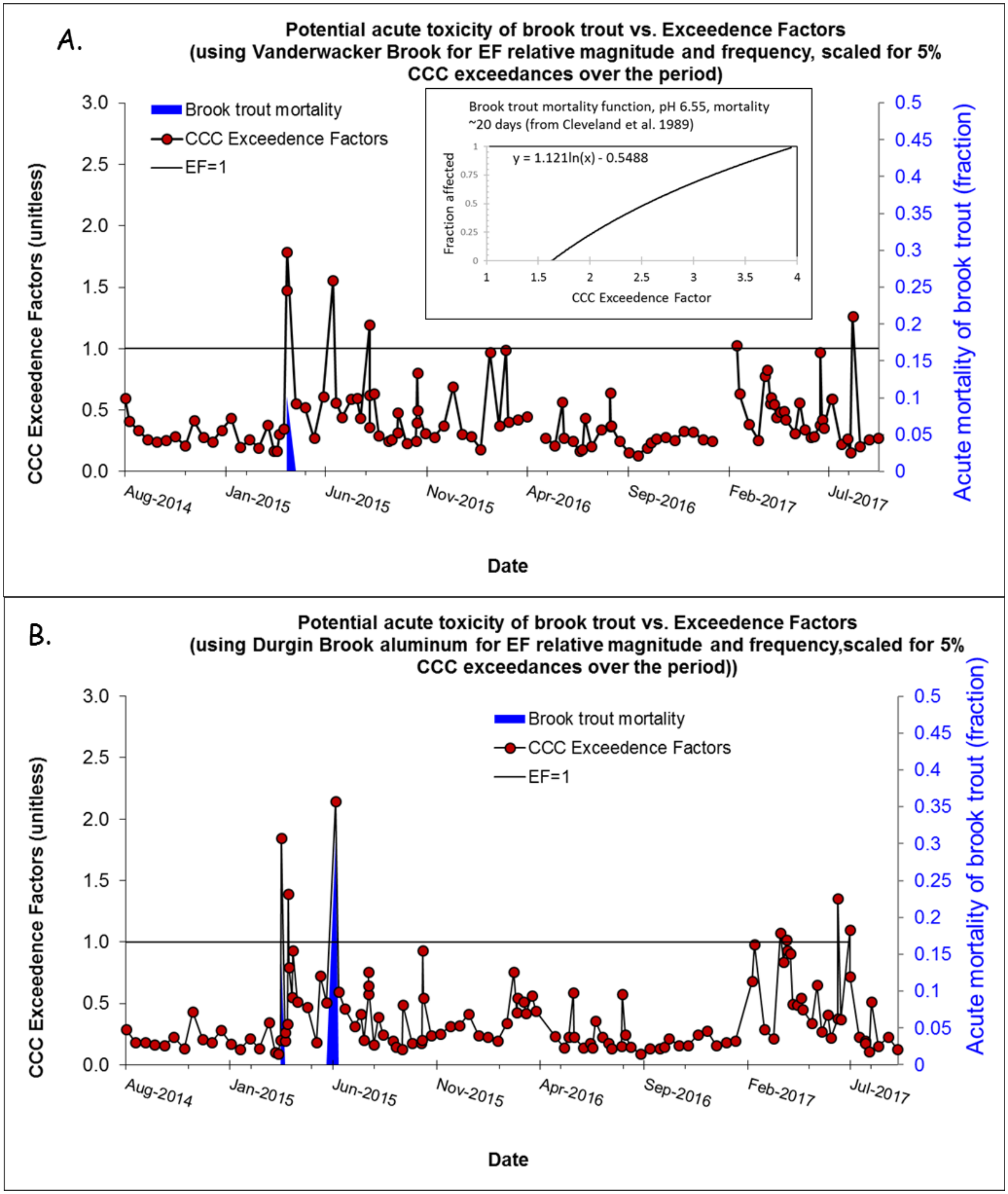


Figure 17. The potential biological implications of a short-term allowable exceedence scheme can also be visualized by overlaying exceedence factors (EF) versus acute toxicity of a locally important species or a function. Here brook trout mortalities are expressed as a function of exceedence factors. Using a piecewise linear regression, the onset of acute mortalities (>LC0) begins at 1.6 CCC EFs and reaches 100% by EF 4.0. In this fashion, scenario A ("Vanderwacker Brook") would be expected to experience a single 10% mortality event and scenario B ("Durgin Brook") would be expected to experience a single 30% mortality event during the period.

A summary and list of overall conclusions drawn from this review were listed at the front of this document.

References

Adams, S.B. and M.L. Warren, Jr. 2005. Recolonization by warmwater fishes and crayfishes after severe drought in upper Coastal Plain hill streams. *Transactions of the American Fisheries Society*. 134(5): 1173-1192. https://doi.org/10.1577/T04-089.1

Allen, M. and T. Gast. 2005. Recovery of fish populations in the upper Sacramento River following the 1991 Cantara spill. Report to the California Department of Fish and Game. Thomas R. Payne and Associates, Arcata, CA. 260 pp.

Arce, E., V. Archaimbault, C.P. Mondy, and P. Usseglio-Polatera. 2014. Recovery dynamics in invertebrate communities following water-quality improvement: taxonomy- vs trait-based assessment. *Freshwater Science*. 33(4): 1060-1073. https://doi.org/10.1086/678673

Arnekleiv, J.V. and L. Størset. 1995. Downstream effects of mine drainage on benthos and fish in a Norwegian river: a comparison of the situation before and after river rehabilitation. *Journal of Geochemical Exploration*. 52(1): 35-43. https://doi.org/10.1016/0375-6742(94)00029-B

Baldigo, B.P., K.M. Roy, and C.T. Driscoll. 2016. Response of fish assemblages to declining acidic deposition in Adirondack Mountain lakes, 1984–2012. *Atmospheric Environment*. 146: 223-235. https://doi.org/10.1016/j.atmosenv.2016.06.049

Baldwin, D.H., J.A. Spromberg, T.K. Collier, and N.L. Scholz. 2009. A fish of many scales: extrapolating sublethal pesticide exposures to the productivity of wild salmon populations. *Ecological Applications*. 19(8): 2004-2015. https://doi.org/10.1890/08-1891.1

Barnthouse, L.W. 2004. Quantifying population recovery rates for ecological risk assessment. *Environmental Toxicology and Chemistry*. 23(2): 500-508. https://doi.org/10.1897/02-521

Beisner, B.E., D.T. Haydon, and K. Cuddington. 2003. Alternative stable states in ecology. *Frontiers in Ecology and the Environment*. 1(7): 376-382. https://doi.org/10.1890/1540-9295(2003)001[0376:assie]2.0.co;2

Beketov, M.A., R.B. Schäfer, A. Marwitz, A. Paschke, and M. Liess. 2008. Long-term stream invertebrate community alterations induced by the insecticide thiacloprid: Effect concentrations and recovery dynamics. *Science of the Total Environment*. 405(1–3): 96-108. https://doi.org/10.1016/j.scitotenv.2008.07.001

Bender, E.A., T.J. Case, and M.E. Gilpin. 1984. Perturbation Experiments in Community Ecology: Theory and Practice. *Ecology*. 65(1): 1-13. https://doi.org/10.2307/1939452

Benoit, D.A., E.N. Leonard, G.M. Christensen, and J.T. Fiandt. 1976. Long-term effects of cadmium exposure on three generations of brook trout (*Salvelinus fontinalis*). *Transactions of the American Fisheries Society*. 108(4): 550-560. https://doi.org/10.1577/1548-8659(1976)105<550:TEOCOT>2.0.CO;2

Benson, W.H., H.E. Allen, J.P. Connolly, C.G. Delos, L.W. Hall Jr, S.N. Luoma, D. Maschwitz, J.S. Meyer, J.W. Nichols, and W.A. Stubblefield. 2003. Exposure Analysis. Pages 15-51 *in* M. C. Reiley, W. A. Stubblefield, W. J. Adams, D. M. Di Toro, P. V. Hodson, R. J. Erickson, and F. J. Keating, editors. *Reevaluation of the State of the Science for Water-Quality Criteria Development*. SETAC Press, Pensacola, FL.

Besley, C.H. and B.C. Chessman. 2008. Rapid biological assessment charts the recovery of stream macroinvertebrate assemblages after sewage discharges cease. *Ecological Indicators*. 8(5): 625-638. https://doi.org/10.1016/j.ecolind.2007.09.003

Beverton, R.J.H. and S.J. Holt. 1959. A Review of the Lifespans and Mortality Rates of Fish in Nature, and Their Relation to Growth and Other Physiological Characteristics. Pages 142-180 *in Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing)*. John Wiley & Sons, Ltd, http://dx.doi.org/10.1002/9780470715253.ch10.

Binns, N.A. 1967. Effects of rotenone treatment on the fauna of the Green River, Wyoming. Fisheries Research Bulletin Number 1, Wyoming Game and Fish Commission, Cheyenne, Wyoming. 115 pp.

Birstein, V.J. 1993. Sturgeons and Paddlefishes: Threatened Fishes in Need of Conservation. *Conservation Biology*. 7(4): 773-787. https://doi.org/10.1046/j.1523-1739.1993.740773.x

Blanchfield, P.J., K.A. Kidd, M.F. Docker, V.P. Palace, B.J. Park, and L.D. Postma. 2015. Recovery of a wild fish population from whole-lake additions of a synthetic estrogen. *Environmental Science & Technology*. 49(5): 3136-3144. https://doi.org/10.1021/es5060513

Buys, D.J., A.R. Stojak, W. Stiteler, and T.F. Baker. 2015. Ecological risk assessment for residual coal fly ash at Watts Bar Reservoir, Tennessee: Limited alteration of riverine-reservoir benthic invertebrate community following dredging of ash-contaminated sediment. *Integrated Environmental Assessment and Management*. 11(1): 43-55. https://doi.org/10.1002/ieam.1577

Cairns, J., Jr., J.S. Crossman, K.L. Dickson, and E.E. Herricks. 1971. The recovery of damaged streams. *ASB Bulletin*. 18(3): 79-106

Capel, P.D., W. Giger, P. Reichert, and O. Wanner. 1988. Accidental input of pesticides into the Rhine River. *Environmental Science & Technology*. 22(9): 992-997. https://doi.org/10.1021/es00174a001

CDPHE. 2017. Section 303(d) Listing Methodology 2018 Listing Cycle. Colorado Department of Public Health and Environment, Water Quality Control Division, Denver. 78 pp. https://www.colorado.gov/pacific/sites/default/files/303d_LM_2018.pdf.

Chadwick, J.W., S.P. Canton, and R.L. Dent. 1986. Recovery of benthic invertebrate communities in Silver Bow Creek, Montana, following improved metal mine wastewater treatment. *Water, Air, and Soil Pollution*. 28(3): 427-438. https://doi.org/10.1007/bf00583505

Clearwater, S.J., P.G. Jellyman, B.J.F. Biggs, C.W. Hickey, N. Blair, and J.S. Clayton. 2011. Pulse-dose application of chelated copper to a river for *Didymosphenia geminata* control: Effects on macroinvertebrates and fish. *Environmental Toxicology and Chemistry*. 30(1): 181-195. https://doi.org/10.1002/etc.369

Clements, W.H. 1999. Metal tolerance and predator–prey interactions in benthic macroinvertebrate stream communities. *Ecological Applications*. 9(3): 1073-1084. https://doi.org/10.1890/1051-0761(1999)009[1073:mtappi]2.0.co;2

Clements, W.H., J.H. Van Hassel, D.S. Cherry, and J. Cairns, Jr. 1989. Colonization, variability, and the use of substratum-filled trays for biomonitoring benthic communities. *Hydrobiologia*. 173(1): 45-53. https://doi.org/10.1007/bf00008597

Clements, W.H., N.K.M. Vieira, and S.E. Church. 2010. Quantifying restoration success and recovery in a metal-polluted stream: a 17-year assessment of physicochemical and biological responses. *Journal of Applied Ecology*. 47(4): 899-910. https://doi.org/10.1111/j.1365-2664.2010.01838.x

Cleveland, L., E.E. Little, and R.H. Wiedmeyer. 1989. Chronic no-observed-effect concentrations of aluminum for brook trout exposed in low calcium dilute acidic water. Pages 229-246 *in* T. E. Lewis, editor. *Environmental Chemistry and Toxicology of Aluminum*. Lewis Publishers, Inc.,, Chelsea, MI.

Cook, K.V., R.J. Lennox, S.G. Hinch, and S.J. Cooke. 2015. Fish out of water: how much air is too much? *Fisheries*. 40(9): 452-461. https://doi.org/10.1080/03632415.2015.1074570

Coombs, J., J.L. Wilson, J. Burr, S.J. Fraley, and T.R. Russ. 2010. Pigeon River Recovery Project: Bringing Back Aquatic Diversity. Presentation (53 slides). https://www.se-eppc.org/2010/pdf/Coombs.pdf [Accessed November 2018].

Copeland, C. 1993. Toxic Pollutants and the Clean Water Act: Current Issues. Congressional Research Service Report No. 93-849. https://digital.library.unt.edu/ark:/67531/metacrs89/m1/1/high_res_d/93-849_1993Sep21.html.

Crawford, C.G., D.J. Wangsness, and J.D. Martin. 1992. Recovery of benthic-invertebrate communities in the White River near Indianapolis, Indiana, USA, following implementation of advanced treatment of municipal wastewater. *Archiv für Hydrobiologie - Advances in Limnology*. 126(1): 67-84

Crutchfield, J.U., Jr. 2000. Recovery of a power plant cooling reservoir ecosystem from selenium bioaccumulation. *Environmental Science and Policy*. 3(S1): S145-S163 https://doi.org/10.1016/S1462-9011(00)00042-3

Danylchuk, A.J. and W.M. Tonn. 2003. Natural disturbances and fish: local and regional influences on winterkill of fathead minnows in boreal lakes. *Transactions of the American Fisheries Society*. 132(2): 289-298. https://doi.org/10.1577/1548-8659(2003)132<0289:NDAFLA>2.0.CO;2

Dauwalter, D.C., F.J. Rahel, and K.G. Gerow. 2009. Temporal variation in trout populations: implications for monitoring and trend detection. *Transactions of the American Fisheries Society*. 138(1): 38-51. https://doi.org/10.1577/T07-154.1

Davies, P.E. and L.S.J. Cooke. 1993. Catastrophic macroinvertebrate drift and sublethal effects on brown trout, *Salmo trutta*, caused by cypermethrin spraying on a Tasmanian stream. *Aquatic Toxicology*. 27(3-4): 201-224. https://doi.org/10.1016/0166-445X(93)90055-6

DeForest, D.K., S. Pargee, C. Claytor, S.P. Canton, and K.V. Brix. 2016. Biokinetic food chain modeling of waterborne selenium pulses into aquatic food chains: Implications for water quality criteria. *Integrated Environmental Assessment and Management*. 12(2): 230-246. https://doi.org/10.1002/ieam.1696

Delos, C.G. 2008. Modeling framework applied to establishing an allowable frequency for exceeding aquatic life criteria (workshop materials). Technical Workshop on Population-Level Ecological Risk Assessment [June 2008]. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, D.C.

Detenbeck, N.E., P.W. DeVore, G.J. Niemi, and A. Lima. 1992. Recovery of temperate-stream fish communities from disturbance - a review of case studies and synthesis of theory. *Environmental Management*. 16(1): 33-53. https://doi.org/10.1007/BF02393907

Dill, W.A. and A.J. Cordone. 1997. History And Status of Introduced Fishes In California, 1871 – 1996. Fish Bulletin 178, State of California, Department of Fish and Game [Special Collections & Archives, UC San Diego, La Jolla, 92093-0175 (https://lib.ucsd.edu/sca)].

https://oac.cdlib.org/view?docId=kt8p30069f&query=&brand=oac4 [Accessed 8 February 2019].

Divino, J.N. and W.M. Tonn. 2007. Effects of reproductive timing and hatch date on fathead minnow recruitment. *Ecology of Freshwater Fish*. 16(2): 165-176. https://doi.org/10.1111/j.1600-0633.2006.00208.x

Duggan, S.B., C.J. Kotalik, and W.H. Clements. 2018. Integrating results of field biomonitoring and mesocosm experiments to validate postspill impacts of petroleum hydrocarbons on stream benthic communities. *Environmental Science & Technology*. 52(22): 13584-13590. https://doi.org/10.1021/acs.est.8b04807

Dunham, J.B., S.B. Adams, R.E. Schroeter, and D.C. Novinger. 2002. Alien invasions in aquatic ecosystems: toward an understanding of brook trout invasions and potential impacts on inland cutthroat trout in western North America. *Reviews in Fish Biology and Fisheries*. 12: 373-391. https://doi.org/10.1023/A:1025338203702

Dunham, J.B., M.K. Young, R.E. Gresswell, and B.E. Rieman. 2003. Effects of fire on fish populations: landscape perspectives on persistence of native fishes and nonnative fish invasions. *Forest Ecology and Management*. 178(1-2): 183-196. https://doi.org/10.1016/S0378-1127(03)00061-6

Eagle, J. and B.H. Thompson, Jr. . 2003. Answering Lord Perry's question: dissecting regulatory overfishing. *Ocean and Coastal Management*. 46(6-7): 649-680. https://doi.org/doi:10.1016/S0964-5691(03)00040-1

Effler, S.W., S. Litten, S.D. Field, T. Tong-Ngork, F. Hale, M. Meyer, and M. Quirk. 1980. Whole lake responses to low level copper sulfate treatment. *Water Research*. 14(10): 1489-1499. https://doi.org/10.1016/0043-1354(80)90015-9

Ensign, W.E., K.N. Leftwich, P.L. Angermeier, and C.A. Dolloff. 1997. Factors influencing stream fish recovery following a large-scale disturbance. *Transactions of the American Fisheries Society*. 127(6): 895–907. https://doi.org/10.1577/1548-8659(1997)126<0895:FISFRF>2.3.CO;2

Ferson, S. and L.R. Ginzburg. 1996. Inferring ecological risk from toxicity bioassays. *Water, Air, and Soil Pollution*. 90(1): 71-82. https://doi.org/10.1007/BF00619269

Finley, K. and R. Garrett. 2007. Learned discourses. *Integrated Environmental Assessment and Management*. 3(2): 297-299. https://doi.org/10.1002/ieam.5630030216

Fisher, S.G., L.J. Gray, N.B. Grimm, and D.E. Busch. 1982. Temporal Succession in a Desert Stream Ecosystem Following Flash Flooding. *Ecological Monographs*. 52: 93–110. https://doi.org/10.2307/2937346

Fredeen, F.J.H. 1975. Effects of a single injection of methoxychlor black-fly larvicide on insect larvae in a 161-km (100-mile) section of the North Saskatchewan River. *The Canadian Entomologist*. 107(8): 807-817. https://doi.org/10.4039/Ent107807-8

Frost, T.M., J.M. Fischer, J.L. Klug, S.E. Arnott, and P.K. Montz. 2006. Trajectories of zooplankton recovery in the Little Rock Lake whole-lake acidification experiment. *Ecological Applications*. 16(1): 353-367. https://doi.org/doi:10.1890/04-1800

Gergs, A., S. Classen, T. Strauss, R. Ottermanns, T.C.M. Brock, H.T. Ratte, U. Hommen, and T.G. Preuss. 2016. Ecological recovery potential of freshwater organisms: consequences for environmental risk assessment of chemicals. *Reviews of Environmental Contamination and Toxicology*. 236: 259-294. https://doi.org/10.1007/978-3-319-20013-2_5

Giddings, J.M., K.R. Solomon, and S.J. Maund. 2001. Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies. *Environmental Toxicology and Chemistry*. 20(3): 660-668. https://doi.org/10.1002/etc.5620200327

Giger, W. 2009. The Rhine red, the fish dead—the 1986 Schweizerhalle disaster, a retrospect and long-term impact assessment. *Environmental Science and Pollution Research*. 16(1): 98-111. https://doi.org/10.1007/s11356-009-0156-y

Ginzburg, L.R., S. Ferson, and H.R. Akçakaya. 1990. Reconstructibility of density dependence and the conservative assessment of extinction risks. *Conservation Biology*. 4(1): 63-70. https://doi.org/10.1111/j.1523-1739.1990.tb00268.x

Gore, J. and A.M. Milner. 1990. Island biogeographical theory: Can it be used to predict lotic recovery rates? *Environmental Management*. 14(5): 737-753. https://doi.org/10.1007/bf02394722

Greathouse, E.A., J.G. March, and C.M. Pringle. 2005. Recovery of a tropical stream after a harvest-related chlorine poisoning event. *Freshwater Biology*. 50(4): 603-615. https://doi.org/10.1111/j.1365-2427.2005.01344.x

Griffith, J.S. and D.A. Andrews. 1981. Effects of a small suction dredge on fishes and aquatic invertebrates in Idaho streams. *North American Journal of Fisheries Management*. 1(1): 21-28. https://doi.org/10.1577/1548-8659(1981)1<21:EOASSD>2.0.CO;2

Ham, K.D. and T.N. Pearsons. 2000. Can reduced salmonid population abundance be detected in time to limit management impacts? *Canadian Journal of Fisheries and Aquatic Sciences*. 57(1): 17-24. https://doi.org/10.1139/f99-175

Hamilton, S.K. 2012. Biogeochemical time lags may delay responses of streams to ecological restoration. *Freshwater Biology*. 57: 43-57. https://doi.org/10.1111/j.1365-2427.2011.02685.x

Hanson, M.L., D.W. Graham, E. Babin, D. Azam, M.A. Coutellec, C.W. Knapp, L. Lagadic, and T. Caquet. 2007. Influence of isolation on the recovery of pond mesocosms from the application of an insecticide. I. Study design and planktonic community responses. *Environmental Toxicology and Chemistry*. 26(6): 1265-1279. https://doi.org/doi:10.1897/06-248R.1

Hardy, R.W., L. Oram, and G. Möller. 2010. Effects of dietary selenomethionine on cutthroat trout (*Oncorhynchus clarki bouvieri*) growth and reproductive performance over a life cycle. *Archives of Environmental Contamination and Toxicology*. 58(1): 237-245. https://doi.org/10.1007/s00244-009-9392-x

Hastings, E., W.H. Kittams, and J.H. Pepper. 1961. Repopulation by aquatic insects in streams sprayed with DDT. *Annals of the Entomological Society of America*. 54(3): 436–437. https://doi.org/10.1093/aesa/54.3.436

Hawkins, C.P. and J.R. Sedell. 1990. The role of refugia in the recolonization of streams devastated by the 1980 eruption of Mount St. Helens. *Northwest Science*. 64(5): 271-274

Herbst, D.B., R.B. Medhurst, and N.P.R. Black. 2018. Long-term effects and recovery of streams from acid mine drainage and evaluation of toxic metals threshold ranges for community re-assembly. *Environmental Toxicology and Chemistry*. 37(10): 2575-2592. https://doi.org/10.1002/etc.4217

Hermanutz, R.O., K.N. Allen, N. Detenbeck, and C.E. Stephan. 1996. Exposure of Bluegills (*Lepomis macrochirus*) to Selenium in Outdoor Experimental Streams. U.S. Environmental Protection

Agency, National Health and Environmental Research Laboratory, Mid-Continent Ecological Division, Duluth. 43 pp.

Hesthagen, T., A. Fjellheim, A.K. Schartau, R.F. Wright, R. Saksgård, and B.O. Rosseland. 2011. Chemical and biological recovery of Lake Saudlandsvatn, a formerly highly acidified lake in southernmost Norway, in response to decreased acid deposition. *Science of the Total Environment*. 409(15): 2908-2916. https://doi.org/10.1016/j.scitotenv.2011.04.026

Hobbs, W.O., J.M.R. Hobbs, T. LaFrançois, K.D. Zimmer, K.M. Theissen, M.B. Edlund, N. Michelutti, M.G. Butler, M.A. Hanson, and T.J. Carlson. 2012. A 200-year perspective on alternative stable state theory and lake management from a biomanipulated shallow lake. *Ecological Applications*. 22(5): 1483-1496. https://doi.org/10.1890/11-1485.1

Hoiland, W.K., F.W. Rabe, and R.C. Biggam. 1994. Recovery of macroinvertebrate communities from metal pollution in the South Fork and mainstem of the Coeur d'Alene River, Idaho. *Water Environment Research*. 66(1): 84-88. https://doi.org/10.2307/25044410

Hokanson, K.E.F., C.F. Kleiner, and T.W. Thorslund. 1977. Effects of constant temperatures and diel temperature fluctuations on specific growth and mortality rates and yield of juvenile rainbow trout, *Salmo gairdneri [Oncorhynchus mykiss]*. *Journal of the Fisheries Research Board of Canada*. 34(5): 639-648. https://doi.org/10.1139/f77-100

Holcombe, G.W., D.A. Benoit, and E.N. Leonard. 1979. Long-term effects of zinc exposure on three generations of brook trout (*Salvelinus fontinalis*). *Transactions of the American Fisheries Society*. 108(1): 76-87. https://doi.org/10.1577/1548-8659(1979)108<76:LEOZEO>2.0.CO;2

Holcombe, G.W., D.A. Benoit, E.N. Leonard, and J.M. McKim. 1976. Long-term effects of lead exposure on three generations of brook trout (*Salvelinus fontinalis*). *Journal of the Fisheries Research Board of Canada*. 33(8): 1731-1741. https://doi.org/10.1139/f76-220

Hornberger, M.I., S.N. Luoma, M.L. Johnson, and M. Holyoak. 2009. Influence of remediation in a mine-impacted river: metal trends over large spatial and temporal scales. *Ecological Applications*. 19(6): 1522-1535. https://doi.org/10.1890/08-1529.1

Huckins, C.J., E.A. Baker, K.D. Fausch, and J.B.K. Leonard. 2008. Ecology and Life History of Coaster Brook Trout and Potential Bottlenecks in Their Rehabilitation. *North American Journal of Fisheries Management*. 28(4): 1321-1342. https://doi.org/10.1577/M05-191.1

IDEQ. 2016. The Idaho Department of Environmental Quality Water Body Assessment Guidance, Third Edition. Idaho Department of Environmental Quality, Boise, Idaho. 118 pp. http://www.deq.idaho.gov/water-quality/surface-water/monitoring-assessment.aspx.

Iftikar, Fathima I., V. Matey, and C.M. Wood. 2010. The Ionoregulatory Responses to Hypoxia in the Freshwater Rainbow Trout *Oncorhynchus mykiss*. *Physiological and Biochemical Zoology*. 83(2): 343-355. https://doi.org/10.1086/648566

Jackson, D. and G. Marx. 2016. Spills of pig waste kill hundreds of thousands of fish in Illinois (August 5, 2016). Chicago Tribune. https://www.chicagotribune.com/news/watchdog/pork/ct-pig-farms-pollution-met-20160802-story.html [Accessed November 12, 2018].

Jackson, J. 2018. Low oxygen levels responsible for fish kill below American Falls Dam. Idaho Department of Fish and Game (August 7, 2018). https://idfg.idaho.gov/blog/2018/08/low-oxygen-levels-responsible-fish-kill-below-american-falls-dam.

Jager, T., C. Albert, T.G. Preuss, and R. Ashauer. 2011. General Unified Threshold model for Survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental Science & Technology*. 45(7): 2529-2540. https://doi.org/10.1021/es103092a

Janz, D.M., D.K. DeForest, M.L. Brooks, P.M. Chapman, G. Gilron, D. Hoff, W.A. Hopkins, D.O. McIntyre, C.A. Mebane, V.P. Palace, J.P. Skorupa, and M. Wayland. 2010. Selenium toxicity to aquatic organisms. Pages 139-230 *in* P. M. Chapman, W. J. Adams, M. L. Brooks, C. G. Delos, S. N. Luoma, W. A. Maher, H. M. Ohlendorf, T. S. Presser, and D. P. Shaw, editors. *Ecological Assessment of Selenium in the Aquatic Environment*. Society of Environmental Toxicology and Chemistry (SETAC), Pensacola, Florida, http://dx.doi.org/10.1201/EBK1439826775-c6.

Joachim, S., H. Roussel, J.-M. Bonzom, E. Thybaud, C.A. Mebane, P.J. Van den Brink, and L. Gauthier. 2017. A long-term copper exposure on a freshwater ecosystem using lotic mesocosms: Invertebrate community responses. *Environmental Toxicology and Chemistry*. 36(10): 2698–2714. https://doi.org/10.1002/ETC.3822

Jones, A.R. 2003. Ecological Recovery of Amphipods on Sandy Beaches Following Oil Pollution: An Interim Assessment. *Journal of Coastal Research*. 66-73

Jones, H.P. and O.J. Schmitz. 2009. Rapid recovery of damaged ecosystems. *PLoS ONE*. 4(5): e5653. https://doi.org/10.1371/journal.pone.0005653

Jones, N.E. and I.C. Petreman. 2012. The Capacity to Detect Change Stream Fish Communities Characteristics at the Site-Level in the Lake Ontario Basin. *Environmental Management*. 50(1): 77-88. https://doi.org/10.1007/s00267-012-9859-9

Katano, O., J.I. Toi, K. Maekawa, and K. Iguchi. 1998. Colonization of an artificial stream by fishes and aquatic macro-invertebrates. *Ecological Research*. 13(1): 83-96. https://doi.org/10.1046/j.1440-1703.1998.00248.x

Kattwinkel, M., M. Liess, M. Arena, S. Bopp, F. Streissl, and J. Römbke. 2015. Recovery of aquatic and terrestrial populations in the context of European pesticide risk assessment. *Environmental Reviews*. 23(4): 382-394. https://doi.org/10.1139/er-2015-0013

Keller, A.A. and L. Cavallaro. 2008. Assessing the US Clean Water Act 303(d) listing process for determining impairment of a waterbody. *Journal of Environmental Management*. 86(4): 699-711. https://doi.org/10.1016/j.jenvman.2006.12.013

Keller, W., N.D. Yan, J.M. Gunn, and J. Heneberry. 2007. Recovery of acidified lakes: Lessons from Sudbury, Ontario, Canada. *Water, Air, and Soil Pollution: Focus*. 7(1-3): 317-322. https://doi.org/10.1007/s11267-006-9061-2

Kingsbury, P.D. and D.P. Kreutzweiser. 1987. Permethrin treatments in Canadian forests. Part 1: Impact on stream fish. *Pesticide Science*. 19(1): 35-48. https://doi.org/10.1002/ps.2780190106

Kowalik, R.A., D.M. Cooper, C.D. Evans, and S.J. Ormerod. 2007. Acidic episodes retard the biological recovery of upland British streams from chronic acidification. *Global Change Biology*. 13(11): 2439-2452. https://doi.org/10.1111/j.1365-2486.2007.01437.x

Kreutzweiser, D.P. 1990. Response of a Brook Trout (*Salvelinus fontinalis*) Population to a Reduction in Stream Benthos Following an Insecticide Treatment. *Canadian Journal of Fisheries and Aquatic Sciences*. 47(7): 1387-1401. https://doi.org/10.1139/f90-158

Kreutzweiser, D.P., R.C. Back, T.M. Sutton, K.L. Pangle, and D.G. Thompson. 2004. Aquatic mesocosm assessments of a neem (azadirachtin) insecticide at environmentally realistic concentrations—2: zooplankton community responses and recovery. *Ecotoxicology and Environmental Safety*. 59(2): 194-204. https://doi.org/https://doi.org/10.1016/j.ecoenv.2003.09.008

Kreutzweiser, D.P. and P.D. Kingsbury. 1987. Permethrin treatments in Canadian forests. Part 2: Impact on stream invertebrates. *Pesticide Science*. 19(1): 49-60. https://doi.org/10.1002/ps.2780190107

Krumholz, L.A. and W.L. Minckley. 1964. Changes in the Fish Population in the Upper Ohio River following Temporary Pollution Abatement. *Transactions of the American Fisheries Society*. 93(1): 1-5. https://doi.org/10.1577/1548-8659(1964)93[1:CITFPI]2.0.CO;2

Kubach, K.M., M.C. Scott, and J.S. Bulak. 2010. Recovery of a temperate riverine fish assemblage from a major diesel oil spill. *Freshwater Biology*. 56(3): 503-518. https://doi.org/10.1111/j.1365-2427.2010.02517.x

Lackmann, A.R., A.H. Andrews, M.G. Butler, E.S. Bielak-Lackmann, and M.E. Clark. 2019. Bigmouth Buffalo *Ictiobus cyprinellus* sets freshwater teleost record as improved age analysis reveals centenarian longevity. *Communications Biology*. 2(1): 197. https://doi.org/10.1038/s42003-019-0452-0

Lake, P.S. 2000. Disturbance, patchiness, and diversity in streams. *Journal of the North American Benthological Society*. 19(4): 573-592. https://doi.org/10.2307/1468118

Landers, M. 2016. 5 years after big fishkill, Ogeechee River has recovered, officials say. The Florida Times-Union (Jacksonville, FL), May 31, 2016. https://www.jacksonville.com/article/20160531/NEWS/801248862 [Accessed November 11, 2018].

Landrum, P.F., P.M. Chapman, J. Neff, and D.S. Page. 2013. Influence of exposure and toxicokinetics on measures of aquatic toxicity for organic contaminants: A case study review. *Integrated Environmental Assessment and Management*. 9(2): 196-210. https://doi.org/10.1002/ieam.1388

Landrum, P.F., J.A. Steevens, D.C. Gossiaux, M. McElroy, S. Robinson, L. Begnoche, S. Chernyak, and J. Hickey. 2004. Time-dependent lethal body residues for the toxicity of pentachlorobenzene to *Hyalella azteca*. *Environmental Toxicology and Chemistry*. 23(5): 1335-1343. https://doi.org/10.1897/03-164

Langford, T.E.L., P.J. Shaw, A.J.D. Ferguson, and S.R. Howard. 2009. Long-term recovery of macroinvertebrate biota in grossly polluted streams: Re-colonisation as a constraint to ecological quality. *Ecological Indicators*. 9(6): 1064-1077. https://doi.org/10.1016/j.ecolind.2008.12.012

LDWF. 2015. Pearl River Fish Kill Post Incident Monitoring Report 2012 – 2014. Louisiana Department of Wildlife and Fisheries (undated report, ca. 2015). . 66 pp. https://catalog.data.gov/dataset/pearl-river-fish-kill-post-incident-monitoring-report-2012-2014 [Accessed November 10, 2018].

Lefcort, H., J. Vancura, and E.L. Lider. 2010. 75 years after mining ends stream insect diversity is still affected by heavy metals. *Ecotoxicology*. 19(8): 1416-1425. https://doi.org/10.1007/s10646-010-0526-8

Leland, H.V., S.V. Fend, T.L. Dudley, and J.L. Carter. 1989. Effects of copper on species composition of benthic insects in a Sierra Nevada, California, stream. *Freshwater Biology*. 21(2): 163-179. https://doi.org/10.1111/j.1365-2427.1989.tb01356.x

Lemly, A.D. 1997. Ecosystem recovery following selenium contamination in a freshwater reservoir. *Ecotoxicology and Environmental Safety*. 36(3): 275-281. https://doi.org/10.1006/eesa.1996.1515

Liess, M., B.J. Pieters, and S. Duquesne. 2006. Long-term signal of population disturbance after pulse exposure to an insecticide: Rapid recovery of abundance, persistent alteration of structure. *Environmental Toxicology and Chemistry*. 25(5): 1326-1331. https://doi.org/10.1897/05-466R.1

Lytle, D.A. and B.L. Peckarsky. 2001. Spatial and temporal impacts of a diesel fuel spill on stream invertebrates. *Freshwater Biology*. 46(5): 693-704. https://doi.org/10.1046/j.1365-2427.2001.00695.x

Mangel, M. and M.V. Abrahams. 2001. Age and longevity in fish, with consideration of the ferox trout. *Experimental Gerontology*. 36(4): 765-790. https://doi.org/10.1016/S0531-5565(00)00240-0

Maxwell, D. and S. Jennings. 2005. Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. *Journal of Applied Ecology*. 42(1): 25-37. https://doi.org/doi: 10.1111/j.1365-2664.2005.01000.x

McDonald, D.G. and C.M. Wood. 1993. Branchial mechanisms of acclimation to metals in freshwater fish. *Fish Ecophysiology*. 297-321. https://doi.org/10.1007/978-94-011-2304-4_12

McFadden, J.T., G.R. Alexander, and D.S. Shetter. 1967. Numerical changes and population regulation in brook trout *Salvelinus fontinalis*. *Journal of the Fisheries Research Board of Canada*. 24(7): 1425-1459. https://doi.org/10.1139/f67-121

McKim, J.M., G.F. Olson, G.W. Holcombe, and E.P. Hunt. 1976. Long-term effects of methylmercuric chloride on three generations of brook trout (*Salvelinus fontinalis*): toxicity, accumulation, distribution, and elimination. *Journal of the Fisheries Research Board of Canada*. 33(12): 2726-2739

MDC. 2018. Missouri Pollution and Fish Kill Investigations, 2017. Missouri Department of Conservation. Central Regional Office and Conservation Research Center. Columbia, MO. 43 pp. www.mostreamteam.org [Accessed 31 January 2019].

MDE. 2017. 2016 Fish Kill Summary. Maryland Department of the Environment, Science Services Administration, Fish Kill Investigation Section. 15 pp. https://mde.state.md.us/programs/water/FishandShellfish/Documents/2016_FK_ANNUAL_Report.pdf.

Meade, R. 2004. Fish and invertebrate recolonization in a Missouri prairie stream after an acute pollution event. *North American Journal of Fisheries Management*. 24(1): 7-19. https://doi.org/10.1577/M02-094

Mebane, C.A. 2006. Cadmium risks to freshwater life: derivation and validation of low-effect criteria values using laboratory and field studies. U.S. Geological Survey Scientific Investigation Report 2006-5245 (2010 rev.). 130 pp. https://doi.org/10.3133/sir20065245.

Mebane, C.A. 2010. Relevance of risk predictions derived from a chronic species-sensitivity distribution with cadmium to aquatic populations and ecosystems. *Risk Analysis*. 30(2): 203-223. https://doi.org/10.1111/j.1539-6924.2009.01275.x

Mebane, C.A. and D.L. Arthaud. 2010. Extrapolating growth reductions in fish to changes in population extinction risks: copper and Chinook salmon. *Human and Ecological Risk Assessment*. 16(5): 1026-1065. https://doi.org/10.1080/10807039.2010.512243

Mebane, C.A., R.J. Eakins, B.G. Fraser, and W.J. Adams. 2015. Recovery of a mining-damaged stream ecosystem. *Elementa: Science of the Anthropocene*. 3(1): 000042. https://doi.org/10.12952/journal.elementa.000042

Mebane, C.A., J.P. Sumpter, A. Fairbrother, T.P. Augspurger, T.J. Canfield, W.L. Goodfellow, P.D. Guiney, A. LeHuray, L. Maltby, D.B. Mayfield, M.J. McLaughlin, L. Ortego, T. Schlekat, R.P. Scroggins, and T.A. Verslycke. 2019. Scientific Integrity Issues in Environmental Toxicology and Chemistry: improving research reproducibility, credibility, and transparency. *Integrated Environmental Assessment and Management*. 15(3): 320-344. https://doi.org/10.1002/ieam.4119

Meffe, G.K. and A.L. Sheldon. 1990. Post-Defaunation Recovery of Fish Assemblages in Southeastern Blackwater Streams. *Ecology*. 71(2): 657-667. https://doi.org/10.2307/1940320

Melaas, C.L., K.D. Zimmer, M.G. Butler, and M.A. Hanson. 2001. Effects of rotenone on aquatic invertebrate communities in prairie wetlands. *Hydrobiologia*. 459(1): 177-186. https://doi.org/10.1023/a:1012514124430

Miller, D.H., J.E. Tietge, M.E. McMaster, K.R. Munkittrick, X. Xia, D.A. Griesmer, and G.T. Ankley. 2015. Linking mechanistic toxicology to population models in forecasting recovery from chemical stress: A case study from Jackfish Bay, Ontario, Canada. *Environmental Toxicology and Chemistry*. 34(7): 1623-1633. https://doi.org/10.1002/etc.2972

Mills, K.H., S.M. Chalanchuk, and D.J. Allan. 2000. Recovery of fish populations in Lake 223 from experimental acidification. *Canadian Journal of Fisheries and Aquatic Sciences*. 57(1): 192-204. https://doi.org/10.1139/f99-186

Minshall, G.W., D.A. Andrews, and C.Y. Manuel-Faler. 1983. Application of island biogeographic theory to streams: macroinvertebrate recolonization of the Teton River, Idaho. Pages 279-297 *in* J. R. Barnes, and G. W. Minshall, editors. *Stream Ecology: Application and Testing of General Ecological Theory*. Plenum Press, New York and London.

Monteith, D.T., A.G. Hildrew, R.J. Flower, P.J. Raven, W.R.B. Beaumont, P. Collen, A.M. Kreiser, E.M. Shilland, and J.H. Winterbottom. 2005. Biological responses to the chemical recovery of acidified fresh waters in the UK. *Environmental Pollution*. 137(1): 83-101. https://doi.org/https://doi.org/10.1016/j.envpol.2004.12.026

Munkittrick, K.R., C.J. Arens, R.B. Lowell, and G.P. Kaminski. 2009. A review of potential methods for determining critical effect size for designing environmental monitoring programs. *Environmental Toxicology and Chemistry*. 28(7): 1361-1371. https://doi.org/10.1897/08-376.1

Munthe, J., R.A. Bodaly, B.A. Branfireun, C.T. Driscoll, C.C. Gilmour, R. Harris, M. Horvat, M. Lucotte, and O. Malm. 2007. Recovery of mercury-contaminated fisheries. *Ambio*. 36(1): 33-44. https://doi.org/10.1579/0044-7447(2007)36[33:ROMF]2.0.CO;2

Murphy, J.F., J.H. Winterbottom, S. Orton, G.L. Simpson, E.M. Shilland, and A.G. Hildrew. 2014. Evidence of recovery from acidification in the macroinvertebrate assemblages of UK fresh waters: A 20-year time series. *Ecological Indicators*. 37, Part B(0): 330-340. https://doi.org/10.1016/j.ecolind.2012.07.009

Myers, R.A., K.G. Bowen, and N.J. Barrowman. 1999. Maximum reproductive rate of fish at low population density. *Canadian Journal of Fisheries and Aquatic Sciences*. 56(12): 2404–2419. https://doi.org/10.1139/cjfas-56-12-2404

NCDEQ. 2016. Water Quality Assessment Methods for Toxics. North Carolina Department of Environmental Quality, Division of Water Resources, Water Sciences Section and Planning Section. 11 pp.

https://files.nc.gov/ncdeq/Water%20Quality/Planning/TMDL/303d/2016/Assessment%20Method%20for%20Toxics-April-1-2016.pdf [Accessed Nov 10, 2018].

Nelson, S.M. and R. Roline. 1996. Recovery of a stream macroinvertebrate community from mine drainage disturbance. *Hydrobiologia*. 339(1-3): 73-84. https://doi.org/10.1007/bf00008915

Netherland, M.D., K.D. Getsinger, and J.D. Skogerboe. 1997. Mesocosm Evaluation of the Species-Selective Potential of Fluridone. *Journal of Aquatic Plant Management*. 35: 41-50

Niemi, G.J., P. DeVore, N. Detenbeck, D. Taylor, A. Lima, J. Pastor, J.D. Yount, and R.J. Naiman. 1990. Overview of case studies on recovery of aquatic systems from disturbance. *Environmental Management*. 14(5): 571-587. https://doi.org/10.1007/BF02394710

O'Connor, T.P. 2001. Comparing episodic, chronic, fishing, and non-fishing impacts on resource populations. *Marine Pollution Bulletin*. 42(7): 532-535

Parsons, J., K. Hamel, J.D. Madsen, and K.D. Getsinger. 2001. The use of 2,4-D for selective control of an early infestation of Eurasian watermilfoil in Loon Lake, Washington. *Journal of Aquatic Plant Management*. 39: 117-124

Paulsson, K. and K. Lundbergh. 1991. Treatment of mercury contaminated fish by selenium addition. *Water Air & Soil Pollution*. 56(1): 833-841. https://doi.org/10.1007/bf00342320

Payer, R.D. and C.G. Scalet. 1978. Population and Production Estimates of Fathead Minnows in a South Dakota Prairie Wetland. *The Progressive Fish-Culturist*. 40(2): 63-66. https://doi.org/doi:10.1577/1548-8659(1978)40[63:PAPEOF]2.0.CO;2

Peterman, R.M. and M. M'Gonigle. 1992. Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin*. 24(5): 231-234. https://doi.org/10.1016/0025-326X(92)90559-O

Phinney, D.E. 1975. Repopulation of an Eradicated Stream Section by Brook Trout. *Transactions of the American Fisheries Society*. 104(4): 685-687. https://doi.org/doi:10.1577/1548-8659(1975)104<685:ROAESS>2.0.CO;2

Piller, K.R. and A.D. Geheber. 2015. Black liquor and the hangover effect: fish assemblage recovery dynamics following a pulse disturbance. *Ecology and Evolution*. 5(12): 2433-2444. https://doi.org/10.1002/ece3.1530

Rabeni, C.F., S.P. Davies, and K.E. Gibbs. 1985. Benthic invertebrate response to pollution abatement: structural changes and functional implications. *JAWRA Journal of the American Water Resources Association*. 21(3): 489-497. https://doi.org/10.1111/j.1752-1688.1985.tb00161.x

Raddum, G.G. and A. Fjellheim. 2003. Liming of River Audna, Southern Norway: A large-scale experiment of benthic invertebrate recovery. *Ambio*. 32(3): 230-234. https://doi.org/10.1579/0044-7447(2003)032[0230:lorasn]2.0.co;2

Resh, V.H., A.V. Brown, A.P. Covich, M.E. Gurtz, H.W. Li, G.W. Minshall, S.R. Reice, A.L. Sheldon, J.B. Wallace, and R.C. Wissmar. 1988. The Role of Disturbance in Stream Ecology. *Journal of the North American Benthological Society*. 7(4): 433-455. https://doi.org/10.2307/1467300

Roberts, S. 2016a. Animas River 2015 Benthic Macroinvertebrate (BMI) Report. Prepared for: U.S. Environmental Protection Agency, Denver, CO. Mountain Studies Institute, Durango, CO. 138 pp. http://www.mountainstudies.org/goldkingspill/ [Accessed 30Nov2018].

Roberts, S. 2016b. Animas River Water Quality at Rotary Park, Durango, Colorado: Gold King Mine Release Monitoring. Prepared for: U.S. Environmental Protection Agency, Denver, CO. Mountain Studies Institute, Durango, CO. 103 pp. http://www.mountainstudies.org/goldkingspill/ [Accessed 30Nov2018].

Roghair, C.N. and C.A. Dolloff. 2005. Brook trout movement during and after recolonization of a naturally defaunated stream reach. *North American Journal of Fisheries Management*. 25(3): 777–784. https://doi.org/10.1577/M04-034.1

Root, K.V. 2002. Evaluating risks for threatened aquatic species: the shortnose sturgeon in the Connecticut River. *American Fisheries Society Symposium*. 28: 45-54

Rose, K.A. and J.H. Cowan, Jr. 2000. Predicting fish population dynamics: compensation and the importance of site-specific considerations. *Environmental Science and Policy*. 3(Supplement): S433-S443. https://doi.org/10.1016/S1462-9011(00)00054-X

Rowe, C.L. 2008. "The Calamity of So Long Life": Life histories, contaminants, and potential emerging threats to long-lived vertebrates. *BioScience*. 58(7): 623-631. https://doi.org/10.1641/b580709

Ryon, M.G. 2011. Recovery of fish communities in a warm water stream following pollution abatement. *Environmental Management*. 47(6): 1096-1111. https://doi.org/10.1007/s00267-010-9596-x

SAB. 2006. EPA Science Advisory Board (SAB) Consultation on a Proposed Framework for Revising the Guidelines for Deriving Water Quality Criteria for Protection of Aquatic Life. Science Advisory Board, U.S. Environmental Protection Agency U.S Environmental Protection Agency EPA-SAB-CON-06-004. 151 pp.

Schmidt, T.S., H.A. Rogers, J.L. Miller, C.A. Mebane, and L.S. Balistrieri. 2018. Understanding the captivity effect on invertebrate communities transplanted into an experimental stream laboratory. *Environmental Toxicology and Chemistry*. 37(11): 2820-2834. https://doi.org/10.1002/etc.4237

Schneider, J. 2010. White River Fish Kill: 10 years of recovery (January 6, 2010). The Herald-Bulletin (Anderson, IN). http://www.heraldbulletin.com/news/business/white-river-fish-kill-years-of-recovery/article_4db0b8f8-04b5-57db-a39d-154d30dd51d0.html [Accessed November 11, 2018].

Schrank, A.J., F.J. Rahel, and H.C. Johnstone. 2003. Evaluating Laboratory-Derived Thermal Criteria in the Field: An Example Involving Bonneville Cutthroat Trout. *Transactions of the American Fisheries Society*. 132(1): 100–109. https://doi.org/10.1577/1548-8659(2003)132<0100:ELDTCI>2.0.CO;2

Scott, W.B. and E.J. Crossman. 1998. Freshwater fishes of Canada. Galt House Publishing, Oakville, Ontario. 966 pp

Sedell, J.R., G.H. Reeves, F.R. Hauer, J.A. Stanford, and C.P. Hawkins. 1990. Role of refugia in recovery from disturbances; modern fragmented and disconected river systems. *Environmental Management*. 14(5): 711-724

Sibley, P.K., N.K. Kaushik, and D.P. Kreutzweiser. 1991. Impact of a pulse application of permethrin on the macroinvertebrate community of a headwater stream. *Environmental Pollution*. 70(1): 35-55. https://doi.org/10.1016/0269-7491(91)90130-O

Smith, J.G. 2003. Recovery of the Benthic Macroinvertebrate Community in a Small Stream After Long-Term Discharges of Fly Ash. *Environmental Management*. 32(1): 77-92. https://doi.org/10.1007/s00267-003-2962-1

Stanford, B., H. Jones, and E. Zavaleta. 2018. Meta-analysis of the effects of upstream land cover on stream recovery. *Conservation Biology*. https://doi.org/10.1111/cobi.13189

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman, and W.A. Brungs. 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of

Aquatic Organisms and their Uses. U.S. Environmental Protection Agency, EPA 822-R-85-100, NTIS PB85 227049, Duluth, Narragansett, and Corvallis. 98 pp.
Strong, D.R., Jr. 1972. Life history variation among populations of an amphipod (*Hyalella azteca*). *Ecology*. 53(6): 1103-1111. https://doi.org/10.2307/1935422
Sweetapple, C., M. Astaraie-Imani, and D. Butler. 2018. Design and operation of urban wastewater systems considering reliability, risk and resilience. *Water Research*. 147: 1-12. https://doi.org/https://doi.org/10.1016/j.watres.2018.09.032
Swift, M.C. 2002. Stream ecosystem response to, and recovery from, experimental exposure to selenium. *Journal of Aquatic Ecosystem Stress and Recovery*. 9(3): 159 - 184. https://doi.org/10.1023/A:1021299003516
Trudel, M. and J.B. Rasmussen. 1997. Modeling the Elimination of Mercury by Fish. *Environmental Science & Technology*. 31(6): 1716-1722. https://doi.org/10.1021/es960609t
USEPA. 1985. Technical support document for water quality-based toxics control. Office of Water, U.S. Environmental Protection Agency, EPA 440/4-85-032, Washington, D.C. 117 pp.
USEPA. 1991. Technical support document for water quality-based toxics control. Office of Water, U.S. Environmental Protection Agency, EPA 505/2-90-001, Washington, D.C. 143 pp. http://www.epa.gov/npdes/pubs/owm0264.pdf.
USEPA. 1997. Guidelines for Preparation of the Comprehensive State Water Quality Assessments (305(b) Reports) and Electronic Updates: Supplement. U.S. Environmental Protection Agency EPA-841-B-97-002B. 132 pp. https://www.epa.gov/waterdata/guidelines-preparation-comprehensive-state-water-quality-assessments-305b-reports-and [Accessed November 39, 2018].
USEPA. 2003. EPA Region 10 guidance for Pacific Northwest state and tribal temperature water quality standards. U.S. Environmental Protection Agency, EPA 910-B-03-002, Seattle, WA. 49 pp. http://yosemite.epa.gov/r10/water.nsf/Water+Quality+Standards/WQS+Temperature+Guidance.
USEPA. 2016. Aquatic Life Ambient Water Quality Criterion for Selenium – Freshwater 2016. 822-R-16-006, U.S. Environmental Protection Agency. 807 pp. https://www.epa.gov/wqc/aquatic-life-criterion-selenium [Accessed February 21, 2018].
USEPA. 2017. Draft Aquatic Life Ambient Water Quality Criteria for Aluminum. U.S. Environmental Protection Agency, EPA-822-P-17-001. 278 pp. https://www.epa.gov/wqc/2017-draft-aquatic-life-criteria-aluminum-freshwater.
Van Urk, G., F. Kerkum, and C.J. Van Leeuwen. 1993. Insects and insecticides in the Lower Rhine. *Water Research*. 27(2): 205-213. https://doi.org/https://doi.org/10.1016/0043-1354(93)90077-U
Van Walleghem, J.L.A., P.J. Blanchfield, L.E. Hrenchuk, and H. Hintelmann. 2013. Mercury Elimination by a Top Predator, *Esox lucius*. *Environmental Science & Technology*. 47(9): 4147-4154. https://doi.org/10.1021/es304332v
Vandenbos, R.E., W.M. Tonn, and S.M. Boss. 2006. Cascading life-history interactions: alternative density-dependent pathways drive recruitment dynamics in a freshwater fish. *Oecologia*. 148(4): 573-582. https://doi.org/10.1007/s00442-006-0410-7
Wallace, J.B., D.S. Vogel, and T.F. Cuffney. 1986. Recovery of a Headwater Stream from an Insecticide-Induced Community Disturbance. *Journal of the North American Benthological Society*. 5(2): 115-126. https://doi.org/10.2307/1467866

Ward-Campbell, B., K. Cottenie, N.E. Mandrak, and R. McLaughlin. 2017. Fish assemblages in agricultural drains are resilient to habitat change caused by drain maintenance. *Canadian Journal of Fisheries and Aquatic Sciences*. 74(10): 1538-1548. https://doi.org/10.1139/cjfas-2016-0361

Watanabe, N.C., S. Harada, and Y. Komai. 2000. Long-term recovery from mine drainage disturbance of a macroinvertebrate community in the Ichi-kawa River, Japan. *Hydrobiologia*. 429(1-3): 171-180. https://doi.org/10.1023/A:1004027201667

Webster, N.I., W.B. Keller, and C.W. Ramcharan. 2013. Restoration of zooplankton communities in industrially damaged lakes: Influences of residual metal contamination and the recovery of fish communities. *Restoration Ecology*. 21(6): 785–792. https://doi.org/10.1111/rec.12013

White, J.N. 2016. Assessment of the Fish Community in the Animas River After the Gold King Mine Spill (slide deck). Colorado Department of Parks and Wildlife. 59 pp. https://animas.nmwrri.nmsu.edu/wp-content/uploads/SpeakerPresentations/JimNWhite.pdf [Accessed 30Nov2018].

Willmsen, C. and L.V. Mapes. 2017. A disaster years in the making: A cascade of errors led to catastrophe at the West Point treatment plant. *Seattle Times*. https://projects.seattletimes.com/2017/west-point/ [Accessed February 7, 2018].

Winemiller, K.O. 2005. Life history strategies, population regulation, and implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*. 62(4): 872-885. https://doi.org/10.1139/f05-040

Wood, C.M. 2012. An introduction to metals in fish physiology and toxicology: Basic principles. *Fish Physiology: Homeostasis and Toxicology of Essential Metals*. 31(PART A): 1-51. https://doi.org/10.1016/S1546-5098(11)31001-1

Yount, J.D. and G.J. Niemi. 1990a. Preface [to a collection of 17 articles on recovery in aquatic ecosystems]. *Environmental Management*. 14(5): 515-516. https://doi.org/10.1007/bf02394706

Yount, J.D. and G.J. Niemi. 1990b. Recovery of lotic communities and ecosytems from disturbance - a narrative review of case studies. *Environmental Management*. 14(5): 547-569. https://doi.org/10.1007/BF02394709

Zahner, H.M. 2009. Recovery of fathead minnows (*Pimephales promelas*) following episodic copper exposure: a biochemical, physiological, individual, and population perspective. PhD Dissertation. 247 pp. Environmental Toxicology, Clemson University, http://tigerprints.clemson.edu/all_dissertations/360.

Zahner, H.M., E.J. Van Genderen, J.R. Tomasso, and S.J. Klaine. 2006. Whole-body sodium concentration in larval fathead minnows (*Pimephales promelas*) during and following copper exposure. *Environmental Toxicology and Chemistry*. 26(6): 1635-1639. https://doi.org/10.1897/05-543R.1

Appendix 1: Freshwater disturbance and recovery case study summaries

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Fish community, crayfishes	Warmwater streams in N. Mississippi	Severe drought (Press, 2)	Fish and crayfish abundance and richness	9 mo, abundance and richness similar to pre-desiccation; >1 yr - size structure	1 year	Despite resilience to stream desiccation, effects on fish species composition and fish and crayfish size structure persisted 1 year after the drought.	(Adams and Warren 2005)
Fish community	Upper Sacramento River, CA	Fungicide spill from train derailment virtually eliminated life for a 35 mile reach (Pulse, 2)	Time for abundances of common fish taxa to reach carrying capacity and pre-spill abundance	Rainbow trout, 3-4 years; Spotted bass, 4 yrs; Riffle sculpin, >10 yrs; suckers, >10 yrs	10 years	Unaffected tributaries provided colonization sources for trout and sculpin. Highly mobile trout were present within 1-year; Source populations more distant for long-lived suckers	(Allen and Gast 2005)
Benthic invertebrates	Vistre River, France	Wastewater effluent (press, 3)	Trait-based or gross taxonomic composite measures	3 mo - Major taxa endpoints; ~2 years for traits	4 years	Major taxa (diversity, % mayflies, % flies, etc) increased w/in 3 mo; although when grouped by function (e.g., feeding type) some took 2 years	(Arce et al. 2014)
Fish community	Adirondack lakes (NY)	Acidification (Press, 1)	Fish-community richness, total fish abundance, and brook trout	Incomplete recovery despite increased pH and decreased Al	>28 years	Limited connectivity between lakes constrains natural recolonization. May have reached alternative stable state and will not return to pre-acidic conditions.	(Baldigo et al. 2016)
Benthic invertebrate community	Artificial streams	Insecticide pulse (4)	Taxonomic similarity to reference	> 0,6 years for long-lived uni- and semivoltine taxa	0.6 years	Short lived multivoltine taxa recovered within 0.2 years.	(Beketov et al. 2008)
Benthic macroinvertebrates	Small streams in Blue Mountains, Australia	Sewage abatement (press, 3)	Multivariate community analyses	~1 yr	7 years	Regulatory water chemistry limits recovered in only 1 month, presence of other urban nonpoint pollutants likely delayed recover	(Besley and Chessman 2008)
Benthic macroinvertebrates	Green River and tributaries, Wyoming	Large scale (~120 miles) rotenone fish	Taxa re-occurrence	8 months for most taxa; >2 years for some long-lived taxa	2 years	Slowest taxa to return were among the most tolerant taxa to many pollutants – Tipula and Hydropsychidae caddisflies.	(Binns 1967)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
		eradication (Pulse, 2)				Longer-live span assumed as the reason.	
Fish populations	Green River and tributaries, Wyoming	Large scale (~120 miles) rotenone fish eradication (Pulse, 2)	Taxa re-occurrence and age structure	No recovery of some species (which was their purpose); 2 years for game species	2 years	Native Colorado pikeminnow and humpback sucker and nonnative channel catfish and bullheads had not been detected and were expected to be permanently eradicated as a dam blocked recolonization.	(Binns 1967)
Fish population	Experimental lake	Estrogenic chemicals (EE2) (Press, 3)	Physiologic abnormalities, abundance, size structure	4 years	7 years	Estrogenic chemicals added to experimental lake; time to recovery was from cessation of additions.	(Blanchfield et al. 2015)
Fish and invertebrate community	Stream in NZ	1-h pulse-dose of chelated Cu (Pulse, 3)	Biofilm, invertebrate, and fish community	Biofilms-6 wk Insects: 77dd Trout – 10 mo	77 d (10-mo for fish)	Mayflies had a delayed response at some sites (21 d post application); unexpected persistent trout mortality/reduced numbers	(Clearwater et al. 2011)
Benthic invertebrate community	Rocky streams in Virginia and Colorado	Colonization of denuded rock baskets (Pulse, 4)	Taxa occurrence and abundance	30 days	~30 days	Colonization times of completely denuded, small areas (<1 m2) is considered here to reflect recovery times for locally severe impacts from disturbances such as small-scale, suction dredge mining or operating construction equipment in a stream	(Clements et al. 1989; Clements 1999; Schmidt et al. 2018)
Fish population and Invertebrate community	Arkansas River, CO	Mining contaminated streams (Press, 2/3)	Brown trout abundance, and invertebrate community structure	Trout – 3 yrs Mayfly richness – 1 yr	17 yrs	Mayfly richness increased within a year at sites close to colonizing sources following pollution controls. Several metals elevated, but Zn is probably the dominant toxicant/	(Clements et al. 2010)
Fish populations	Pigeon River, NC & TN)	Chronic industrial wastewater (Press, 2)	Increase in species richness	~5 years after mill modernization	21 years	"Recovery" consisted of increased numbers of native fish taxa and index of biotic integrity scores, but many missing or depressed taxa	(Coombs et al. 2010)
Benthic invertebrate community	White River, Indiana	Wastewater treatment	Similarity to upstream reference	3 years after upgrades came online	7 years (2 pre, 5 post)	Downstream community became similar to upstream taxa (dominated by mayflies and caddisflies which are	(Crawford et al. 1992)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
		upgrade (Press, 2)				generally intolerant of organic pollution). Prior to upgrades, dominated by oligochaete worms and chironomid midges	
Fish populations	Hyco Reservoir, NC	Selenium in powerplant discharges (Press, 3)	Fish abundances and biomass compared to reference	5 years	8 years	Little change in overall fish numbers because of replacement by tolerant species	(Crutchfield 2000)
Fish population and Invertebrate community	Stream in Tasmania	Pyrethroid insecticide spray drift	Benthic metrics, fish	Benthos abundance, 6 months; Structure >1 year, direct effects to fish, 1 month	1 year	Indirect effects to fish (growth delays) following reduction in food.	(Davies and Cooke 1993)
Benthic invertebrate community	Stream in Colorado	Tanker truck petroleum spill (Pulse, 3)	Species occurrences and abundances	>2.5 years	2.5 years	Tolerant groups replaced sensitive, long-lived taxa (e.g., perlodid stoneflies, and other mayflies, stoneflies, and caddisflies. Slow recovery despite nearby, upstream source populations attributed to slow reproduction of long-lived stoneflies	(Duggan et al. 2018)
Fish community	Stream in Colorado	Tanker truck petroleum spill (3)	Species occurrences and abundances	>5 years	>5 years	Little recovery of brown trout and mottled sculpin 5-years after spill	(Duggan et al. 2018)
Fish populations	Forest streams in Idaho	Wildfire	Reoccupation and abundance	<1 year	4 years	In larger interconnected systems, fish populations were resilient to the effects of massive fire. Extirpated populations in small streams were recolonized by dispersal from refuges or nearby streams	(Dunham et al. 2003)
Phyto- and zooplankton	Cazenovia Lake, New York,	Cu sulfate algicide; low-level and short duration (pulse, 4)	Bacteria and phytoplankton populations	1 week	3 months	2-5 d treatments of 5 µg/L Cu. No clear effects on primary productivity, zooplankton, or macrophytes.	(Effler et al. 1980)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Fish community	Small stream, Virginia, USA	Manure spill killed all fish in 6 km section of stream (2)	5 of 11 fish species recovered in the time period	>11 months	11 months	Species with high parental care showed rapid recovery within observation period.	(Ensign et al. 1997)
Ecological structure and function	Desert stream in Arizona	Flash flood (3)	Benthic and algal composition and biomass; algal numerous	1 month	2 months	Algal responses and functional measures recovered within 2-weeks; flood eliminated algae and 98% of benthics	(Fisher et al. 1982)
Benthic invertebrate community	Saskatchewan River	Large scale (100 mile) insecticide application (Pulse, 2)	Major taxa abundance	4-5 weeks for Plecoptera; 2 to 10 weeks for Simuliidae	10 weeks	Single high concentration slug of methoxychlor. Within 1 to 3 weeks, populations of most taxa had surpassed pre-treatment abundances, except for blackflies (Simulium, the target) and non-target stoneflies (Plecoptera). Chironomidae, Ephemeroptera, and Trichoptera were least affected and recovered more quickly	(Fredeen 1975)
Zooplankton community	Little Rock Lake, Wisconsin	Experimental acidification (Press, 2)	Species trajectories and multivariate community analyses	10 years	16 years	Recovery of 40% of zooplankton was delayed by 1-6 years by community inertia from tolerant species that thrived during acidification.	(Frost et al. 2006)
Fish and invertebrate community	Lentic mesocosms	Pyrethroid insecticide applications (Pulse, 3)	Multiple aquatic invertebrate major taxa; fish populations	< 1 year; most populations recovered within weeks	Varied, up to 1 year	Rapid recovery of zooplankton; amphipods slowest, insects intermediate. Hypothesized indirect effects to fish through loss of food not detected	(Giddings et al. 2001)
Freshwater Shrimp populations	Tropical stream (Puerto Rico)	Chlorine-bleach harvest poisoning (Pulse, 3)	Population abundance and size structure; leaf breakdown	<3 months	3 months	Rapid recovery attributed to nearby sources of organisms for colonization, mobility of dominant organisms, unimpaired habitat, and rapid flushing and processing of chlorine	(Greathouse et al. 2005)
Benthic invertebrate community	Coldwater stream	Suction dredging for small-scale gold mining (4)	Recolonization of dredged plots	<38 days	38 days	Authors considered stream recolonization to be best case because of limited turbidity plumes and absence of disturbance outside treatment plots	(Griffith and Andrews 1981)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4-mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Zooplankton community	Lentic mesocosms	Pyrethroid insecticide applications	Community composition; taxa abundances	77-105 days for daphnids; 28 days for community composition	14 months	More isolated ponds slower to recover	(Hanson et al. 2007)
Benthic invertebrate communities	Streams in Yellowstone National Park (Hellroaring and Tower Creeks)	Large scale, indiscriminate, aerial DDT watershed spraying (Pulse, 2)	Major taxa abundance	3 years for most major taxa; recovery mostly complete within 1 year in the smaller, less affected stream	5 years	Aerial spraying of DDT over a large portion of Yellowstone National Park eliminated most aquatic insect life in study streams within a few days. Recovery of major taxa occurred within 3 years although composition varied; one formerly abundant caddisfly family (Leptoceridae) was not found within the study period	(Hastings et al. 1961)
Fish populations	Streams	Mt. St. Helens eruption (Pulse, 1)	Trout and sculpin abundances	Sculpins 5 years; trout >8 years	8 years	Recolonization dependence on proximity to refugia; post-eruption habitat remained unfavorable to trout	(Hawkins and Sedell 1990)
Benthic invertebrate communities	Coldwater stream	Major ions and metals from mine drainage (press, 2)	Benthic invertebrate taxa richness, multivariate analyses	12-14 years	17 years	Long recovery times resulted from incomplete and incremental pollution controls. Recovery interpreted as time for taxa richness to consistently exceed 10th percentile of reference sites	(Herbst et al. 2018)
Fish, insect, and Daphnia populations	Acidified lake, S. Norway	Acidification (press, 2)	Brown trout, acid-tolerant caddisfly, and acid sensitive Daphnia	1-2 years	12 years	Biological recovery 1-2 years post chemical suitability possible because physical barriers to repopulation were absent	(Hesthagen et al. 2011)
Fish populations	Beaver Creek, IL	Manure land application (pulse, 2)	Recolonization of fish species	>2 years	2 years	9 fish species had not been detected since the fish kill and 18 others had not returned to previous levels.	(Jackson and Marx 2016)
Fish populations	Belews Lake, NC	Selenium discharges (Press, 2)	Fish community composition stabilized	20 years	30	~2-3 years after Se loading reduced, overall biomass recovered and sensitive taxa began to recolonize; Species richness reached baseline at ~10 years, but relative composition in flux. By 20 years assemblage had stabilized in a new equilibrium	(Janz et al. 2010)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Invertebrate community	Lotic mesocosms	Copper dosing for 2 years (Press, 2)	Recovery of zooplankton and macroinvertebrate abundances	1 year	1.5 years	Partial recovery of some taxonomic groups during the 2nd year of exposure, most likely from acquired tolerance	(Joachim et al. 2017)
Amphipod populations	Sydney Harbor	Oil pollution (pulse, 3-4)	Recovery of population abundances of *Exeodiceros fossor*	4 months for lightly oiled beaches; > 9 months if more severely oiled	9 months	Beaches that had been subject to cleanup efforts showed less recovery than naturally recovering beaches.	(Jones 2003)
Fish and invertebrate community	Artificial stream	Desiccation (pulse, 3)	Colonization of fish and invertebrate	100 days	120 days	Species richness and diversity in the experimental stream reached almost the same level as that in the permanent streams.	(Katano et al. 1998)
Zooplankton communities	Lakes near Sudbury, Ontario	Acidification from legacy smelter emissions (press, 2)	Species trajectories and multivariate community analyses	>25 years	25 years	Systems may have reached alternative stable states which differ from reference.	(Keller et al. 2007; Webster et al. 2013)
Fish populations	Slow and fast moving coldwater streams (New Brunswick, Canada)	Direct insecticide (permethrin) application (3)	Abundance of salmonids, cyprinids, and sculpin; growth of salmonids	4 months	4 years	Reductions in salmonid growth rates and reductions in fish densities in treated areas, presumably due to emigration, following severe impacts on aquatic invertebrates. No apparent effect on sculpin and cyprinids	(Kingsbury and Kreutzweiser 1987)
Fish populations	Headwater forest stream (Icewater Cr, ON, Canada)	Direct insecticide (permethrin) application (3)	Brook trout numbers, size structure, and growth	1 year	4 years	Reductions in growth were an apparent indirect effect following catastrophic drift of macroinvertebrates.	(Kreutzweiser 1990)
Benthic invertebrate communities	Headwater forest streams (Quebec and New Brunswick, Canada)	Direct insecticide (permethrin) application at differing rates and single or double doses (3)	Drift; abundance of major taxa	3-6 weeks (lower rates, single pulse); 1.5 years; higher rates or double pulses	"Several field seasons"	Lower rates: Despite massive drift increase and rapid depletion of benthic fauna, benthos density recover to or above pre-spray levels, or to levels comparable with those of untreated controls, by 3-6 weeks; higher or double rates, up to 1.5 years.	(Kreutzweiser and Kingsbury 1987)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Zooplankton communities	Lentic Mesocosms	Insecticide (neem)	Zooplankton community structure	>1 yr	1 yr	Selective toxicity (copepods much more sensitive than cladocerans)	(Kreutzweiser et al. 2004)
Fish community	Ohio River locks	Steel manufacturing effluents	Species presence	11 days	11 days	A labor strike abruptly shut down mills and discharges, and fish re-occupied newly inhabitable reaches, presumably from tributary source populations	(Krumholz and Minckley 1964)
Fish community	Medium river, South Carolina, USA	Diesel oil pipeline spill killed all fish in 37 km section of stream (2A)	Fish assemblage similarity	4 years	9.3 years	Recovery of the uppermost disturbed site was faster than the other disturbed sites because of its proximity to the undisturbed main stem fish assemblage, whereas the downstream sites were slower to recover largely because of isolation by anthropogenic barriers.	(Kubach et al. 2010)
Benthic invertebrate communities	Streams in the English Midlands	Severe, watershed wide industrial pollution that diminished over time (Press,1)	Community metrics and family occurrence	> 30 years	ca. 50 years	At the site most isolated from potential sources of colonizing taxa, no clean-water macroinvertebrate taxa were recorded 30 years after the major sources of pollution ceased.	(Langford et al. 2009)
Benthic invertebrate communities	Streams in the English Midlands	Severe industrial pollution that diminished over time (Press,2)	Community metrics and family occurrence	2- 5 years	ca. 50 years	Where clean-water colonizers were readily available, significant improvements in ecological quality followed within 2–5 years of the improvements in chemical quality.	(Langford et al. 2009)
Fish and mussel populations	Pearl River, Louisiana	Wastewater upset at a pulp and paper mill resulting in a large (40 mile) fish kill (pulse, 3)	Comparisons to pre-spill occurrence and numbers of fish	2 months – overall fish assemblage; >4 years for mussel taxa	4 years	Of the 20 most common mussel species found in 2007 pre-spill surveys, 18 were less abundant in a 2011 post-spill survey. Many long-lived paddlefish and sturgeon were killed in the spill but were too rare to assess in general assemblage sampling. Side channel refuges mitigated effects	(LDWF 2015; Piller and Geheber 2015)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Benthic invertebrate communities	Stream in Sierra Nevada mountains, California	Field experiment with Cu (Press, 4)	Taxa abundance	>0.92 years	0.92 years	Limited recovery after 11 months, owing to lack of recolonization by several semivoltine taxa (e.g., stoneflies and elmid beetles) whereas univoltine mayflies and caddisflies thrived in their absence	(Leland et al. 1989)
Daphnia magna populations	Laboratory	Pyrethroid insecticide pulse (pulse, 3)	Abundance and size structure	16 days – abundance; >60 days – age structure	60 days	Abundance in affected treatments recovered within 2 generations (1 generation = 8d); but age structure was still altered at test end	(Liess et al. 2006)
Benthic invertebrate communities	Forest stream in New York	Diesel spill (2)	Community similarity to reference	>15 months for species richness	15 months	Contamination not measured but delayed recovery at locations nearest spill suggests ongoing contamination prevented recovery	(Lytle and Peckarsky 2001)
Fish and invertebrate community	Prairie stream, Missouri, USA	31 km chicken manure spill (2A)	Sport fish population metrics and invertebrate diversity and abundance	>4.6 years for fish community; ~ 1 year for invertebrates.	4.6 years	Large fish (Largemouth bass and bluegill) abundance remained low and the end of the study	(Meade 2004)
Fish and invertebrate community	Stream in Idaho	1-week pulse of Cd (4)	Aquatic insect community and fish populations above and below pipeline break	<2 months	2 months	Pulse from a mining-mill water pipeline break in waste rock exceeded Cd chronic criteria by about 3X for about a week. No effects on fish populations and very minor effects to invertebrates when surveyed 2 months later.	(Mebane 2006)
Benthic macroinvertebrate communities	1st to 5th order coldwater streams, Idaho	Elevated metals downstream of mining disturbances (Press, 2)	Benthos – colonization and abundances of sensitive taxa;	~ 4 years most recovery occurred within about 4 years after criteria mostly attained.	21 years	Invertebrates steadily increased in diversity and abundance as water quality improved, but without obvious recovery milestones. Plateau in recovery at about 80% of reference diversity appears related to Co, a substance without ALC. Some taxa appear constrained at Cu concentrations <1X chronic criteria.	(Mebane et al. 2015)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Fish communities	1st to 5th order coldwater streams, Idaho	Elevated metals downstream of mining disturbances (Press, 2)	Benthos – colonization and abundances of sensitive taxa; fish – colonization, abundances, and size structure of populations	Salmonids – about 4 years after criteria attainment; Sculpin –>10 years for sites most distant from sources	21 years	Trout began moving into new habitats when Cu dropped below ~3X chronic criteria and numbers fully recovered within 3 generations after first appearance; sculpin were slower to recover especially at sites distant from source populations.	(Mebane et al. 2015)
Fish communities	Small, warmwater streams (South Carolina)	All fish removed from ~10m pool/riffle unites (pulse, 4)	Species richness, density, biomass, and mean mass of fish	<1 year	1 year	Few differences after a year in any of the 37 sites manipulated and studied	(Meffe and Sheldon 1990)
Zooplankton and benthic community	Prairie wetlands, MN	Rotenone (pulse, 3)	Species abundance and multivariate analyses	3 weeks-benthos; 9-months zooplankton	12 months	9 months recovery time for zooplankton;	(Melaas et al. 2001)
Fish population	Freshwater estuary (Lake Superior)	Pulp mill effluent (Press, 3)	Population status (modeled)	3 years to recovery post-remediation	Modeled (not applicable)	Modeling exercise linking chemically induced alterations in molecular and biochemical endpoints to adverse outcomes in whole organisms and populations	(Miller et al. 2015)
Fish populations	Experimental Lake, Ontario	Experimental acidification (press, 2)	Abundance and age structure	5 years for lake trout condition & survival rates; >10 years for species having to recolonize	20 years	Small, seasonally dry connections slowed recolonization of extirpated fathead minnow and slimy sculpin. Acid tolerant pearl dace and white sucker became more abundant than reference owing to predatory lake trout declines.	(Mills et al. 2000)
Benthic invertebrate communities	Teton River, Idaho	Channel dewatering that killed all aquatic life for over 10 km (Press, 1)	Species trajectories and composite metrics	1 to 3.5 years for abundance or trophic equilibrium respectively	4 years	Potential colonists (especially Baetis and Chironomidae) can move over long distances (km) in a short time (days). Predators slowest to appear. Simple metrics (abundance, richness, combinations) gave misleadingly short estimates of recovery.	(Minshall et al. 1983)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Fish, macroinvertebrate, diatom, and macrophyte communities	Upland lakes and streams in the UK	Acidification (press, 1)	First occurrence and abundances of acid sensitive taxa	>10 years	~30 years	Very gradual recovery is linked to gradual improvements in pH, community inertia,	(Monteith et al. 2005)
Mercury in fish tissue	Various (review)	Not applicable	Hg in difference species	>10 years	varied	Declines in Hg concentrations in fish following point source controls are slow, may reflect the incomplete controls	(Munthe et al. 2007) -
Fish community	Small streams in the UK	Acid rain (increased pH and decreased Al)	Community similarity to reference	>20 years	20 years	Residual contamination and/or to ecological inertia may have created alternative stable states preventing the reassembly of acid-sensitive faunas.	(Murphy et al. 2014)
Benthic invertebrate communities	Acidified lakes and streams in the UK	Acidification (press, 1)	Community similarity to reference	>20 years	20 yrs	Biological recovery lags chemical recovery. Infrequent episodic pollution in streams and/or ecological inertia impeding full recover	(Murphy et al. 2014)
Benthic invertebrate community	Arkansas River, CO	Mining contaminated streams (press, 2)	Invertebrate community structure	1 year for most taxa; probably complete in 2 yrs	2 yrs post intervention	Rapid recovery attributed to proximity to upstream colonizing sources and absence of physical habitat alteration (no "yellow boy" substrate coatings).	(Nelson and Roline 1996)
Aquatic plant assemblage	Mesocosms in Texas	Herbicide (fluridone) at different strengths and duration (pulse, 3)	Plant regrowth as biomass	<90 days	90 days	Most non-target plants recovered substantially by 60 days, those hit longer or with higher concentrations took longer to recover.	(Netherland et al. 1997)
Aquatic plant assemblage	Loon Lake, Washington	Herbicide (2,4,D) (pulse, 3)	Biomass and frequency in littoral surveys	>1 year	1 year	Application targeted Eurasian watermilfoil stands; effective at preventing rebound for 1 year. Few effects to nontarget plants.	(Parsons et al. 2001)
Mercury in fish tissue	Lakes in Sweden	Not applicable	Hg in different species	2 years for perch; >2 years for pike	2 years	Added Se to lakes to reduce Hg in fish. Perch approached reference 2 years after additions but little change in pike	(Paulsson and Lundbergh 1991)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
Fish populations	Forest stream	Rotenone extirpation of fish (Pulse, 2)	Fish populations	Brook trout, <1 year; sculpin, no recovery	1 year	Abundant upstream source populations	(Phinney 1975)
Benthic invertebrate communities	Forest stream in S. Norway	Abrupt elimination of acidity by liming	First occurrence; species trajectories	1 years – first occurrence of sensitive taxa; 11-years, 90% recovery	19 yrs	Baseline pH was 4.6 – 5.4 and Al 100-200 µg/L from acid rain. Liming immediately increased pH to 6 – 7 and dropped Al to 20-70 µg/L. Sensitive species steadily appeared, reaching 90% of maximum richness after 11 years.	(Raddum and Fjellheim 2003)
Benthic invertebrate communities	Animas River, CO	Major spill from mine remediation mishap	Community similarity with baseline			Communities were exposed to a severe but brief pulse of metals and turbidity. Community composition and structure do not appear to have been altered by the Gold King Mine release.	(Roberts 2016a)
Brook trout	Forest stream	Debris flow extirpating fish for a 2 km section (Pulse, 2)	Brook trout recolonization and population density	3 years	5 years	Favorable conditions for recolonization included a nearby source population, no insurmountable physical barriers, and suitable habitat in the affected area.	(Roghair and Dolloff 2005)
Fish assemblage	Warm water forest stream (EF Poplar Cr., Tennessee)	Chronic industrial waste (Press, 3)	Species richness, abundance of indicator species, multivariate community analyses	6 to 10 years	20 years	Slow, progressive increases in species richness and other measures coincident with pollution abatement efforts. No data on magnitude of pollution disturbances or abatement effectiveness presented	(Ryon 2011)
Benthic invertebrate communities	Coldwater stream (Icewater Cr, ON, Canada)	Direct insecticide (permethrin) application (Pulse, 3)	Species richness, abundance of indicator species, multivariate	6 weeks	2 years	Applied as a highly concentrated, single slug. Despite severe, local effects, recovery was rapid	(Sibley et al. 1991)
Fish and invertebrate tissue residues and invertebrate community	Coolwater mesocosm streams, Northfield, MN	Selenium dosed for 1 or 2 years (Press, 3)	Tissue residues and abundance of affected taxa	1 yr - isopods	1 or 2 years post treatment	Isopods were severely affected and recovered after 1 year. Removal of the selenium source resulted in gradual recovery of the ecosystem. Selenium residues decreased slowly in the	(Swift 2002)

Community or endpoint of concern	Setting	Disturbance & severity rating (1- severe; 4- mild) †	Recovery endpoints	Recovery time	Duration observed	Remarks	Reference
						sediments, plants, macroinvertebrates, and fishes over the two or three years studied after selenium dosing ended	
Benthic invertebrate and fish communities	Rhine River, Switzerland to the Netherlands	Pesticide spill (pulse, 2)	Taxa reoccurrence	1 year	At least 2 years; details sparse	All insect taxa reappeared within one year; most fish reappeared "within a few months;" eel age structure and diversity slower to recover. Baseline ecosystem was already severely affected by chronic chemical contamination.	(Capel et al. 1988; Van Urk et al. 1993; Giger 2009)
Benthic invertebrate communities	Forest headwater streams, Coweeta, GA	Direct insecticide (methoxychlor) repeated applications (pulse, 2)	Structural and functional measures	> 2 years	2 years post treatment	Slow recolonization (>2 years) was attributed to the more severe effects of repeated pulses and absence of upstream colonizing populations in the headwaters streams	(Wallace et al. 1986)
Stream fish	Streams ("drains") in cultivated watersheds	Habitat disturbance (drainage maintenance) (Pulse, 4)	Fish populations, community composition	10-12 months	2 years	Highly resilient cyprinid dominated streams; recovery time is for most affected stream, others had no obvious effect	(Ward-Campbell et al. 2017)
Benthic invertebrate communities	Forest streams	Mining effluent (Zn dominated) (Press, 2)	Increased richness and abundance	1 year	8 yrs, plus a year 25 revisit	Substantial (but incomplete) recovery with the first year of ceased discharges (Zn declined from ~350 to 125 µg/L). Observations of tolerant mayflies and sensitive caddisflies were opposite the patterns reported from many other studies.	(Watanabe et al. 2000)
River fish community	Animas River, CO	Major spill from mine remediation mishap	Fish populations, pre-and post spill	< 1 month	1 month	Abundance and composition of fish community was similar to pre-spill conditions; prior disturbance, low bioavailability, and rapid dissipation appear to have been mitigating factors.	(White 2016)

† A pulse disturbance allows an ecosystem to remain within its normal bounds or domain and to recover to conditions that were present prior to the disturbance. A press disturbance forces an ecosystem to a different domain or set of conditions (Bender et al. 1984; Yount and Niemi 1990b)

<u>Disturbance severity ratings</u> (after Gore and Milner 1990):

Level 1: Disturbance completely destroys communities along the entire stream length leaving no upstream or downstream sources to colonize and may result in a new stream channel.

Level 2: Disturbance completely destroys communities in a reach of stream but upstream and downstream colonization sources or hyporheic zone refuges remain, leading to succession and faunal organization. Examples: severe chemical spills, reclaimed or diverted river channels, surface mining effects, intermittent streams

Level 3: Results in reduction of species abundance and diversity from pre-disturbance levels in a section of stream but does not completely eliminate the benthos, leading to secondary succession and secondary faunal organization. Examples: incomplete kills from application of insecticides, chemical spills, wastewater treatments plant malfunctions, floods, chronic nonpoint and point pollution sources including nutrient enrichment.

Level 4: Results in reduction of species abundance and/or diversity or loss of benthos compared to predisturbance levels in discrete patches within a stream section but such that proximal patches are virtually unaffected. Leads to secondary succession and secondary faunal organization at the affected sites. Examples: sediment inputs from highway construction, logging, and introduced substrata, localized fish kills

Appendix 2: Stage-structured population model summaries

Table 1. *Hyalella azteca*, univoltine, (annual) life history scenario

Stage or age	Initial numbers	Survival to next stage	SD	Fecundity	SD
Juvenile	718	0.285	0.203	0	0
Adult	282	0.05	0.0383	3.712	0.912

Model contains females only, initial density, 1000 females/m2, distributed according to the stable age distribution. 1 time step equals 1 year. Density dependence is based on the abundances of all stages and affects fecundities only. Carrying capacity (K) is 1000 (SD 200) females/m2 and is constrained by scramble competition. Stable population growth rate (lambda, λ) is 1.049, maximum growth rate (Rmax) is 2.9. Environmental stochasticity uses a lognormal distribution, fecundity, survival, and carrying capacity are all correlated within the population. Model parameters based on Mebane (2006) but were collapsed to a two-stage juvenile and adult model.

Table 2. Fathead minnow, life history scenario

Stage or age	Initial numbers	Survival to next stage	SD	Fecundity	SD
YOY	176	0.056	0.011	0	
Year 1	2	0.633	0.15	388	109
Year 2	0	0		388	109

Model contains females only, initial abundance was 178 females, distributed according to the stable age distribution, rounded to whole numbers. 1 time step equals 1 year. Density dependence is based on the abundances of all stages and affects all vital rates. Carrying capacity (K) is 15,000 (SD 100) females and is constrained by contest competition. Stable population growth rate (lambda, λ) is 4.9, maximum growth rate (Rmax) is 5.0. Environmental stochasticity uses a lognormal distribution, fecundity, survival, and carrying capacity are all correlated within the population. Basic vital rates were estimated from Divino and Tonn (2007), overlain with carrying capacity estimates derived from Vandebos et al (2006). Both studies were from the same groups of ponds.

Table 3. Brook trout life history scenario

Stage or age	Initial numbers	Survival to next stage	SD	Fecundity	SD
YOY	4235	0.3972	0.0541	0	0
Year 1	2325	0.1439	0.0287	1.7188	0.5022
Year 2	383	0 0.066	0.0214	6.1795	2.0534
Year 3	24	0.0114	0.0314	13.1349	5.9618
Year 4	0	0		6.867	(no estimate)

Model contains females only, initial abundance was 6967 females. 1 time step equals 1 year. Density dependence is based on the abundances of all stages and affects all vital rates. Carrying capacity (K) is 8527 (SD 361) females and is constrained by contest competition. Stable population growth rate (lambda, λ) is 1.03, maximum growth rate (Rmax) is 1.88. Fecundity is expressed as spawner to spawner, replacement. Environmental stochasticity uses a lognormal distribution, fecundity, survival, and carrying capacity are all correlated within the population. Basic vital rates followed Ferson and Ginzburg's (1996) parametrization, supplemented by stochasticity estimates from McFadden (1967), which was primary source of the Ferson and Ginzburg model.

Table 4. Chinook salmon, 4 to 5-year anadromous life cycle scenario

Stage or age	Stage or age	Initial numbers (SD, range)	Survival to next stage	SD and range	Maternity (fecundity x sex ratio)	SD
Year 0, Early-life stage	Egg to census trap migrant ("smolt")	241,961 (270,610; 0 – 864,620)	0.192	0.0581 (0.11 – 0.26)	0	0
Age 1, Juvenile	Smolt to adult, (natal stream to natal stream to Marsh Cr)	59,689 (73,047; 1412 – 217,832)	0.00497	(0.0065; 0 – 1.6)	0	0
Age 2-5, Adult	Total adults from brood (includes year 4 & 5 fish)	145 (182; 0 – 518)	0	0	2375	318

Model contains both sexes (Female : male sex ratio 0.467), 1 time step equals 5 years. Density dependence is based on the abundances of ELS and juvenile stages and affects survival only. Carrying capacity (K) is 518 (SD 182) and is constrained by contest competition. Threshold for depensation (scarcity decreases finding mates) is 25 adults. Stable population growth rate (lambda, λ) is 1.31, maximum growth rate (Rmax) is 1.78. Environmental stochasticity uses a lognormal distribution, fecundity, survival, and carrying capacity are all correlated within the population. Model parameters are from Mebane and Arthaud (2010)

Table 5. Shortnose sturgeon life history scenario

Stage or age	Initial numbers	Survival to next stage	SD	Fecundity	SD
Year 2	6952	0.1	0.01		
Year 3	695	0.52	0.052		
Year 4	153	0.62	0.062		
Year 5	78	0.88	0.088		
Year 6	48	0.88	0.088		
Year 7	43	0.88	0.088		
Year 8	37	0.88	0.088		
Year 9	33	0.88	0.088		
Year 10+	241	0.88	0.088	32.92	6.584

Model contains females only, initial abundance was 400 adult females, extrapolated to each year class according to the stable age distribution. Fecundity is expressed as through survivors to age 2, which was the youngest abundances at age that could be estimated from survey data. 1 time step equals 1 year. Vital rates are not density dependent, and demographic stochasticity assumed survival rates SD of 10% of the rate estimates. Stable population growth rate (lambda, λ) is 1.0084. Basic vital rates followed Root's (2002) parametrization.