Chapter 17
CADDIS: The Causal Analysis/Diagnosis Decision Information System

Susan B. Norton, Susan M. Cormier, Glenn W. Suter II, Kate Schofield, Lester Yuan, Patricia Shaw-Allen, and C. Richard Ziegler

Abstract Biological monitoring and assessment methods have become indispensable tools for evaluating the condition of aquatic and terrestrial ecosystems. When an undesirable biological condition is observed (e.g., a depauperate fish assemblage), its cause (e.g., toxic substances, excess fine sediments, or nutrients) must be determined in order to design appropriate remedial management actions. Causal analysis challenges environmental scientists to bring together, analyze, and synthesize a broad variety of information from monitoring studies, models, and experiments to determine the probable cause of ecological effects. Decision-support systems can play an important role in improving the efficiency, quality and transparency of causal analyses.

CADDIS (http://www.epa.gov/caddis) is an on-line decision framework for identifying the stressors responsible for undesirable biological conditions in aquatic systems. CADDIS was developed in response to requirements under the U.S. Clean Water Act to develop plans for restoring impaired aquatic systems. CADDIS is based on U.S. EPA’s 2000 Stressor Identification Guidance document, and draws from multiple types of eco-epidemiological evidence. A major update in 2007 added summaries of commonly encountered causes of biological impairment: metals, sediments, nutrients, flow alteration, temperature, ionic strength, low dissolved oxygen, and toxic chemicals. These reviews are designed to help practitioners choose which causes to consider, based on sources, site information, and observed biological effects. A series of conceptual models illustrates connections between sources, stressors and effects. Another major new section provides advice and tools for analyzing data and interpreting results as causal evidence; these tools help quantify associations between any cause and any biological impairment using innovative methods such as species-sensitivity distributions, biological inferences, conditional probability analysis, and quantile regression analysis.
An essential part of the development strategy for CADDIS has been the use of case studies to test the process and tools in different regions, and with different causal factors. Case studies have been conducted in streams on the urbanized east coast and the agriculturally-dominated mid-west to the arid west, and have considered causes including low dissolved oxygen, increased temperature, toxic substances, altered food resources and fine sediments. Lessons learned from the case studies include the importance of a structure for organizing the large variety of evidence that is often available, the need for well-matched reference sites for comparison, the benefits of iterative and directed data collection, and the frequency of surprising results. The case studies illustrate the promise of CADDIS: by building on the foundation of biological monitoring, we can provide a powerful means for improving the health of our aquatic systems.

17.1 Introduction

The Causal Analysis/Diagnosis Decision Information System (CADDIS: http://www.epa.gov/caddis) is an on-line decision support system to help scientists identify the stressors responsible for undesirable biological conditions in aquatic systems.

The development of CADDIS was motivated by the increased use of biological monitoring and survey methods to evaluate aquatic ecosystems (Davis 1995; Ohio Environmental Protection Agency 1987; Plafkin et al. 1989). Fifty-seven U.S. states and tribes currently use biological assessments in water resource management (U.S. Environmental Protection Agency 2002). Biological assessments have also become an integral part of programs in the United Kingdom, Europe, Australia, Canada, New Zealand and South Africa (Marchant 1997; Metcalfe-Smith 1994; Stark and Maxted 2007). Biological assessments often reveal impairments previously overlooked by water quality measurements and that were not necessarily resolved by controlling point source emissions. However, biological assessments do not identify the cause of impairment; they only indicate where conditions are unacceptable. So, when biological assessments indicate that a water body is impaired, the cause of the change in the biological community needs to be identified before it can be rectified.

CADDIS aims to improve the practice of causal analysis of biological effects, by providing a formal inferential methodology and technical content useful for implementing the method. A formal method for making decisions about causation has many benefits; it provides a structure for organizing data and thinking when a situation is complex and provides transparency when a situation is contentious. When remedial alternatives are costly, a formal method can increase confidence that a proposed remedy will truly improve environmental condition.

Although the primary application of CADDIS has been to lotic systems, it is based on principles that are applicable to any ecosystem; lakes, estuaries, and terrestrial systems. CADDIS is designed to be prompted by the results of biological monitoring programs, but the principles can be applied to assessments prompted
by concerns over sources, such as a non-point source inputs, a particular industrial outfall or a hazardous waste site, or over other types of observed effects, such as mass mortalities. In particular, the principles have been applied to assessments of fish kills and impaired populations and communities on contaminated sites.

### 17.2 Content

CADDIS is designed to help practitioners find, analyze and use information to produce causal evaluations in aquatic systems. It contains an inferential process and information needed to apply that process. The inferential process is based on U.S. EPA’s 2000 Stressor Identification Guidance document and draws from multiple types of eco-epidemiological evidence (Section 17.2.1). A major update in 2007 provides tools and reviews that make information useful for causal analysis more accessible (Section 17.2.2).

#### 17.2.1 The Step-by-Step Guide to Stressor Identification

The Step-by-Step Guide to Stressor Identification provides a formal process for making decisions about causation at specific sites (Fig. 17.1). It is a general

![The Stressor Identification Process](image)

**Fig. 17.1** The Stressor Identification Process (shown in the *darker gray box*), within the broader management context
framework that can be applied to the great range of causal scenarios and data availability that investigators encounter.

A formal process for causal analysis can mitigate many of the cognitive shortcomings that arise when we try to make decisions about complex subjects. Common errors include clinging to a favorite hypothesis when it should be doubted; using default rules of thumb that are inappropriate for a particular situation; and favoring data that are conspicuous (Kahneman et al. 1982; Nisbett and Ross 1980; Norton et al. 2003). The general attributes of a good decision process have been the subject of study for several decades. A good decision process provides a means for:

1. Choosing the most appropriate frame or scope for the analysis
2. Collecting the right information for the analysis
3. Organizing the information
4. Reaching conclusions
5. Obtaining feedback on the effectiveness of the decision (Russo and Schoemaker 1989).

The sections below discuss the five steps of the Step-by-Step guide in the context of these characteristics.

### 17.2.1.1 Choosing the Most Appropriate Frame or Scope for the Analysis

Step 1: Defining the Case, and
Step 2: Identifying Candidate Causes

---

**Box 17.1. The Little Scioto River, OH, USA** The Little Scioto River case study was developed to illustrate the application of the Stressor Identification process (Cormier et al. 2002, Norton et al. 2002, Cormier and Ferster 2007). The case study involves a 15-km reach of a river in north-central Ohio (Fig. 17.2).

Many point and non-point sources of pollutants are associated with the Little Scioto River. Point sources include a wastewater treatment plant and combined sewer overflows that enter between 9.5 and 10.5 km, respectively, upstream of its confluence with the Scioto River. Non-point sources include runoff from agricultural land uses and from the city of Marion. Releases may also originate from several contaminated industrial areas, including an abandoned wood treatment plant, a landfill, an appliance plant, and a rail facility (Ohio Environmental Protection Agency 1994). Finally, the stream was channelized in the early 1900s starting from river kilometer 15 and continuing downstream to the confluence with the Scioto River.

Several sites in the Little Scioto Case study area were considered to be biologically impaired based on the results of fish and macroinvertebrate surveys. Specifically, the values of two multimetric indices, the Index of Biotic Integrity (IBI) and the Invertebrate Community Index (ICI) were below criteria set by the Ohio Environmental Protection Agency.
Methods for causal analysis have frequently considered a very general spatial frame, but a targeted candidate cause frame. Examples include: Does smoking cause cancer (Hill 1965)? Can chlorinated dioxins, furans and biphenyls cause deformities in wildlife in the Great Lakes region of the United States (Fox 1991)? The types of causal analyses addressed by CADDIS reverse this scope.

Fig. 17.2 Map of the Little Scioto River case study area (adapted from Cormier et al. 2002)
They focus instead on a more localized spatial frame (e.g., a stream reaches), but consider a full range of candidate causes. They ask questions like: Did excess fine sediments, low dissolved oxygen, or chemical contaminants cause the loss of mayflies in this stream reach?

A causal analysis is prompted by the observation of an undesirable biological effect; a fish kill, a decline in a biological index, or a high incidence of anomalies (e.g., Box 17.1). The evaluation of ecological condition, including selection of appropriate biological indicators and sampling designs is a complex subject in itself and not addressed by CADDIS. Rather, causal analysis begins when these conditions assessments or other observations indicate that something is amiss.

In Step 1 of the CADDIS guide, practitioners begin scoping the analysis by defining the case that will be investigated (e.g., Box 17.2). First, the specific biological effects that will be analyzed are defined. For example, as mentioned above, the biological impairment triggering causal analysis may be a decline in a biological index score. The specific biological effects or metrics that contribute to that decline may include decreases in the abundance of larval stoneflies. Describing the effects in terms of what is actually happening biologically makes it easier to use information on the mechanisms behind the cause, and to use supporting evidence from other areas (e.g., similar situations in other locales, scientific literature, etc.).

**Box 17.2. Defining the Case: The Little Scioto River** The IBI and ICI indices were disaggregated to gain additional insights into the changes occurring in the Little Scioto River. A subset of individual metrics were identified that indicated distinctive changes in the assemblage at different points along the stream reach: the weight of fish normalized to 1 km distance (relative fish weight); the percent of fish having deformities, eroded fins, lesions or tumors (DELT anomalies); percent of macroinvertebrate individuals that were mayflies (percent mayflies); and the percent of macroinvertebrates that were taxa considered to be tolerant of stress (percent tolerant invertebrates).

The pattern of effects changed at different locations on the Little Scioto, indicating that different causes may be operating (Fig. 17.3). For example, the weight of fish normalized to 1 km distance (relative fish weight) increased at Site A, whereas the percent of fish having deformities, eroded fins, lesions or tumors (DELT anomalies) did not increase dramatically until Site B. For this reason, separate causal analyses were performed for each site. The reference sites used for comparison moved incrementally down stream. That is, the furthest upstream site (Site U Rkm 14.9) was used as a baseline for comparison for Site A (Rkm 12.5); both Sites U and Site A were used as baselines for comparison to Site B (Rkm 10.5).

Second, the geographic scope of the analysis is defined. The geographic scope of the case has two parts: (1) the impaired stream reach (or similar
stream part), and (2) other sites within the same aquatic system (e.g., the same stream, watershed, bay, or reservoir) that are either unimpaired or impaired in a different way, for use as a reference for comparison. Whenever possible all locations within a case should be part of the same system. They also should be located relatively close together, and aside from anthropogenic effects, should be as similar as possible physically, chemically, and biologically.

In Step 2, the scope of the analysis is further defined in terms of the candidate causes that will be analyzed (e.g., Fig. 17.4). Rather than trying to prove or disprove a particular candidate cause, CADDIS instead identifies the most probable cause from a list of candidates. Candidate causes are the stressors the organisms either contact (e.g., increased metals) or co-occur with (e.g., lack of suitable habitat). The list of candidate causes is compiled by reviewing available information from the site and from the region. People who have an interest in the assessment may have insights and opinions on candidate causes; these candidate causes should be included in the list so that they can be appropriately addressed.

Fig. 17.3 Patterns of selected fish and macroinvertebrate metrics at different locations within the Little Scioto River case study; percent of macroinvertebrate individuals from mayfly taxa (% mayflies); percent of macroinvertebrate individuals from taxa considered to be tolerant of stress (percent tolerant invertebrates); the percent of fish having deformities, eroded fins, lesions or tumors (DELT anomalies); the weight of fish normalized to 1 km distance (relative fish weight) (adapted from Cormier et al. 2002)
An important part of describing candidate causes is the construction of a conceptual model that describes the linkages between potential sources, stressors or candidate causes, and biological effects in the case (Fig. 17.4). The models show in graphical and narrative form the working hypotheses and assumptions about how and why effects are occurring. They also provide a framework for keeping track of what information is available and relevant to each candidate cause, setting the stage for the next steps of the analysis.
17.2.1.2 Collecting and Organizing Information

Step 3: Evaluating Data from the Case
Step 4: Evaluating Data from Elsewhere

A wide variety of arguments and data analyses can be used to support causal analyses. The objective is to show that fundamental characteristics of a causal relationship are indeed present at the case under investigation; for example, that the effect is associated with a sequential chain or chains of events; that the organisms are exposed to the causes at sufficient levels to produce the effect; that manipulating or otherwise altering the cause will change the effect; and that the proposed cause-effect relationship is consistent with general knowledge of causation in ecological systems.

The Step-by-Step guide walks practitioners through fifteen different types of evidence (Tables 17.1 and 17.2). Confidence in conclusions increases as more types of evidence are evaluated for more candidate causes. Although

<table>
<thead>
<tr>
<th>Table 17.1</th>
<th>Types of evidence that use data from the case</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of evidence</strong></td>
<td><strong>The concept</strong></td>
</tr>
<tr>
<td>Spatial/Temporal Co-occurrence</td>
<td>The biological effect must be observed where and when the cause is observed, and must not be observed where and when the cause is absent.</td>
</tr>
<tr>
<td>Evidence of Exposure or Biological Mechanism</td>
<td>Measurements of the biota show that relevant exposure to the cause has occurred, or that other biological mechanisms linking the cause to the effect have occurred.</td>
</tr>
<tr>
<td>Causal Pathway</td>
<td>Steps in the pathways linking sources to the cause can serve as supplementary or surrogate indicators that the cause and the biological effect are likely to have co-occurred.</td>
</tr>
<tr>
<td>Stressor-Response Relationships from the Field</td>
<td>As exposure to the cause increases, intensity or frequency of the biological effect increases; as exposure to the cause decreases, intensity or frequency of the biological effect decreases.</td>
</tr>
<tr>
<td>Manipulation of Exposure</td>
<td>Field experiments or management actions that increase or decrease exposure to a cause must increase or decrease the biological effect.</td>
</tr>
<tr>
<td>Laboratory Tests of Site Media</td>
<td>Controlled exposure in laboratory tests to causes (usually toxic substances) present in site media should induce biological effects consistent with the effects observed in the field.</td>
</tr>
<tr>
<td>Temporal Sequence</td>
<td>The cause must precede the biological effect.</td>
</tr>
<tr>
<td>Verified Predictions</td>
<td>Knowledge of a cause’s mode of action permits prediction and subsequent confirmation of previously unobserved effects.</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Biological measurements (often at lower levels of biological organization than the effect) can be characteristic of one or a few specific causes.</td>
</tr>
</tbody>
</table>
most assessments will have data for only some of the types of evidence, a ready guide to all of the types of evidence may lead practitioners to seek additional evidence.

The fifteen types of evidence provide a system for organizing the data and information relevant to a causal analysis. Each relevant analysis can be isolated, which is helpful when so much information is being evaluated. Human minds can only retain and process about seven pieces of information at time. Breaking up or chunking information into pieces is a more effective way to manage complex tasks (Nisbett and Ross 1980). Isolating each analysis helps prevent cognitive overloading and our tendencies to give undue weight to information that is easily obtained.

The types of evidence are organized into two sets: those that utilize data from the case itself (Step 3) and those that bring in information from other situations, or biological knowledge (Step 4). Causal analyses often begin with an examination of data from the case at hand (Table 17.1). For example, a field biologist might observe that effects occur when a particular candidate cause is present, but do not occur when it is absent (evidence of spatial co-occurrence). Such associations provide the core of information used for characterizing causes. It is beneficial to evaluate associations from the case first, because they can be powerful enough to eliminate candidate causes from further consideration.

<table>
<thead>
<tr>
<th>Table 17.2</th>
<th>Types of evidence the use data from elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of evidence</strong></td>
<td><strong>The concept</strong></td>
</tr>
<tr>
<td>Stressor-Response Relationships from Other Field Studies</td>
<td>At the impaired sites, the cause must be at levels sufficient to cause similar biological effects in other field studies.</td>
</tr>
<tr>
<td>Stressor-Response Relationships from Laboratory Studies</td>
<td>Within the case, the cause must be at levels associated with related biological effects in laboratory studies.</td>
</tr>
<tr>
<td>Stressor-Response Relationships from Ecological Simulation Models</td>
<td>Within the case, the cause must be at levels associated with effects in mathematical models simulating ecological processes.</td>
</tr>
<tr>
<td>Mechanistically Plausible Cause</td>
<td>The relationship between the cause and biological effect must be consistent with known principles of biology, chemistry and physics, as well as properties of the affected organisms and the receiving environment.</td>
</tr>
<tr>
<td>Manipulation of Exposure at Other Sites</td>
<td>At similarly impacted locations outside the case sites, field experiments or management actions that increase or decrease exposure to a cause must increase or decrease the biological effect.</td>
</tr>
<tr>
<td>Analogous Stressors</td>
<td>Agents similar to the causal agent at the impaired site should lead to similar effects at other sites.</td>
</tr>
</tbody>
</table>
Data from elsewhere may include information from other sites within the region; stressor-response relationships derived from field or laboratory studies; studies of similar situations in other streams, and numerous other kinds of information (Table 17.2). After assembling the information, it must then be related to observations from the case.

CADDIS includes a scoring system, adapted from the system by Susser (Susser 1986), that can be used to summarize the degree to each type of evidence that is available strengthens or weakens the case for a candidate cause. A consistent system for scoring the evidence facilitates the synthesis of the information into a final conclusion. The number of plusses and minuses increases with the degree to which the evidence either supports or weakens the argument for a candidate cause. Evidence can score up to three plusses (+++) or three minuses (----).

There are two other types of scores:

- **Refute (R)** is used for indisputable evidence that disproves that the candidate cause is responsible for the specific effects.
- **Diagnose (D)** is used when a set of symptoms for a particular causal agent or class of agents is, by definition, sufficient evidence of causation, even without the support of other types of evidence.

For example, the scoring table for spatial-temporal co-occurrence is shown in Table 17.3, and is applied in the Little Scioto Case Study in Box 17.3.

<table>
<thead>
<tr>
<th>Finding</th>
<th>Interpretation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>The effect occurs where or when the candidate cause occurs, OR the effect does not occur where or when the candidate cause does not occur.</td>
<td>This finding somewhat supports the case for the candidate cause, but is not strongly supportive because the association could be coincidental.</td>
<td>+</td>
</tr>
<tr>
<td>It is uncertain whether the candidate cause and the effect co-occur.</td>
<td>This finding neither supports nor weakens the case for the candidate cause, because the evidence is ambiguous.</td>
<td>0</td>
</tr>
<tr>
<td>The effect does not occur where or when the candidate cause occurs, OR the effect occurs where or when the candidate cause does not occur.</td>
<td>This finding convincingly weakens the case for the candidate cause, because causes must co-occur with their effects.</td>
<td>- - -</td>
</tr>
<tr>
<td>The effect does not occur where and when the candidate cause occurs, OR the effect occurs where or when the candidate cause does not occur, and the evidence is indisputable.</td>
<td>This finding refutes the case for the candidate cause, because causes must co-occur with their effects.</td>
<td>R</td>
</tr>
</tbody>
</table>
17.2.1.3 Reaching Conclusions

Step 5: Identifying the Probable Cause

After the evidence has been assembled and analyzed, the probable cause may be obvious. However, in many cases, a more systematic approach to synthesizing the evidence is useful for reaching and communicating conclusions. CADDIS provides advice on using the evidence and scores developed in Steps 3 and 4 to identify the probable cause.

Alternative approaches to using a system to reaching causal conclusions include relying on an expert’s knowledge base of patterns and intuition, for example, it “feels” like toxic substances are the cause. In studies of medical diagnoses, intuitive approaches have been shown to yield results that are inconsistent and difficult to replicate (Russo and Schoemaker 1989). Diagnostic accuracy increased when physicians were given the results of probabilistic rules, suggesting that experts can profit from formal analyses (Dawes 2001). Rules of thumb are another alternative; for example, one might apply a rule that any chemical that is above its Ambient Water Quality Criterion is a probable cause. Rules of thumb are most useful when developed and applied to a particular subset of questions; they can be inaccurate and insensitive when applied generally.

Box 17.3. Example Analysis of Spatial Co-Occurrence from the Little Scioto River Case Study  Spatial co-occurrence was evaluated by comparing stressor levels at sites upstream (Site U) of the impaired reach with those downstream. At the first site where the impairment was observed (Site A), concentrations of metals, biochemical oxygen demand (BOD), and nutrients were higher and dissolved oxygen was lower than at the upstream site (where no impairment was observed), so these could not be eliminated as potential causes. Ammonia was not detected at the site, but water column measurements of ammonia are highly variable. The concentrations of polycyclic aromatic hydrocarbons (PAHs) in sediments at Site A were not greater than the upstream site (in fact PAHs were not detected at either location), so PAHs were eliminated as a cause. A summary of the scores is shown in the following table:

<table>
<thead>
<tr>
<th>Candidate cause</th>
<th>Result</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sediment</td>
<td>Elevated sediment co-occurs with impairment</td>
<td>+</td>
</tr>
<tr>
<td>Pool/riffle</td>
<td>Poor pool/riffle condition co-occurs with impairment</td>
<td>+</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>Reduced DO co-occurs with impairment</td>
<td>+</td>
</tr>
<tr>
<td>Ammonia</td>
<td>Elevated ammonia not detected at site</td>
<td>−</td>
</tr>
<tr>
<td>Metals</td>
<td>Elevated metals concentrations observed at impairment</td>
<td>+</td>
</tr>
<tr>
<td>PAHs</td>
<td>Elevated PAHs not detected at Site A</td>
<td>R</td>
</tr>
</tbody>
</table>
The team found that the Little Scioto River could be divided into three general geographical segments—upper, middle, and lower—based on the biological conditions and causal analysis.

- **Upper (Above Site U)** Biologically unimpaired, the upper regions of the river had a diverse community of fish and invertebrates.

- **Middle (Site A to Site U)** The middle segment exhibited less biological diversity, with a fish community dominated by the presence of large carp. The probable cause for that impairment was attributed to channelization and deepening of the stream, as well as alterations of stream habitat and water quality.

- **Lower (Site B to the Confluence with the Scioto)** The lower reaches of the Little Scioto were the most impaired. The fish community had decreased diversity and fewer fish. In addition, individual fish were smaller, and showed an increased incidence of external anomalies and lesions. The biological impairments observed here were consistent with toxicological effects associated with exposure to PAH, metals, and ammonia. (Heavy creosote contamination began approximately 270 m upstream from the confluence of North Rockswale Ditch.) Although the chemical contamination was sufficient to cause the biological impairments, the contaminated portions of the Little Scioto was also affected by the same channelization problems as the middle segment of the river.

CADDIS uses a strength of evidence approach. Evidence for each candidate cause is weighed, then the evidence is compared across all of the candidate causes. The evidence and scores developed in Steps 3 and 4 provide the basis for the conclusions. The scoring approach is advantageous because it incorporates a wide array of information, and the basis for the scoring can be clearly documented and presented.

One of the challenges commonly faced by causal analyses of stream impairments is that evidence is sparse or uneven. Because information is rarely complete across all of the candidate causes, CADDIS does not employ direct comparison or a quantitative Multi-Criteria Decision Analysis approach. The scores are not added. Rather the scores are used to gain an overall sense of the robustness of the underlying body of evidence and to identify the most compelling arguments for or against a candidate cause (Box 17.4).

In the best case, the analysis points clearly to a probable cause or causes. In most cases, it is possible to reduce the number of possibilities. At the least, the analysis identifies data gaps that need to be filled to increase confidence in conclusions.
17.2.1.4 Obtaining Feedback on the Decision

Mechanisms for obtaining feedback on the effectiveness of decisions are an important part of improving decision making processes. However, it is also very difficult to do so. Causal analysis is only one of several activities required to improve and protect biological condition. In some cases, the most effective management action will be obvious after the probable cause has been identified. In many cases, however, the investigation must identify sources and apportion responsibility among them. This can be even more difficult than identifying the stress in the first place (e.g., quantifying the sources of sediment in a large watershed), and may require environmental process models. The identification and implementation of management alternatives can also be a complex process that requires additional analyses (e.g., economic comparisons, engineering feasibility) and stakeholder involvement (Box 17.5). In the best case, a causal analysis is compelling enough to prompt management action, and follow-up monitoring confirms that the management action improved biological condition (Box 17.6).

Box 17.5. Management Actions in the Little Scioto River  Findings from the case study clarified remediation options and likely environmental outcomes. However, the results of the causal analysis were unknown to the U.S. EPA and Ohio EPA resource managers negotiating the remediation of the Little Scioto. A decision was made to focus remediation on removal of the contaminated sediments.

Managing expectations of recovery potential is critical for evaluating remediation success. For the lower segment of the Little Scioto, removing sediment alone, through dredging could reduce exposure to chemical contamination; however, the biological condition would be expected to improve to levels similar to those in the middle segment of the river. Achieving conditions of the most upstream segment of the river would also require extensive, habitat restoration efforts.

17.2.2 Information and Tools

A second major objective of the CADDIS project is to make information relevant to causal arguments more available. Studies of how people form judgments have shown that we are overly influenced by data that are conspicuous and easy to find (Nisbett and Ross 1980). Using readily available information is not a problem in itself: our goal is to expand the range of readily available information and ensure that it is science-based. Our approach is three-fold. First we provide basic information on commonly encountered causes that is applicable to a broad range of assessments. Second, we provide downloadable tools that make it easier for scientists to analyze their own data. Third, we provide databases of stressor-response relationships that may be difficult to find or generate. An advantage of a Web-based system is that this information can be cross-linked with the causal analysis framework described in Section 17.2.1.
Box 17.6. Management Actions and Environmental Outcomes in the Willimantic River Connecticut, USA

Causal Assessment

Monitoring by the Connecticut Department of Environmental Protection in the autumn of 1999 identified biological impairment in the Willimantic River in northeastern Connecticut, USA. The specific biological impairment defined as low numbers of Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa at a site on the Middle River, and low numbers of EPT and non-EPT taxa downstream on the Willimantic River. The assessors suggested that unreported episodic, acute exposures were causing the impairment because the magnitude of the measured candidate causes were deemed insufficient to cause the severe impairment observed at the site. They recommended sampling in a way to localize the area of the impairment.

Management Action

New biological sampling localized the upper bounds of the impairment near a raceway previously obscured by vegetation. A grey discharge was discovered and was traced to a broken pipe under a loading dock of a textile mill. The episodic toxic discharge was confirmed as the probable cause after rerouting of the illicit discharge and observing an increase in number of EPT and non-EPT taxa at two impaired locations.

Environmental Outcome

Three years after rerouting the illicit discharge, the impaired sites reached acceptable biological conditions as defined by the State’s Department of Environmental Protection. These findings have given confidence to the state agency to apply causal analysis to other rivers and we have been able to demonstrate that scientific information can be presented in a way that results in management action that improves the environment.

17.2.2.1 Candidate Causes

CADDIS provides information on eight commonly encountered candidate causes: metals, sediments, nutrients, flow alteration, temperature, ionic strength, and low dissolved oxygen, and toxic chemicals. Currently these reviews are designed to help practitioners choose which causes to consider, based on sources, site information, and observed biological effects.

A series of generic conceptual models illustrates connections between sources, stressors and effects (Fig. 17.4). Conceptual models are graphic representations of the potential links among sources, stressors, and biological responses. These diagrams support decision-making in several ways. Initial development of conceptual model diagrams provides a framework for brainstorming and prioritizing possible stressors and causal pathways. As the causal assessment progresses, these diagrams can help investigators identify data gaps,
track the likelihood of different candidate causes, and, perhaps most im-
p ortantly, clearly and efficiently communicate the logic of the causal analysis to
stakeholders.

Conceptual models also can be powerful tools for organizing and provid-
ing access to information relevant to stream impairment. CADDIS includes
a prototype Interactive Conceptual Models (ICMs) for phosphorus that
builds on the conceptual model diagram. The diagram serves as a structural
framework, or scaffold, for organizing stressor-specific information. Users
can query the diagram (via hyperlinks) to access detailed information rele-
vant to their own decision-making processes; currently, users can select two
or more shapes in the diagram to retrieve literature citations that support
the hypothesized relationship between those shapes. In the future, other
types of information, such as quantitative stressor-response relationships
and measurement techniques, may be built onto the framework in different
layers.

17.2.2.2 Analyzing Data

Another section of the site provides advice and tools for analyzing data and
interpreting results as causal evidence. CADDIS presents selected methods
and describes how to apply them in a causal assessment. The methods range
from well-established exploratory data analysis methods such as scatter plots,
box plots and correlations, to statistical modeling methods like regression,
conditional probability analysis and species sensitivity distributions. The
Analyzing Data section also reviews fundamental concepts and best practices
for data handling. Advice is provided that discuss how the source, quality and
structure of data (e.g., timing, variability) influence how it should be orga-
nized and analyzed for causal analysis. Best practices for interpreting statist-
cal outputs are also provided.

The application of some of these methods to causal analysis has required
adaptation and extension. For example, paleolimnological methods for infer-
ring environmental concentrations from algal species occurrences have been
adapted for use in analyzing macroinvertebrate species occurrence data that are
often available in causal analyses (U.S. Environmental Protection Agency
2006a; Yuan 2007). The use of these approaches to calculate tolerance values
and to predict environmental conditions from biological assemblage informa-
tion is discussed in detail.

CADDIS provides downloadable tools that can quantify associations
between any cause and any biological impairment, including a Species Sensi-
tivity Distribution generator and a downloadable statistical package (CAD-
Stat) that provides a graphical user interface that makes a variety of exploratory
and statistical methods easier to use. As of the writing of this chapter, methods
in CADStat include scatter plots, box plots, correlation, linear regression,
Fig. 17.5 A simple generic conceptual model for nutrients (Source, Schofield 2007.)
quantile regression, conditional probability analysis and classification and regression tree (CART) analysis.

17.2.2.3 Databases

CADDIS provides a series of databases that house information that users can modify and use in their analyses. A library of conceptual models includes both simple and complex generic models for each of the commonly encountered candidate causes described above (e.g., Fig. 17.5). It also contains conceptual models from case studies. Models are provided in downloadable form in Adobe Acrobat ® (.pdf) and Microsoft Power Point ® (.ppt) formats.

The Databases section also contains three databases that contain quantitative stressor-response information. Two of the databases synthesize laboratory test results for metals, yielding concentration-response curves and species sensitivity distributions for metals (U.S. Environmental Protection Agency 2005) (Fig. 17.6). A third database compiles stressor-response associations from regional data (e.g., Fig. 17.7). Practitioners can use these materials with their site data to develop evidence, in particular, stressor-response from laboratory studies, and stressor-response from regional data.

Fig. 17.6 The Species Sensitivity Distribution Gallery provides a collection of generic SSDs for metals. The example shown plots the proportion of species that have LC50’s less than a given concentration of cadmium. Data from U.S. EPA’s ECOTOX database, following the methods described in U.S. EPA 2005
17.3 Structure

CADDIS is implemented primarily using an Adobe ColdFusion® front end to an Oracle® database. The downloadable tools are provided in a variety of formats, from Microsoft Excel® spreadsheets to programs that can be used with the R statistical package (http://cran.r-project.org/index.html). CADStat was implemented using the Graphical User Interface for R (http://rosuda.org/JGR/). The interactive conceptual model is implemented in Macromedia Flash®. The application includes a password-protected editing function that allows U.S. EPA to keep content up to date.

17.4 Case Studies

An essential part of the development strategy for CADDIS has been the use of case studies to test the process and tools in different regions, and with different causal factors. To date, case studies have been conducted to determine the causes of...
macroinvertebrate or fish community attributes in the states of Ohio, Connecticut, Maine, Iowa, Washington, Mississippi, Maine, West Virginia and Virginia. Case studies conducted early in the development cycle, such as the Little Scioto case study discussed above, led us to make the process more linear, and influenced the direction of tool and information development. We expect that testing the analytical tools and information included in the 2007 release of CADDIS will yield additional insights into making causal analyses more defensible and practicable.

17.5 Future Directions

Our overall objective for CADDIS is to provide an on-line destination for information, methods, and experiences relevant to conducting causal analysis in aquatic systems. Toward that end we intend to continue adding content that can be used to evaluate additional candidate causes, for example altered habitat. Analytical methods of interest include those that address multivariate and spatial issues. The methodology, information and tools useful for causal analysis also have applications in risk assessment and setting of benchmarks and criteria; we are working to take advantage of this nexus. Finally, we intend to make CADDIS a user-supported system, building a community of scientists that share information useful for causal and risk assessment.

17.5.1 Multiple Stressors

The stressor-specific information and statistical methods currently presented in CADDIS have emphasized the analysis of individual candidate causes to determine whether they are sufficient to induce the observed effects. However, interactions among causal agents should also be analyzed. Future editions of CADDIS will contain information, tools and guidance for such analyses. Examples include information on frequency and strength of co-occurrence of different stressors under different watershed land-use scenarios, and stressor-response relationships for combinations of stressors that produce response interactively, i.e., with greater- or less-than response or concentration addition.

17.5.2 CADDIS and Risk Assessment

Although CADDIS was developed to determine causes of impairments identified by biological monitoring programs, it can play two roles in risk assessments for contaminated sites. First, it can determine the cause of impairments identified during biological sampling for a site risk assessment. For example, fish community sampling for the Clinch River unit of the Oak Ridge, Tennessee,
Superfund site revealed that fish abundances were low in a contaminated embayment and the fish that were present displayed physical and physiological impairments (Environmental Sciences Division and Jacobs Engineering Group 1996). An ad hoc analysis of the evidence found that contaminants were likely causes of the impairments, but CADDIS would have eased the analysis and made the conclusions more convincing (Suter et al. 1999). CADDIS has been applied to a stream contaminated by a Superfund site (Cormier et al. 2002; Norton et al. 2002). It has also been applied to as yet unpublished assessments of contaminated sites in California, Colorado, Delaware and Tennessee. CADDIS’s distinction between evidence of causal relationships at the site and evidence from elsewhere, its standard types of evidence and its scoring system are all directly applicable to risk assessments.

Second, the statistical modeling tools in CADDIS and the galleries of exposure-response relationships are useful for risk assessments. CADDIS is designed to support inferences from identified effects to uncertain causes, while risk assessment makes inferences from identified causes to uncertain effects, but the models are the same for both. If biological response data are available for the contaminated site and reference locations, the statistical tools such as regression can be used to develop site-specific exposure-response models for the contaminants of concern. Models of exposure-response relationships from regional monitoring data can be used to determine the credibility of potentially causal relationships at the site. In particular, regional monitoring data are commonly available for common stressors of aquatic communities such as sediment, temperature and dissolved oxygen that are alternatives to contaminants as causes of impairment at contaminated sites. These models can be used to evaluate whether the observations of candidate causes and effects at the site are consistent with regional patterns.

### 17.5.3 CADDIS and Criteria Setting

Some environmental stressors do not lend themselves to the traditional methods of deriving environmental quality criteria, because their effects are not readily tested in the laboratory. The U.S. EPA (U.S. Environmental Protection Agency 2006b) has developed an alternative approach for such stressors that is based on the analysis of multiple types of evidence concerning causal relationships that was inspired by CADDIS. That is, different types of field and laboratory data are analyzed using different statistical methods and the results are compared to arrive at a protective level for a particular location or region. The statistical methods for analysis of field data in CADDIS are useful for developing and evaluating alternative approaches to criteria development.
17.5.4 CADDIS as a Platform for Collaborative Information Sharing

The wide variety of information relevant to causal analysis and the need for region and ecosystem specific data argue for evolving CADDIS towards a collaborative platform. In this way the entire community of investigators can share results and advances. CADDIS currently provides a framework and functional context for collecting information on ecological causal relationships. Increasing the availability of such relationships would make CADDIS a more effective decision support tool. What if any individual or group could input—and felt compelled to do so—details about an ecological causal relationship into an online CADDIS platform, that information was assessed, by peers or otherwise, for accuracy to an appropriate degree, and then harnessed to enhance causal assessment efforts?

Recent advancements in technology and innovative online paradigms bring the potential achievement of this task into view. Information gathering mechanisms have recently gained momentum in today’s information technology world. Various terms for these and related efforts include “peer production” or “commons-based peer production,” “collective intelligence,” “crowdsourcing” and “massively distributed collaboration.” Platforms that currently employ such mechanisms include, for example: Wikipedia, (http://www.wikipedia.org), and the Encyclopedia of Life (http://www.eol.org), a project that aims to allow scientists from around world and from different sectors to develop a Web page that holds information specific to each species on Earth.

The CADDIS project team is at the fledgling stages of moving in the direction of a collaborative platform. We see potential in allowing collaborators to enter information through our conceptual model diagrams, given the intuitive nature of these cause and effect illustrations, improvements in graphical user interface design tools (for example, Macromedia’s Flash® and Java™ technology), and their online aesthetic quality. Whether or not a diagrammatic approach is taken, issues related to informatics offer challenges when, for example, users wish to enter different types of stressor-response relationships, with varying levels of accuracy, different naming conventions, varying sample size, and dataset inconsistencies. As such, flexibility of the user-interface and underlying database will be a critical component of this endeavor.

Open participation often spurs concern about accuracy of information. Research scientists have become accustomed to and comfortable with classic peer-reviewed journal articles, whereby articles are submitted for publication and academically reviewed by an expert panel of, say, three peers, for content and accuracy. Models for the scrutiny of internet-based information submissions, whereby anyone can contribute, are evolving; Wikipedia and the Encyclopedia of Life both address such concerns on their Web sites. The CADDIS team suspects that the potential benefits of a collaborative platform—for example, an exponential increase in available causal relationship
information—far outweigh potential challenges associated with the introduction of innovative review processes. A CADDIS-based mass collaboration platform may in fact outperform more traditional academic peer reviewed mechanisms for amassing causal knowledge in terms of speed, usability, accuracy, and cost.

17.6 Conclusions

Investigating causes of adverse effects observed in aquatic systems poses many challenges. Investigators must have a deep understanding of the many ways that physical, chemical and biological stressors impact aquatic systems. They must be able to organize many types of evidence into a coherent whole. In many investigations, the initial information that is available is quite sparse and investigators must be able to decide whether the information is sufficient to support a decision, and describe the value of additional data collection.

CADDIS is designed to make the formal process of causal analysis more accessible and feasible. A formal process forces assessors to confront unexpected and counterintuitive findings. Our experiences have shown a high frequency of surprising results. For example, in one case study, upstream sites that were expected to serve as reference locations were just as biological degraded as downstream sites. In the Willimantic River, the source was intermittent and spatially disjunct from the effects that were initially observed, making detection difficult.

The identification of the stressors responsible for biological degradation is only one step in a management approach that begins with the detection of a biological impairment and ends with an effective management action that restores desired condition. By making causal analyses more defensible and transparent, we hope to increase the application and utility of biological assessment methods, improve the scientific basis for sound management action and contribute to improving the condition of the world’s waters.

References


Ohio Environmental Protection Agency (1987) Biological Criteria for the Protection of Aquatic Life: Volume II: Users Manual for Biological Assessment of Ohio Surface Waters.WQMA-SWS-6. Division of Water Quality Planning and Assessment, Ecological Assessment Section, Columbus, OH


U.S. Environmental Protection Agency (2005) Methods/Indicators for Determining when Metals are the Cause of Biological Impairments of Rivers and Streams: Species Sensitivity Distributions and Chronic Exposure-Response Relationships from Laboratory Data. NCEA-1494. Office of Research and Development, National Center for Environmental Assessment, Cincinnati, OH

U.S. Environmental Protection Agency (2006a) Estimation and application of macroinvertebrate tolerance values. EPA/600/P-04/116A. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC

U.S. Environmental Protection Agency (2006b) Framework for developing suspended and bedded sediments water quality criteria. EPA-822-R-06-001. Washington, DC