Environmental Management

Using Adaptive Processes and Adverse Outcome Pathways to Develop Meaningful, Robust, and Actionable Environmental Monitoring Programs

Tim J Arciszewski,* †† Kelly R Munkittrick, † Garry J Scrimgeour,§ Monique G Dubé,∥ Fred J Wrona,§ and Rod R Hazewinkel§

†Canada’s Oil Sands Innovation Alliance, Calgary, Alberta, Canada
††Present Address: Alberta Energy Regulator, Calgary, Alberta, Canada
§Alberta Environment and Parks, Edmonton, Alberta, Canada
∥Alberta Energy Regulator, Calgary, Alberta, Canada

ABSTRACT

The primary goals of environmental monitoring are to indicate whether unexpected changes related to development are occurring in the physical, chemical, and biological attributes of ecosystems and to inform meaningful management intervention. Although achieving these objectives is conceptually simple, varying scientific and social challenges often result in their breakdown. Conceptualizing, designing, and operating programs that better delineate monitoring, management, and risk assessment processes supported by hypothesis-driven approaches, strong inference, and adverse outcome pathways can overcome many of the challenges. Generally, a robust monitoring program is characterized by hypothesis-driven questions associated with potential adverse outcomes and feedback loops informed by data. Specifically, key and basic features are predictions of future observations (triggers) and mechanisms to respond to success or failure of those predictions (tiers). The adaptive processes accelerate or decelerate the effort to highlight and overcome ignorance while preventing the potentially unnecessary escalation of unguided monitoring and management. The deployment of the mutually reinforcing components can allow for more meaningful and actionable monitoring programs that better associate activities with consequences.

Keywords: Adaptive monitoring Environmental monitoring Environmental management Trigger Tier

INTRODUCTION

Any planned development of resources includes an expectation that socioeconomic benefits of the activity will be carefully weighed against any potential environmental costs. Typically, evaluating costs and benefits informs consent of stakeholders and is supported by understanding developed through ecological risk assessments (Munns et al. 2003; Suter et al. 2003), integrated modeling (Hamilton et al. 2015), weight-of-evidence approaches (Burton et al. 2002), watershed-scale assessments (Dubé et al. 2013), experience (Hardman-Mountford et al. 2005), and other dedicated research activities (Shotyk et al. 2017). Progressively and sequentially reducing the impacts of point and nonpoint stressors on environmental conditions in many localities shows clear evidence of the benefits of these processes.

Much of the success of reducing impacts of human activities can be attributed to knowledge gained through ecological monitoring. More specifically, much of that success can be attributed to monitoring programs that follow conventional guidance. Conventional monitoring programs are characterized by the use of predetermined tools, study designs, and analyses. When all types and magnitude of possible impacts are known, when minor exposures cause obvious, severe, and unwanted change, conventional programs dedicated to detecting change can be readily designed and implemented (Dubé et al. 2013). Following such a model continues to be useful in many instances, including programs for determining the safety of drinking water (WHO 2006).

While conventional monitoring programs can provide a useful model, they can also fail in important instances (Lindenmayer and Likens 2010; Baum et al. 2016). Often, failure or challenges associated with conventional programs...
applied to ecological monitoring originate from its sometimes-unrecognized multidisciplinary nature (Bammer 2003; Vaughan et al. 2009). Broadly, the challenges can be classified as scientific and social, but other more specific attributes, such as institutional aspects and their interactions, affect ecological monitoring. Common scientific challenges in monitoring include what to measure and how to interpret any differences observed but can also include poor articulation of questions and ineffective design choices (Field et al. 2007; Lindenmayer and Likens 2010). Some social factors that affect ecological monitoring include differences among groups in philosophical views of the environment (Callcott 1999; Huntington 2000) and which aspects are considered socially relevant (Díaz et al. 2011), scientific bias (Burgman 2005), and hubris (Hall et al. 2007). More specifically, institutional factors include both scientific and social aspects and vary from parochial or poorly coordinated monitoring projects, disparate regulatory requirements among programs (Dubé et al. 2013), inertia and inflexibility (Huntington 2000), or degeneration of the program into a data-collection exercise. Although other factors are clearly important, such as the lack of baseline information (Landres et al. 1999), competing demands on limited resources (Postel and Richter 2012), and the time scales considered, the combination of these factors, especially when they are not recognized, can sometimes reduce the efficacy of the monitoring in informing management (Lindenmayer and Likens 2010).

The desire to improve environmental monitoring has been greatly aided by recent discussions of adaptive monitoring (Lindenmayer and Likens 2010), which emphasized many attributes included in existing programs like the environmental effects monitoring (EEM) suite in Canada (Walker et al. 2003). Although EEM made several important advances and heavily influences the current discussion, further improvements can be drawn from additional work to expand and apply a useful monitoring program in other situations. Recent advances in quantitative risk assessments (Bayliss et al. 2012; Landis et al. 2016), addressing cumulative effects at a watershed scale (Dubé et al. 2013), and integrated risk assessment (Munns et al. 2003) are clear sources of advancement. As well, improving the consideration of other philosophical aspects, such as adverse outcome pathways (AOPs) (Ankley et al. 2010), strong inference (Platt 1964), and scientific surprises (Kuhn 1962), can be adopted create a robust monitoring system.

While the general concept of adaptive monitoring is clearly beneficial, more specific suggestions to rectify the weaknesses, regardless of apparent logic, utility, and simplicity, may be either difficult to implement or may trade one problem for another. A poignant example is the need for long-term data. This need has been discussed for many years (Likens 1989) and the continued absence of long-term data is lamented in contemporary work (Lindenmayer et al. 2015). In many cases, progress on the widespread adoption of long-term collections has been prevented by operational constraints, including financial limitations, small-scale scoping, methodological and design inconsistencies among and within programs, and diverse data storage practices. A further challenge is a shifting baseline in which data may have an unknown expiry date and, depending on the objectives of a program, may or may not lose relevance (Hardman-Mountford et al. 2005; Duarte et al. 2009). Long-term data sets may also have unintended consequences, including the anchoring of programs and loss or deterioration of data over time. Within a long-term program, the original purpose can also become obscured. It is clear that even something as seemingly simple as long-term data requires careful consideration, mindfulness, and intent in design, data collection, analysis, and interpretation.

Similar to earlier work (Lindenmayer and Likens 2010), information presented here is intended to inform decision making where long-term monitoring has occurred, is likely to occur, or even where multiple proximate programs can be coordinated or consolidated (even peripherally) into a cohesive regional analysis useful for multiple decision-making purposes.

The intended audience of the present paper includes designers and operators of monitoring programs, but any person, group, or organization that interacts with these professionals will likely benefit from this discussion. The discussion will initially focus on the necessity of separating monitoring and management objectives, defining and focusing monitoring questions, and defining optimal monitoring strategies and approaches to address those questions. While many of these concepts are well described elsewhere, this work will focus on the relationship among the various aspects and steps, and the role of adaptation to provide practical guidance in designing and operating a robust system for ecological monitoring.

UTILITY OF ADAPTIVE MONITORING

In many cases, despite the best efforts and talents of scientists, knowledge of environmental effects of various developments is often incomplete, and this ignorance may not be well managed. Furthermore, relevant signals of change can be difficult to detect and link to a cause. A useful monitoring system can be developed to reaffirm development decisions, to address cumulative effects, to illuminate ignorance and improve environmental performance of existing developments, and to inform future activities. This newer philosophical approach, known as “adaptive monitoring,” is a stepwise process guided by ignorance and based on scientific analysis of hypotheses bearing a strong and intentional resemblance to research. The distinguishing difference between adaptive monitoring and research is often only the objectives. Whereas research can often, but not always, be an end in itself or inform ecological management, monitoring must contribute to fulfilling a management objective. The Canadian EEM program was designed to be adaptive, and its origins and operation have influenced the concepts presented here. Briefly exploring the development and objectives of EEM provides a relevant foundation for advancing adaptive monitoring.

Environmental effects monitoring began after an update to the Canadian Pulp and Paper Effluent Regulations in 1992 (Government of Canada 2017) requiring pulp and paper mills to meet discharge guidelines for acute toxicity, total suspended solids, biochemical O demand, and dioxin.
(Walker et al. 2003). Although improving the quality of effluents was an important step, there was still considerable uncertainty of the efficacy of these regulatory requirements in achieving environmental protection, given the diversity of mills, effluents, and the receiving environments and evidence of significant and persistent environmental impacts near mills achieving the discharge guidelines (Munkittrick et al. 1994). Subsequently, the 1992 effluent regulations were initiated using not only best available scientific and engineering knowledge but also included a mechanism, EEM, to test whether the discharge regulations were sufficiently protective under actual, rather than modeled, field scenarios.

The EEM program included unique features, among them nationally applicable guidance for cyclical, industry-funded monitoring across tiers. The efforts during monitoring adjust in intensity and focus when differences larger than a critical effect size are observed, rather than relying solely on statistical significance. Environmental effects monitoring also required that suspected impacts be confirmed before management action could be initiated. The program has also evolved over time in response to the challenges uncovered during the monitoring (Munkittrick et al. 2010). The EEM process has proven beneficial in many specific and general cases (Lowell et al. 2005; Hewitt et al. 2008; Martel et al. 2011). This program has also been expanded to metal mines (Ribey et al. 2002) and has been considered for other aquatic dischargers, such as sewage facilities (Kilgour et al. 2005).

Similar concepts embedded in EEM have also emerged in guidance for other provincial (Ontario Ministry of the Environment 2014; British Columbia Ministry of Energy and Mines and British Columbia Ministry of Environment 2016) and federal (Canadian Nuclear Safety Commission 2013) programs in Canada. Internationally, the concepts are also found in the Water Framework Directive of the European Union (European Commission 2015) and some programs in the United States (Suter et al. 2003; Ohio Environmental Protection Agency 2011). Although it is clear that the concepts of improvement included in adaptive monitoring are emerging in many programs and have been successfully deployed, the suffusion is also clearly not complete. In addition, the language among the guidance differs slightly, as can the recommended frameworks, if one is provided. Although some of these details may be known to many, there is a valuable opportunity to provide details to a wider audience and those less familiar with beneficial concepts of adaptive monitoring and how it can be helpful in a variety of situations.

Accordingly, adaptive monitoring is ideally simple, practical, and progressive. It is composed of hypothesis-driven questions linked to conceptual effect pathways. The monitoring evolves through feedback loops that accelerate or decelerate activity as necessary, and includes other components to improve the program. An adaptive monitoring program converges toward an ideal design over time, which may eventually include only the most basic adaptive components, such as tiers and triggers. For instance, further adaptations may be necessary as understanding of the study system deepens, including evaluating the efficacy of indicators or redundancy of study sites.

Developing an adaptive monitoring program can be split into its component parts. First, designing an effective monitoring program can be informed by quantitative risk assessment and cumulative effects assessment, understanding the differences between adaptive management and adaptive monitoring, and defining the monitoring questions and philosophy. Once a clear path toward effective monitoring and management has been established, the core aspects of adaptive monitoring, AOPs, triggers, and tiers can be defined and integrated.

**IDEAL INITIAL STEPS FOR IMPLEMENTING AN ADAPTIVE MONITORING PROGRAM**

**Quantitative risk assessment and cumulative effects assessment**

Various jurisdictions may have differing or overlapping requirements for development applications. Environmental impact assessment (EIA) typically begins by identifying where, when, how, and why impacts are likely to occur. This process has, however, been heavily criticized in the past (Duinker and Greig 2006). Many suggestions for improving this process have been made recently, including Bayesian belief networks (Nyberg et al. 2006), comanagement (Plummer et al. 2012), or watershed-scale assessments (Dubé et al. 2013). Other clear examples of advances are also available (Munns et al. 2003; Suter et al. 2003; Newman et al. 2007; Cormier and Suter 2008; Bayliss et al. 2012; Landis et al. 2016). Despite these improvements, a gap in some of the existing preddevelopment approaches, and the one addressed by this discussion, is the structure and linkages to post-EIA activities. Called “adaptive monitoring” here, several important other aspects are also required to attain a truly effective monitoring program. The second, after or during a preddevelopment phase, is delineating adaptive management and adaptive monitoring and the objectives of each.

**Management and monitoring**

Monitoring and management are distinct, but linked processes. “Management,” as used here, concerns the act of intervention, or the decision process used to determine whether intervention is necessary after convincing evidence of an unwanted effect is obtained. “Monitoring,” in contrast, is the process that obtains that evidence and feeds information into a management system regarding the efficacy of existing control mechanisms. Modern conceptions of both often include adaptive components. These activities have been much improved with several recent discussions of approaches for adaptive monitoring (Vugteveen et al. 2015) or related concepts, such as adaptive governance (Chaffin et al. 2014). Many, however, tend to blur the roles and activities of management and monitoring (Westgate et al. 2013). Furthermore, in many instances, the application of adaptive management has drifted from the classic definition and often refers to changing processes in response to new information. In some cases, the term “adaptive monitoring”
may be more appropriate to describe decisions and activities, rather than “adaptive management” (Westgate et al. 2013).

Despite some conflation of the terms, adaptive management and adaptive monitoring can be easily and clearly distinguished. Much of the early literature on adaptive management addresses resource development (Walters 1986). The intent of monitoring in such a system is to manage the direct effects of human activities and their sustainability, including fisheries and timber harvests (Walters 1986). There is, however, room within the monitoring program to seek more definitive information that can withstand deeper scrutiny, provide compelling evidence, and ultimately lead to better management decisions. Modifying monitoring with the intent of optimizing the utility of the information provided by the program could be easily described as managing adaptively, including the lessons learned and actions taken during an initial optimization period (Hall et al. 2009). These latter aspects are more closely aligned with the initial steps or an early phase in the adaptive monitoring scheme described here, rather than the modeling normally included in classic adaptive management.

Although monitoring and management activities are distinct, each have decision points (Cook et al. 2016). Here these decision points are called “triggers.” Monitoring triggers influence future monitoring choices, whereas management triggers influence management choices. Monitoring triggers mark a less severe outcome than ecological thresholds (Groffman et al. 2006) or management triggers. The consequence of exceeding a monitoring trigger is to adapt the program to provide more specific information allowing better input to a potential management decision. Management triggers, in contrast, can be represented by a variety of decision guidelines, including water quality criteria, ecological thresholds, probable effects levels (PEls), or alarm levels (Arciszewski and Munkittrick 2015), but regardless of their origin, reflect an unwanted state; the consequence of exceeding a management trigger is often to intervene.

Despite the relevance of differences between management and monitoring, the linkages can also be further explored. Management ideally identifies the aspects of the environment that can or cannot be altered by the activity, which provides scope to the monitoring. This important linkage emphasizes the difference of each activity as part of a cohesive whole. By definition and intention, adaptive monitoring is one part of an effective management process.

Management objectives may be delineated as economic or social goals, whereas monitoring objectives indicate when those goals are being threatened. Aligning monitoring and management objectives has been articulated in discussions of AOPs (Ankley et al. 2010). Utility of AOPs extends beyond this alignment and is further described below (Unifying management objectives and monitoring philosophies with adverse outcome pathways) as a core aspect of an adaptive monitoring program.

Defining the monitoring questions

If one component could be identified as having the most influence over the success of a monitoring program, it is the clarity and relevance of the questions the program is designed to answer. Defining a carefully worded question articulates the monitoring objectives in an understandable and executable manner. The absence of a clear question and agreement on that question, the use of confusing or misaligned vocabulary, or a misunderstanding of the challenges of monitoring can easily lead to fracture among participants and the failure of the program to accomplish its goals. Failure can also be associated with the tendency to ask complicated or vague questions. Furthermore, specific questions may be hidden within the primary goals that define the overall program, may only emerge later, and may not be answerable with a single approach.

It is important to understand that different components of the program will be required to answer different types of questions. For example, questions about the bioavailability of a contaminant, potential human health effects, or ecological consequences of development require different designs, philosophies, and approaches. Thoughtfully integrating the various approaches and questions is important for acceptance of adaptive monitoring by groups composed of multiple stakeholders (Munkittrick and Sandström 2003). A relevant and common point of friction in the design of a monitoring program is a difference in philosophical approach to monitoring the impacts of human activities on the environment.

Influence of monitoring philosophies on questions and tools

All monitoring programs are derived from a desire, either explicit or implicit, to minimize or prevent change attributable to a given activity and to ensure that ecological change stays within acceptable limits, including the basal rates of change inherent and unique to each ecosystem. Among some programs, indicators are often selected to maximize suspected exposure pathways, such as benthic fish with small home ranges to show the maximum possible response to a stressor (Munkittrick et al. 2002), but other reasons have also been discussed, depending on the goals of the program (Cairns et al. 1993) and philosophical origin of the monitoring design (Dubé and Munkittrick 2001). To achieve the goals of monitoring, there are 3 general philosophical approaches to designing a program: effects-based, stressor-based, and value-based (see Dubé and Munkittrick 2001). Indicators based on these varying philosophies are not mutually exclusive but have been described elsewhere as pressure (stressors), state (effects), and benefits (values) indicators (Vugteveen et al. 2015).

Stressor-based designs equate control of stressors (e.g., concentration of Hg or suspended sediment) with control of impact (e.g., levels of human exposure, recruitment in fish species). These designs primarily use direct measurements of toxicants, nonchemical, or physical stressors (Dubé and Munkittrick 2001). A relevant difference is often evaluated
with permissible limits (Häder 2013) or another guideline. Although measuring and accounting for known toxic constituents have clear value, the perceived regulatory and logistical simplicity and ease of interpretability may explain the primary focus of some programs on stressors. Permissible limits are typically derived from exposure studies of a broad range of test organisms and are often broadly applicable. Such limits may, however, have limited relevance in a particular ecological setting if derivation of the limit does not account for relevant characteristics, including sequestration and bioavailability (Luoma et al. 1999) or aspects of a stressor that may vary over time (Hewitt et al. 2008). The risks of a stressor-based approach are missed effects originating from complex and possibly unique interactions among stressors that may be difficult to disentangle and understand clearly. Despite these weaknesses, programs based on this philosophy can provide meaningful information where risk can be clearly delineated and the role of and need for vigilance is not underestimated (Baum et al. 2016).

Informs an ecosystem using integrative measurements to focus attention on significant areas of change, despite little understanding of a specific cause or causes. By focusing on effects measurements, effects-based approaches detect relevant change that occurs in response to varying conditions potentially operating through unanticipated, unknown, or unmeasured pathways. The main challenges of the effects-based approach are defining relevant changes, defining how they will be detected, and recognizing that not all changes in lower levels of organization lead to changes in higher levels of organization (Munkittrick and McCarty 1995). This approach is further complicated because organisms are not passive receptors and causal relationships can be difficult to disentangle.

Informing the design of a comprehensive monitoring program and providing objectives for stressor- and effects-based programs can be achieved with values-based criteria. Usually these designs focus directly on the use of resources or a point of overlap between ecosystem process and human activities, including population-, community-, and ecosystem-level responses (Ankley et al. 2010). Values-based approaches often use the same measurements as a stressor-based or effects-based approach but focus on specific values or uses. Unfortunately, these valued components may not always be the most sensitive indicators of change. There are, however, clear benefits of incorporating valued ecosystem components (VECs). They provide access points for community members to other information that may not seem applicable to a layperson, acceptance of that other information, ownership and confidence in the monitoring, and a priori management objectives. Integration of VECs may also force important discussions, such as describing change that is acceptable versus change that is sustainable. A risk in exclusively focusing on predefined VECs as monitoring objectives is the inability to detect the occurrence of subtle change early before a VEC is affected beyond an acceptable level. Furthermore, even spatially proximate programs may use incompatible VECs (Ball et al. 2013).

DEFINING CORE ASPECTS OF ADAPTIVE MONITORING

Unifying management objectives and monitoring philosophies with adverse outcome pathways

A key task for a successful adaptive monitoring program is understanding strengths and weaknesses of stressor-, effects-, and values-based monitoring philosophies and how each relates to a management objective. In the adaptive monitoring system described here, alignment of measurements of stressors, effects, and VECs along an AOP may emphasize the utility and linkages between various types of tools and enhance vigilance. Often, these pathways terminate at an endpoint such as lethality or some other unwanted effect of human activity; elsewhere these are known as “assessment endpoints” (Suter 1990). An adaptive monitoring program is most effective when a measurement or monitoring endpoint occurs earlier than an assessment endpoint in an effect pathway. Although the linkages between levels of biological organization may be clearly defined, changes at lower levels do not always elicit a measurable response at higher levels (Munkittrick and McCarty 1995) and the relationships between stresses and expected responses may not be static (Wu et al. 2005). Regardless of the specific challenges, integrating factors from each philosophy into an AOP is also intended to make the monitoring inclusive, to spread risk of failure among tools, and to address concerns of all stakeholders.

Adverse outcome pathways have several advantages in monitoring change. The utility of defined AOPs begins with establishing linkages between monitoring and management endpoints. As discussed in earlier influential work, AOPs are also useful for predicting change, both earlier and later in an effect pathway (Ankley et al. 2010). These predictions can be effective in both understanding the cause of a change and providing information to prevent the progression of effects from one level of biological organization to another. Predictions also provide an explicit mechanism to define when change is progressing and influencing adjacent levels of organization, further providing some assurance that the effect pathway is understood. Aligning monitoring indicators along an AOP is a potentially rich source of information for management.

In some cases, however, understanding of an AOP is likely incomplete (Ankley et al. 2010). In these instances, the AOP can be constructed through hypothesis-driven processes proceeding backwards from a management objective (e.g., no change in fish communities) toward finer levels of biological organization (e.g., physiological indicators or markers of exposure). Although AOPs are likely to be useful in many instances, as a relevant factor begins to influence a biological endpoint, the true complexity of biological responses may emerge; predicting the real impact and its
extent may become impossible. In these cases, probabilistic events may be the optimal targets for knowing if intervention is or is not required.

Aligning measurements along AOPs may also provide a possible solution to the actual occurrence or perception of missed effects, otherwise known as “false negatives.” Adverse outcome pathways, as used here, can enhance confidence in monitoring results by providing a framework to track and encourage the detection of changes, ideally and initially in stressors and a well-chosen and sensitive indicator early in an effect pathway. Trackable change is used in concert with the other core features of an adaptive monitoring program, tiers and triggers, to emphasize detection of change to control false negatives by emphasizing positives.

Triggers

A key challenge in ecological monitoring is separating relevant and undesirable anthropogenic effects from natural environmental variability. Overcoming this challenge in an adaptive monitoring program is aided by predicting future data from past observations and forms the basis of the adaptive monitoring framework advocated here. Based on the concepts of control charting (Anderson and Thompson 2004; Burgman 2005; Montgomery 2009), relative boundaries of expected data (also called normal ranges, lines, thresholds, benchmarks, levels, or critical effect sizes [CESs]) are often defined by marginal percentiles that encapsulate a rate of occurrence expected for future observations, if existing sources of variability maintain their relative importance and influence over time.

The complex task of detecting relevant effects can also be aided by clarifying vocabulary. Unexpected observations beyond a threshold are defined as “changes.” Also called differences, change is distinct from an “effect” or an “impact,” which have stronger associations with being unwanted, malignant, and likely requiring intervention. Effects and impacts, including cumulative and subtle effects are also interpreted as important, but they are often difficult to clearly identify. In most cases, however, effects must first be different than expected; in an adaptive monitoring program, effects must first occur as unpredicted observations, or as changes. The general search for change, facilitated by defining thresholds, has several advantages. Identifying change does not necessarily require either explicit knowledge of all factors causing a change, or the actual importance of any difference encountered. More detailed information, such as the form of a change (e.g., step-changes, monotonic trends, or unexpected patterns within or outside the range of expected variability) can also be defined, but this explicit knowledge is also not necessary a priori. In contrast, searching only for a particular pattern or form of change is likely to weaken the monitoring.

Exceeding a threshold often attracts attention, but as initiators of some further monitoring action, including refinement of questions, thresholds are operationalized as triggers. Elsewhere published CESs can serve as preliminary triggers, but their development and refinement is an important and explicit component of an evolving monitoring program (Arciszewski and Munkittrick 2015). Where refinement is desired, a parametric 95% prediction range is commonly advocated as a trigger and placeholder for change that may be relevant (Munkittrick et al. 2009; Arciszewski and Munkittrick 2015). Virtually any coverage probability can be used but requires acknowledging the accompanying (and expected) error rates. Vigilance is also required, because in some cases, expected and actual error rates can deviate wildly, especially when little data are available and the focus of a study is on rare events.

There are a variety of statistical approaches for developing triggers, including tolerance intervals of a confidence limit (Smith 2002; Kilgour et al. 2017) or bootstrapping (Anderson and Thompson 2004). An initial configuration of triggers may include multiple types used to check for agreement. The technique used to define triggers depends heavily on the characteristics of a data set. For example, tolerance intervals rely on normality, whereas nonparametric bootstrapping does not. In contrast, nonparametric bootstrapping is limited to the empirical distribution, whereas parametric methods can approximate missing data using an assumed distribution. Similar to sampling in a given season, triggers can be further refined to account for the influence of environmental covariates (McLaughlin and Flinders 2016). Triggers can be developed for specific study sites, localities, and regions, but an optimal initial configuration focuses on change at individual study sites over time (Arciszewski and Munkittrick 2015).

Families of triggers. Although described as theoretically simple, in practice, accurately predicting future observations will likely be complicated. Foremost among these complicating factors is that observations at a study site of interest are the result of several sources of variability, most of them unknown. Commonly, phases such as “background variability,” “historical variability,” and “natural variability” are used to describe existing or expected conditions. Each, however, makes various assumptions regarding the status of a given study site. “Natural variability” is often used as an umbrella term, but in its strictest definition includes only natural and regular rates and direction of change and infrequent events, including influences operating at a time scale of seconds to those changing over millennia. Most importantly these changes would likely occur in the absence of human activity. Often lumped with “natural variability” and the source of much discomfort with the term and difficulty of identifying relevant changes are any preexisting anthropogenic influences (Landres et al. 1999). In contrast, the terms “historical variability” and “background variability” typically do not assume the absence of the influence of humans but often require the absence of a particular activity. Natural variability is the more powerful tool but is also more difficult to define (Hardman-Mountford et al. 2005).

Using statistical approaches, 2 general families of triggers can be developed. Where abundant data from an undisturbed baseline period is available, a fixed (or static) trigger can be defined. These fixed triggers will likely be rarely
available but would have clear utility in many instances, including paleolimnological studies (Kurek et al. 2013). Fixed triggers are most useful where anthropogenic activity does not mimic a natural process and instead has some unique attribute, or signature. Where a signature can be developed, fixed triggers can be powerful markers of specific anthropogenic activity and potential influence.

Where no reliable baseline data are available, which is the more common state, drift from that unknown baseline cannot be easily addressed. These instances require a different question and a different approach for detecting change, such as hindcasting an expected estimate of historical conditions (Kilgour and Stanfield 2006). Alternatively, instead of focusing on estimating a predevelopment scenario, an adaptive trigger can be used. Using an adaptive trigger first requires defining an interim “baseline” and periodically updating the range of expected values with new data, often annually, only after testing for the occurrence of change. If no change is likely, the new data can be collapsed into the existing interim baseline. In contrast, if changes are detected with an adaptive trigger, the interim baseline is fixed and confirmation is initiated. In addition, sampling at other study sites, if it is not already being done, will be necessary to evaluate context (Arciszewski and Munkittrick 2015). Importantly, however, if no changes are detected, the conditions are presumably stable and sustainable but may not be acceptable, especially if existing current stable conditions represent a marked departure from probable historical conditions. Although an anthropogenic source of variation may be present, adaptive triggers assume the indicator is capable of responding to further anthropogenic change in the environment.

An adaptive trigger is advantageous where the status of a study site of interest (in or out of baseline) is not specifically known and can be applied to understand if current conditions are stable. When conditions are stable, the integration of new data will have diminishing influence on the trigger, which will likely stabilize after between 8 and 12 observations or collections but may take as few as 3 or as many as 20 (Parker and Berman 2003; Arciszewski and Munkittrick 2015). Consequently, when background conditions are stable, or the analysis adequately matches the scale of a stressor and its variability (local vs regional) with the scale of the observation, the data from early years of a program will provide the largest proportion of information (Parker and Berman 2003).

In contrast, when background conditions influencing the indicator are not stable, or when local assessments are affected by unknown regional or intermittent stressors, the triggers reflect this instability. More specifically, changes in some measure of variability over time, such as standard deviation (SD) and its predicted patterns, including normal distribution of errors, can signify change. The effect of changing variability can be examined using simulations (Figure 1). A stable estimate of variability (SD = 1; Figure 1A) is generally flat over “time,” but some slight deviations, related to sampling error, are expected. In the examples of stable SD presented here (Figures 1A and 1C), little variation occurs after 10 simulated years. Where variability is increasing (Figure 1B) or decreasing (Figure 1D) over time, measures of spread will not conform to the expected patterns. Although the changes in variability in the examples are large to illustrate a point, similar types of analyses may be useful in operational programs to test the performance of indicators. When the variability of a trigger has not stabilized after 20 collections, a focused study examining the possible origins is an appropriate response. Noticing a lack of stability is, however, likely to occur earlier than after 20 collections. Comparing patterns of variability among study sites is also useful for detecting unique attributes at an exposed site (Magnuson et al. 1990).

When is a change an effect? A specific topic within the general discussion of triggers is their capacity to detect real

---

**Figure 1.** Twenty simulated examples per pane of stable (A, C) and unstable (B, D) running SD of grand means over simulated years: Estimated SD of a grand mean with an SD = 1 for all collections (A); effect of progressive drift of SD every 5 y between 1, 2, 3, and 4 on running SD (B); estimated (stable) running SD expected for a grand mean estimated for a population with SD = 4 (C); progressive decline in SD from 4 to 1 every 5 y (D).
effects. In the absence of observing a clear problem, the adaptive monitoring system described here, applied with both fixed and adaptive triggers, relies on serial probabilities of rare observations to infer changes and possible effects. This decision framework creates challenges regarding the number of rare observations in a sequence that requires an escalating response. Although a fixed number of exceedances may be operationally simple regardless of the measurement, the number of unusual observations may need to be scaled to the relevance of a given indicator. For instance, many exceedances of water chemistry thresholds may be necessary before action is taken to account for known and/or expected rates of variability (Barnett and O'Hagan 1997), especially in the absence of concurrent (and associated) changes in a biological indicator or valued ecosystem component.

Much of the current focus of triggers has been on marginal percentiles (Arciszewski and Munkittrick 2015; Barrett et al. 2015). Attention has also focused on developing rules similar to those used in industrial control charting for application to ecological data (Anderson and Thompson 2004; Burgman 2005). Adopting such an interpretive structure means analytical tools are not restricted to an a priori definition of a change. For example, trends can be used as a trigger, but singular focus on this or any other analytical technique at the expense of all others is not a recommended practice (Arciszewski and Munkittrick 2015). Instead, control charting is a tool to identify several types of unexpected patterns within the range of described variation (Nelson 1984). In such a system, multiple triggers would be associated with each indicator. Expected rates of exceedance of percentiles from an earlier period would be used to predict the frequency of values above or below various percentiles in the future. For instance, the long-run rate of exceedances of a 75th percentile is 25% of observations, if conditions are stable and the baseline conditions were well characterized. The advantage of triggers based on less extreme percentiles is that they can be more accurately defined, but the disadvantage is that more sequential observations beyond a central percentile, such as the median, are needed to ascribe a truly rare pattern.

To detect unusual patterns within a range of expected variability, program operators must numerically define the value of “rare.” Multiple exceedances of a marginal percentile, such as 2.5\text{th}, have been useful in other work (Arciszewski and Munkittrick 2015), but the serial probability of 0.025\textsuperscript{2} (0.00156\%) can be applied to the chance occurrence of other patterns, including values above a median or within an interquartile range (Table 1). For instance, 8 means above the 75\textsuperscript{th} percentile in 8 sequential years is at least as statistically rare as 3 sequential means above the 97.5 percentile. This same logic also applies to percentiles more extreme than the bounds of a central 95\% interval and may be useful for defining PELs where no other information is available.

Functionally, triggers as proposed here are simple but predictive models of the ranges of expected future observations. Testing those predictive models against observed data is a primary mechanism to develop scientific and hypothesis-driven monitoring. Despite this scientific approach, the monitoring triggers are not infallible and can be challenging to implement (Kilgour and Somers 2017), especially in the early stages of monitoring program. Design of triggers also includes built-in false positive error rates. When a trigger is exceeded, other mechanisms are needed to separate true and false positives. More broadly, further mechanisms are needed to determine which subtle changes are relevant, and which are not. Some suggestions have already been presented, including multiple triggers and the use of AOPs.

Triggers are meant to provide further context, albeit imperfect, for evaluating change by focusing attention from the broad to the specific, thereby providing a linkage between an activity and its management in the midst of uncertainty. Further information is sought to ensure that the change originates from a novel driver in the study system, rather than a poorly defined trigger or normal variation, and to separate (likely) true from (likely) false positives using tiered analyses.

### Tiering

Tiering is a formalized segmentation of a complex environmental management question (is there a problem and must it be fixed?) into measured and manageable pieces (Hodson et al. 1996). Similar to its application in other aspects of environmental management (Suter et al. 2003), each tier addresses simple and answerable questions, rather than attempting to accomplish multiple goals, such as detecting change and attributing cause with a single comprehensive and complex study design. The main functions of tiering are

| Percentile | Consecutive exceedances |
|------------|-------------------------|
| 1 or 99    | 3                       |
| 2.5 or 97.5\textsuperscript{a} | 3 |
| 5 or 95    | 4                       |
| 10 or 90   | 5                       |
| 15 or 85   | 6                       |
| 20 or 80   | 7                       |
| 25 or 75   | 8                       |
| 30 or 70   | 10                      |
| 35 or 65   | 11                      |
| 40 or 60   | 13                      |
| 45 or 55   | 14                      |
| 50         | 16                      |

\*Conventional range.
to moderate and temper the response of a monitoring program to observed and potentially small and subtle exceedances and to separate real change from poor understanding of variability (Hodson et al. 1996).

Tiering has several advantages. Tiering discourages reliance on collecting data first and asking questions later but can also provide opportunities the activity of “doing science backwards” (Lindenmayer and Likens 2010) can still provide useful, but limited, information. Similar to triggers, tiering also provides transparency to regulators, industry, and stakeholders with an understanding of how, when, why, and what changes in the design of a monitoring program would occur. Tiering also necessarily represents a de facto level of scientific concern. Tiers can be distinguished by frequency of measurements, types of measurements, and effort. The effectiveness of a tiered approach is supported by 2 assumptions. First, subtle change can be difficult to detect, but will leave clues. Second, tiering necessarily requires a program that will continue for multiple years.

Movement between tiers is guided by the exceedance of a trigger in one or several indicators (Figure 2). Large differences or unexpected changes (Manley et al. 2004) that exceed some PEL may also be used to skip tiers in adaptive monitoring and proceed directly to some detailed and difficult investigation. The recommended tiers for an adaptive monitoring program adapted from EEM are these:

- baseline and initial effects monitoring,
- surveillance,
- confirmation,
- focused study,
- investigation of cause and investigation of solutions (IOC/IOS), and
- minimal monitoring.

**Baseline and initial effects monitoring.** EIA or predevelopment phases of monitoring will ideally collect information useful for evaluating the occurrence of change after an activity begins. At worst, this information will not be available and baseline and initial effects monitoring is required to initiate an adaptive program. During the baseline phase, collecting baseline data at an exposure site to predict observations during surveillance is a main goal. The baseline collections could conclude after as little as a single year, but to develop a trigger, the baseline collections would likely require multiple years (see Families of triggers section). Deciding how many years to sample during the baseline can be guided by the stabilization of the described variability. After suitable baseline data are available for study sites of interest, surveillance monitoring can begin.

Prior to the initiation of surveillance monitoring, comparisons to determine if change is occurring can, however, also be done against local reference sites during the baseline period. Although not ideal, comparisons to reference sites are adequate for interim analyses. These comparisons may also be used to progress through tiers but may rely on an onerous burden of proof, such as a probable effect or alarm

---

**Figure 2.** Core tiers and generic activities associated with a general and simplified adaptive monitoring program assuming sufficient baseline data are available prior to initiating surveillance; many details found in text omitted for clarity. PEL = probable effects level.
level (Arciszewski and Munkittrick 2015). This initial effects monitoring is commonly done and can provide information to initiate an adaptive program.

**Surveillance.** Regular monitoring at a study site of interest over time is considered “surveillance” monitoring (Figure 2). Surveillance monitoring often includes the basic monitoring requirements determined through the predevelopment phases, such as risk assessment, and measurement of indicators most likely to show responses to novel or changing conditions in the study system. Within surveillance monitoring, multiple indicators may also be mutually reinforcing, especially those aligned along an AOP.

Other mechanisms to evaluate change are also used in a surveillance phase. Evidence of an unusual change at a study site of interest can also be evaluated by comparing patterns from, ideally, multiple sites from a broad spatial scale. When examining point-source contamination that produces strict and unidirectional exposure patterns, such as effluent outfall discharging to a river, spatial stratifying sites into exposed and local and regional references can aid in determining the rareness of an observed change at the exposed site (Arciszewski and Munkittrick 2015). The occurrence of broad regional change at multiple sites is likely related to a broad regional factor, such as climate, or a stressor influenced by a diffusive process, such as atmospheric dispersal. This type of simple analysis will likely be useful in the future as climate regimes shift and affect indicators in various ways.

Large-scale designs may include more surveillance sites than can be reliably sampled in a given year. Financial constraints or other limiting factors, such as a short sampling window (Arciszewski and Munkittrick 2015) will affect the operation a large-scale program. To compensate for these constraints, study sites may be sampled on a rotating basis. However, a problem with rotating sampling designs is that typical analytical approaches, such as analysis of variance (ANOVA) that do not include mixed effects may have difficulty addressing missing data. Even where study designs are planned perfectly, they are not likely to be executed perfectly; some data from study sites, samples, or years are likely to be missed. Where CESs can be defined, mixed-effects ANOVAs should be adequate. Where CESs have not been defined, default options can be used or effort can be directed to define CES.

**Confirmation.** A study designed to confirm a rare observation is ultimately designed to separate likely from unlikely effects and reduce overintervention (Figure 2). Confirmation can, however, also be used to reaffirm the absence of change where none is expected, or to reduce the risk of underintervention.

The basis and utility of confirmation originates from the increasing improbability of sequentially observing rare data but also to compensate for the expected error rates included in triggers. Necessity for confirmation is, however, mediated by the purpose of a given trigger. A change beyond a well-defined management trigger, such as a PEL, especially one determined from reliable baseline data or robust toxicity thresholds, does not necessarily need to be confirmed. Similar to surveillance, when large and unwanted changes occur, often defined by a PEL, the adaptive system may proceed directly to a causal investigation. Where a subtle difference greater than a trigger value but smaller than a PEL has been seen for a second time, the program proceeds to a focused study. Confirmation may also include additional sampling at reference sites and may include additional sampling or analyses in other indicators in an AOP.

**Focused study.** A focused study is a mechanism to further test the occurrence of change and an increasing likelihood of an effect (Figure 2). Focused studies in monitoring typically fall under 2 types, based on the measurement of indicators. The first type uses the same indicators measured throughout the surveillance and confirmation monitoring. In this simple configuration, a focused study provides a second confirmation and may examine other aspects, including the spatial extent of a change (Hodson et al. 1996). Study designs could include an expansion of sampling at local and regional reference sites to evaluate the spatial uniqueness of an observed change or a simple increase in sampling effort at the study site of interest. The second general type of focused study includes measurements of the predicted progression of an effect along an AOP, or additional changes predicted to occur if the suspected effect is real. Regardless of the type of focused monitoring study, the work uses more specific tools and onerous thresholds to better understand the reality and severity of a suspected effect.

Where disagreements regarding any aspect of the monitoring program are encountered, a third type of research-oriented focused study can be invoked to reconcile any disputes. In many respects, including these special and short-term studies, focused studies are a main avenue for integrating flexibility or research priorities into monitoring. These special investigations are intended to be temporary.

**Investigation of cause and investigation of solutions.** Investigation of cause and investigation of solutions or a specific research program is invoked if the occurrence of rare observations has exceeded the critical threshold or some other state has occurred to convince investigators that unwanted effects are occurring, including exceedance of a PEL (Hewitt et al. 2005) (Figure 2). In most, if not all cases, the observed effect likely originates from multiple sources and is therefore cumulative. The purpose of an investigation of cause is to identify the contribution of individual and hypothesized sources responsible for the (likely cumulative) effects observed during monitoring. Investigation of cause is the most scientifically demanding aspect of monitoring but also represents the initial phases of management. Through an ideal IOC/IOS, the contributing factor or factors will be identified and a solution will become apparent and implemented. Following a successful IOC/IOS, monitoring should reinitiate at the surveillance tier to determine whether
the implemented solutions rectify the change that originally triggered the more detailed monitoring.

**Minimal monitoring.** Triggers are tied to important design features of adaptive programs, such as the inclusion and exclusion of study sites over time, changes in frequency or intensity of monitoring, and expansion or reduction in the number of samples, organisms, or environmental endpoints being evaluated. Within the paradigm of adaptive monitoring, study sites that do not enhance understanding of the spatial pattern or variation in the response can be sampled less frequently, but no study sites should be permanently “off” unless deemed redundant. A study site or sites that have shown no exceedances can be downgraded into a minimal program.

A possible scenario for the reduction of sampling frequency of a study site is to match the criteria for upgrading a study site into IOC (i.e., 3 consecutive observations of no change following the stabilization of the estimate of variability, such as SD). Triggering study sites into a lower level of priority and frequency of sampling is meant to avoid the problem of adding new demands, expansion of monitoring and workload, and the consequent need for additional investments. Study sites and measurements added to a (likely) constrained funding envelope in the absence of a mechanism to remove, replace, refine, or optimize the resources for monitoring will inevitably strain the financial and institutional support of the program.

If study sites are de-emphasized following analysis, a process to reinitiate detailed sampling at the original site is required. Reinitiating detailed sampling at a study site could occur after a change in discharge rate, quality or dilution of an effluent, additional development, or an unexpected change in some basic measurement or at some adjacent stations. Changing environmental conditions at other sites is also a clear indication that more information is required. Conceptually, observing exceedances at critical sites, or the disappearance of simple relationships among other stations that originally invoked redundancy, could also reinitialize sampling at a station or inclusion of more detailed endpoints. Alternatively, some sampling at sites where no changes are expected may play a relevant role in testing predictions. Additionally, simple measurements can be maintained and tested for evidence of change. For instance, an observed discontinuity in a relationship between conductivity and alkalinity may serve such a purpose (Bodo 1992). More research is clearly needed prior to the implementation of such candidate tools as triggering mechanisms and for achieving the goals of a particular program, but the approach does offer other advantages. The simplicity of the measurements likely removes barriers to entry among stakeholders and is recommended to integrate community acceptance, participation, and ownership of monitoring activities (Aceves-Bueno et al. 2015). Reduced effort at individual sites may also allow increasing the number of locations included in the program.

If some sampling is occurring at most sites each year, then activating additional Surveillance sampling within a field season may also be possible. This mechanism requires defining a minimum (core) program, which can be defined using surrogate measures of environmental change that quickly and easily activate further sampling if predefined triggers are exceeded. This may also be advantageous where the importance of a given study site has diminished over time but requires a mechanism to reinitialize more extensive monitoring should conditions change. This process can be facilitated by automatic analyses of digitized data (see Data management: tracking change and reporting section).

**Further mechanisms to reduce inference errors**

While the details of the adaptive process may differ among programs, many aspects of adaptive monitoring are tools and mechanisms to reduce inference errors. The main mechanism to control inference errors is by tiered analysis. The control of inference errors is also supported by leveraging several types of analyses and better understanding the assumptions of the program.

**Leveraging types of analysis.** In any monitoring program, analysis of data is often prescribed by a fixed set of tools. These conventional analyses are often based on the most current methodology supported by detailed and careful scientific studies. Directed analyses can have important advantages. Techniques and decisions can be well established and understanding can be wide and immediate. However, analytical rigidity may also have disadvantages, including missing impacts and insensitivity to novelty.

An important advantage of adaptive monitoring, especially for programs envisioned to continue for decades and occur over a large spatial scale, is the availability of data. With the large data sets, there will be opportunities for 2 additional types of analyses: exploratory and experimental. The existence of large data sets provides an opportunity to develop and explore novel analytical techniques, to more deeply scrutinize results and data of previous monitoring as knowledge, technology, capacity, and capability increase, and to search for corroborating evidence of change and linkages and mechanisms among processes. Exploratory analyses outside the routine analysis of the monitoring data or the use of data to detect change in ways unanticipated during the initial design (Arciszewski and Munkittrick 2015) may be critical to the overall success of the program. Despite risks of exploratory analysis, including the Sharpshooter Fallacy and the multiple comparisons problem (Simmons et al. 2011), the remedies are to document choices and limit results of exploration to hypothesis generation (Zuur et al. 2010). Experimentation directed by those hypotheses can provide invaluable information for interpretation, expansion, or contraction of monitoring programs (Munkittrick et al. 2002). An additional benefit of the variety of components (conventional, exploratory, experimental) is to maintain the contributions of various partners and stakeholders who may have different experience, perspectives, and desired outcomes (Gibbons et al. 2008). The intentional integration of additional analytical approaches may also provide.
important access points for junior scientists or students, or for senior researchers from other disciplines with different expertise.

Addressing assumptions. A common source of confusion in monitoring is the misunderstanding of underlying design and/or interpretive assumptions. There will be assumptions associated with design (pseudoreplication, representativeness of reference sites, adequacy of baseline sampling, and inevitable presence of confounding factors), sampling (consistency and standardization of sampling, influence of seasonal and annual variability, appropriateness and sensitivity of the indicator, disturbances), and ecology (site fidelity, species interactions, and indirect effects). Any assumptions need to be recognized and stated, but they may be hidden or obscured. In many cases, the testing of assumptions can be addressed through focused studies.

MANAGING THE MONITORING

Adapting a monitoring program, as defined here, is the progression among tiers based on the occurrence of trigger exceedances. Other aspects beyond tiers and triggers are likely necessary for the successful conceptualization, design, and operation of an adaptive monitoring program. These components include data management and its influence on tracking changes and reporting, accounting for adjusted measurements, and auditing.

Data management: Tracking change and reporting

Depending on the complexity of the study design, the number of specific programs under a monitoring umbrella, and other factors, there may be heavy administrative burden in managing the tiering status of individual sites and of the program, interpreting and understanding differences, and reporting to stakeholders. Similarly, ensuring the quality, long-term integrity, utility, strength, and comparability of data in an adaptive program can present challenges. A variety of resources are available for planning data management (USEPA 2006). They will not be repeated here, other than to emphasize that a robust data management system is a key feature for facilitating success of the monitoring program that begins with tracking differences and reporting.

An obvious approach to achieve data integrity and other aspects is immediate digital entry, processing, and storage of data. This approach has many advantages. At its most basic, digital entry can be used to check for outliers and data entry errors. When such a data entry system is adopted, comparisons can be made in near-real time (Zerger and McDonald 2012). Near-real time analyses can also provide opportunities to increasing effort during a field season when warranted. Furthermore, algorithms could be developed and either partially or fully automated. Irrespective of how quickly the analyses can be done, the intent in near-real time or annual analyses is to expose weaknesses and identify emerging questions as early as possible and enhance the utility of the program.

Digitization of data has further benefits. Presentation of results, including even minimal functionality of historical patterns of particular measurements at particular study sites with an expected range, on websites is an obvious mechanism. Report cards have been used as accessible and comprehensive engagement and actionable tools (Healthy Waterways Report Card 2014). Similarly, unconventional yet simple presentation of data is already a widely appealing option, while also necessarily integrating a variety of analytical approaches and leveraging computational skills and imaginations of data scientists (Yau 2011).

Accounting for adjusted or replaced measurements

As a monitoring program matures, there is a possibility that new measurements may be added and tested, whereas some may be abandoned. However, rigor applied in the initial stages of design, including preoperational monitoring, should resolve some of these types of issues. Regardless, change of techniques over time may be an inevitable component of adaptive monitoring. Common reasons for replacing measurements may be redundancies or inefficiencies. For example, the general pursuit of better, cheaper, faster ways of measuring chemical, physical, and biological components of ecosystems should be expected. Where the quality of data is not suspect and not the reason for replacing a measurement, or set of measurements, meta-analyses can be used to extract and preserve information from the older data sets (Lowell et al. 2005).

Audits

An important component of an environmental monitoring program weaved throughout the discussion thus far is self-audit. Self-audit is necessary to prevent confirmation bias and is a basic aspect of the scientific method. In support of auditing, external reviews can be used to determine if the program is performing as expected (Cundill and Fabricius 2009) or to access the expertise of others not already involved in a particular monitoring program. Regular external reviews have the benefits of evaluating or integrating an esoteric technology or method and integrating alternative expertise to improve the program over time.

CONCLUSIONS AND RECOMMENDATIONS

Articulation of the adaptive monitoring paradigm is an important contribution to detecting and evaluating the relevance of environmental change as part of a cohesive, comprehensive, and strong monitoring program. The design of a useful monitoring system, rather than the pursuit of a perfect and singular monitoring tool (Boyles and Nielsen 2017), has clear benefits. Adaptive monitoring can allow work to begin where little is initially known and allows data to guide design decisions of future programs. The emphasis of hypothesis-driven studies has, as a corollary, an emphasis on data and science-based argumentation and limits consequences of professional bias and hubris (Burgman 2005). The general approach can also provide a foundation for later studies without reinventing design. There may be, however, a heavier administrative burden of an adaptive
program relative to conventional monitoring. For instance, tracking trigger exceedances and following the tier site-specific sampling may require substantive effort.

Although these ideas have been implicitly included in the design and operation of multiple programs, such as EEM (Munkittrick et al. 2002) or the Southern California Coastal Waters Research Program (SCCWRP) (Schiff 2000), they have not been entirely included in others. The information and discussion included here are not exhaustive and are meant to clarify some of the problems of monitoring and their common causes, and to present some solutions. Undoubtedly, some unanticipated issues will arise following the development and operation of long-term adaptive programs, but using the guidance included here and elsewhere, these issues will form the basis of success and improvement, rather than a source of failure. Instead, small mistakes and failures, recognized early in a monitoring program, can be constructive by focusing attention and initiate learning where understanding is incomplete.

Acknowledgment—We would like to thank the various colleagues who have implicitly contributed to this work, including (but not limited to) Frank and Mary Cunningham. We would also like to thank the COSIA Monitoring Priority Area Committee and the presenters at the 2015 COSIA Innovation Summit in Banff, Alberta, Canada, and the peer reviewers for their extremely thorough and valuable contributions. All authors participated in writing and editing of this manuscript. The authors declare no conflict of interest.

Data Accessibility—This paper does not rely on any data, and therefore there is no source to disclose or share.

REFERENCES

Aceves-Bueno E, Adeleye AS, Bradley D, Brandt WT, Callery P, Feraud M, Gander KL, Gentry R, Huang Y, McCullough I. 2015. Citizen science as an approach for overcoming insufficient monitoring and inadequate stakeholder buy-in in adaptive management: criteria and evidence. Ecosystems 18:493–506.

Anderson MJ, Thompson AA. 2004. Multivariate control charts for ecological and environmental monitoring. Ecol Appl 14:1921–1935.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Homung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmidt PK. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–741.

Arciszewski TJ, Munkittrick KR. 2015. Development of an adaptive monitoring framework for long-term programs: An example using indicators of fish health. Integr Environ Assess Manag 11(4):701–718.

Ball MA, Noble BF, Dube MG. 2013. Valued ecosystem components for watershed cumulative effects: An analysis of environmental impact assessments in the South Saskatchewan River watershed, Canada. Integr Environ Assess Manag 9:469–479.

Bammer G. 2003. Integration and implementation sciences: Building a new specialization. Ecol Soc 10:95–107.

Barnett V, O’Hagan A. 1997. Setting environmental standards: The statistical approach to handling uncertainty and variation. London (UK): Chapman & Hall. 111 p.

Barrett TJ, Hille KA, Sharpe RL, Harris KM, Machtans HM, Chapman PM. 2015. Quantifying natural variability as a method to detect environmental change: Definitions of the normal range for a single observation and the mean of m observations. Environ Toxicol Chem 34:1185–1195.

Baum R, Bartram J, Hrudey S. 2016. The Flint water crisis confirms that US drinking water needs improved risk management. Environ Sci Technol 50:5436–5437.

Baylis P, Van Dam R, Bartolo R. 2012. Quantitative ecological risk assessment of the Magela Creek Floodplain in Kakadu National Park, Australia: Comparing point source risks from the ranger uranium mine to diffuse landscape-scale risks. Hum Ecol Risk Assess 18:115–151.

Boiling BA. 1992. Statistical analyses of regional surface water quality in southeastern Ontario. Environ Monit Assess 23:165–187.

Boyles E, Nielsen CK. 2017. Is the ideal sentinel a unicorn? Integr Environ Assess Manag 13:451–452.

British Columbia Ministry of Energy and Mines & British Columbia Ministry of Environment. 2016. Joint Application Information Requirements for Mines Act and Environmental Management Act Permits. Victoria (BC).

Burgman M. 2005. Risks and decisions for conservation and environmental management. Cambridge (UK): Cambridge Univ. 487 p.

Burton GA, Chapman PM, Smith EP. 2002. Weight-of-evidence approaches for assessing ecosystem impairment. Hum Ecol Risk Assess 8:1657–1673.

Cairns Jr J, McCormick PV, Niederlehner B. 1993. A proposed framework for developing indicators of ecosystem health. Hydrobiologia 263:1–44.

Callicott JB. 1999. Beyond the land ethic: More essays in environmental philosophy. Albany (NY): SUNY. 437 p.

Canadian Nuclear Safety Commission. 2013. Environmental protection: Environmental protection policies, programs and procedures. REGDOC-2.9.1. Ottawa (ON).

Chaffin BC, Gosnell H, Cosens BA. 2014. A decade of adaptive governance scholarship: Synthesis and future directions. Ecol Soc 19(3):56. dx.doi.org/10.5751/ES-06824-190356.

Cook CN, de Bie K, Keith DA, Addison PF. 2016. Decision triggers are a critical part of evidence-based conservation. Biol Conserv 195:46–51.

Cormier SM, Suter GW. 2008. A framework for fully integrating environmental assessment. Environ Manag 42:543–556.

Cundill G, Fabricius C. 2009. Monitoring in adaptive co-management: Toward a learning based approach. J Environ Manag 90:3205–3211.

Díaz S, Quétier F, Cáceres DM, Tarnow SF, Pérez-Harguindeguy N, Bret-Harte MS, Finegan B, Peña-Claros M, Poorter L. 2011. Linking functional diversity and social actor strategies in a framework for interdisciplinary analysis of nature’s benefits to society. Proc Natl Acad Sci USA 108:895–902.

Duarte CM, Conley DJ, Carstensen J, Sánchez-Camacho M. 2009. Return to Neverland: Shifting baselines affect eutrophication restoration targets. Estuar Coasts 32:29–36.

Dube M, Munkittrick K. 2001. Integration of effects-based and stressor-based approaches into a holistic framework for cumulative effects assessment in aquatic ecosystems. Hum Ecol Risk Assess 7:247–258.

Dube MG, Duinker P, Greig L, Carver M, Servos M, McMaster M, Noble B, Schreier H, Jackson L, Munkittrick KR. 2013. A framework for assessing cumulative effects in watersheds: An introduction to Canadian case studies. Integr Environ Assess Manag 9:363–369.

Duncker PN, Greig LA. 2006. The impotence of cumulative effects assessment in Canada: Aliments and ideas for redeployment. Environ Manag 37:153–161.

[EC] European Commission. 2015. Communication from the Commission to the European Parliament and the Council. The Water Framework Directive and the Floods Directive: Actions towards the ‘good status’ of EU water and to reduce flood risks. Brussels (BE). [cited 2016 Dec 22]. http://ec.europa.eu/environment/water/water-framework/pdf/4th_report/MS%20Annex%20-%20Belgium.pdf

Field SA, O’Connor PJ, Tyre AJ, Possingham HP. 2007. Making monitoring meaningful. Austral Ecol 32:485–491.

Gibbons P, Zammit C, Youngentob K, Possingham HP, Lindenmayer DB, Bekessy S, Burgman M, Colyvan M, Considine M, Felton A. 2008. Some practical suggestions for improving engagement between researchers and policy-makers in natural resource management. Ecol Manag Restor 9:182–186.

Government of Canada. 2017. Pulp and Paper Effluent Regulations. SOR/92-269. Groffman PM, Baron JS, Blett T, Gold AJ, Goodman I, Gunderson LH, Levinson BM, Palmer MA, Paerl HW, Peterson GD. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? Ecosystems 9:1–13.
Postel S, Richter B. 2012. Rivers for life: Managing water for people and nature. Washington (DC): Island Pr. 220 p.

Ribey SC, Munkittrick KR, McMaster ME, Courtenay S, Langlois C, Munger S, Rosaasen A, Whitley G. 2002. Development of a monitoring design for examining effects in wild fish associated with discharges from metal mines. Water Qual Res J Can 37:229–249.

Schiff K. 2000. Sediment chemistry on the mainland shelf of the Southern California Bight. Mar Pollut Bull 40:268–276.

Shotyk W, Bicalho B, Cuss CW, Donner MW, Grant-Weaver I, Haas-Neill S, Javed MB, Krachler M, Noernberg T, Pelletier R. 2017. Trace metals in the dissolved fraction (< 0.45 μm) of the lower Athabasca River: Analytical challenges and environmental implications. Sci Tot Environ 580:660–669. https://doi.org/10.1016/j.scitotenv.2016.12.012

Simmons JP, Nelson LD, Simonsohn U. 2011. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 22:1359–1366.

Smith RW. 2002. The use of random-model tolerance intervals in environmental monitoring and regulation. J Agric Biol Envr S 7:74–94.

Suter GW. 1990. Endpoints for regional ecological risk assessments. Environ Manag 14:9–23.

Suter GW, Norton SB, Bamthouse LW. 2003. The evolution of frameworks for ecological risk assessment from the Red Book ancestor. Hum Ecol Risk Assess 9:1349–1360.

[USEPA] US Environmental Protection Agency. 2006. Guidance on systematic planning using the data quality objectives process. Washington (DC): EPA QA/G-4. p 120.

Vaughan I, Diamond M, Gumell A, Hall K, Jenkins A, Milner N, Naylor L, Sear D, Woodward G, Ormerod S. 2009. Integrating ecology with hydromorphology: A priority for river science and management. Aquat Conserv 19:113–125.

Vugteveen P, van Katwijk MM, Rouwette E, Lenders HR, Hånssen L. 2015. Developing an effective adaptive monitoring network to support integrated coastal management in a multiuser nature reserve. Ecol Soc 20:59.

Walker S, Ribey S, Trudel L, Porter E. 2003. Canadian environmental effects monitoring: Experiences with pulp and paper and metal mining regulatory programs. Environ Monitor Assess 88:311–326.

Walters C. 1986. Adaptive management of renewable resources. New York (NY): Macmillan. 374 p.

Westgate MJ, Likens GE, Lindenmayer DB. 2013. Adaptive management of biological systems: A review. Biol Conserv 158:128–139.

[WHO] World Health Organization. 2006. Guidelines for drinking-water quality: Incorporating first addendum. 3rd ed. Vol 1, Recommendations. Geneva (CH): WHO.

Wu RS, Siu WH, Shin PK. 2005. Induction, adaptation and recovery of biological responses: implications for environmental monitoring. Mar Pollut Bull 51:623–634.

Yau N. 2011. Visualize this! Indianapolis (IN): Wiley. 358 p.

Zerger A, McDonald W. 2012. Ecoinformatics solutions to support monitoring for improved biodiversity conservation. In: Lindenmayer D, Gibbons P, editors. Biodiversity monitoring in Australia. Collingwood (AU): CSIRO. p 23–32.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1:3–14.