1-1-2006

Setting expectations for the ecological condition of running waters: the concept of reference condition

J. L. Stoddard

P. Larsen

C. P. Hawkins
Utah State University

R. K. Johnson

R. H. Norris

Ecological Applications, 16(4), 2006, pp. 1267–1276

SETTING EXPECTATIONS FOR THE ECOLOGICAL CONDITION OF STREAMS: THE CONCEPT OF REFERENCE CONDITION

John L. Stoddard,[1,5] David P. Larsen,[1] Charles P. Hawkins,[2] Richard K. Johnson,[3] and Richard H. Norris[4]

[1]*U.S. Environmental Protection Agency, Office of Research and Development, Western Ecology Division, 200 SW 35th Street, Corvallis, Oregon, USA*
[2]*Department of Aquatic, Watershed, & Earth Resources, Utah State University, Logan, Utah, USA*
[3]*Department of Environmental Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden*
[4]*Cooperative Research Centre for Freshwater Ecology, University of Canberra, Canberra, Australian Capital Territory, Australia*

Abstract. An important component of the biological assessment of stream condition is an evaluation of the direct or indirect effects of human activities or disturbances. The concept of a "reference condition" is increasingly used to describe the standard or benchmark against which current condition is compared. Many individual nations, and the European Union as a whole, have codified the concept of reference condition in legislation aimed at protecting and improving the ecological condition of streams. However, the phrase "reference condition" has many meanings in a variety of contexts. One of the primary purposes of this paper is to bring some consistency to the use of the term. We argue the need for a "reference condition" term that is reserved for referring to the "naturalness" of the biota (structure and function) and that naturalness implies the absence of significant human disturbance or alteration. To avoid the confusion that arises when alternative definitions of reference condition are used, we propose that the original concept of reference condition be preserved in this modified form of the term: "reference condition for biological integrity," or RC(BI). We further urge that these specific terms be used to refer to the concepts and methods used in individual bioassessments to characterize the expected condition to which current conditions are compared: "minimally disturbed condition" (MDC); "historical condition" (HC); "least disturbed condition" (LDC); and "best attainable condition" (BAC). We argue that each of these concepts can be narrowly defined, and each implies specific methods for estimating expectations. We also describe current methods by which these expectations are estimated including: the reference-site approach (condition at minimally or least-disturbed sites); best professional judgment; interpretation of historical condition; extrapolation of empirical models; and evaluation of ambient distributions. Because different assumptions about what constitutes reference condition will have important effects on the final classification of streams into condition classes, we urge that bioassessments be consistent in describing the definitions and methods used to set expectations.

Key words: best attainable condition; bioassessment; Clean Water Act; consistency of terminology needed; historical condition; least disturbed condition; minimally disturbed condition; monitoring; reference condition defined.

Introduction

Human beings, through their great range of activities, have altered the global landscape in a variety of ways. Describing the effects of these activities on the structure and function of aquatic ecosystems and their biota is a fundamental objective of biological assessments, whether the effects are considered singly or in combination, and whether they are local, regional, national, or international. Conducting a biological assessment involves an evaluation of the biota, and should include the environmental factors that have direct and indirect effects on the temporal and spatial variation in the biota. Since the primary focus of a biological assessment is an evaluation of the effect of human activity, a critical element in the process is estimating biological status in the absence of human disturbance.

Most biological assessments are based, either directly or indirectly, on the concept of comparing current condition to natural conditions (structure, composition, function, diversity) in the absence of human disturbance or alteration (i.e., comparison to a pristine, unpolluted, or anthropogenically undisturbed state; Steedman 1994, Hughes 1995, Jackson and Davis 1995, Davies and Jackson 2006). The term *reference condition* has been used to describe the state used to gauge the effects of human activity (Karr and Chu 1999), and the term

Manuscript received 30 September 2004; revised 1 July 2005; accepted 8 August 2005. Corresponding Editor: E. H. Stanley. For reprints of this Invited Feature, see footnote 1, p. 1249.
[5] E-mail: stoddard.john@epamail.epa.gov

reference-condition approach has been applied to the development and use of the concept (Reynoldson et al. 1997, Bailey et al. 1998, Reynoldson and Wright 2000, Bailey et al. 2004). However, one might argue that a reference condition could be described for a variety of purposes related to biological assessment. For example, it could be used to describe the best remaining condition in a region heavily modified by human activity, or the upstream condition in an assessment of the effects of a point source discharge into a stream.

In this paper, we argue the need for a term reserved to describe the condition in the absence of human disturbance and the need for terms that describe conditions at varying levels of human disturbance. We take this stance in response to many hours of confusing discussion and dialogue on the topic among scientists, managers, and government regulators. All of the authors of this paper, and by inference many of its readers, have had the experience of being well engaged in discussions among a mixed group, only to discover that participants use the words "reference condition" to refer to very different biological states, including the condition of ecosystems at some point in the past; the best of today's existing conditions; the condition of systems in the absence of significant human disturbance; or the condition that today's sites might achieve if they were better managed. We define these four different aspects of reference condition as *historical condition*, *least disturbed condition*, *minimally disturbed condition*, and *best attainable condition*, and discuss them in detail below. We also suggest that it is important to reserve a term to refer specifically to the biological condition in the absence of human disturbance. Common arguments about whether it is even possible to determine a reference condition for naturalness in regions that are heavily influenced by human landuse often result from different perspectives on what "reference condition" really means. This paper came about in an effort to provide some consistent terms and definitions on the issue of reference condition. Much of the science of ecological assessment depends on our ability to set expectations, against which the current condition of aquatic ecosystems can be compared. It seems very unlikely that we will all agree on exactly how expectations should be set, but it is important that we all be able to describe how we determine expectations using a consistent set of terms and definitions.

Davies and Jackson (2006), in a companion paper, propose a general model for the common pattern of biological degradation observed across aquatic ecosystems when exposed to increasing stressor loads. That model uses "naturalness" as one end of a biological-condition gradient with "heavily altered" at the other end. The model is used in the context of the U.S. Clean Water Act which allows states to designate uses for aquatic ecosystems (including levels or tiers of aquatic-life uses) and to develop criteria (e.g., physical, chemical, and biological criteria) by which to judge attainment of those uses. Part of the difficulty in the past has been the lack of a standardized way by which levels of aquatic-life uses designated in one state could be compared with those in another. A related issue has been the need for benchmarks, or a "reference condition" (which might be expressed as a biocriterion) by which to judge attainment of different levels of designated uses.

Background

Legislative mandates

Societal concern about human effects on the environment is embodied in a variety of legislative mandates. In the United States, this concern over the condition of aquatic ecosystems is reflected in the Clean Water Act of 1972 (and as amended, U.S. Code title 33, sections 1251–1387). The often-repeated objective of the Act, to "restore and maintain the chemical, physical, and biological integrity of [the] Nation's waters" (U.S. Code title 33, chapter 26, subchapter 1, section 1251(a)) focuses restoration and protection efforts in the United States on the concept of biological (as well as physical and chemical) integrity. Although the Act does not define biological integrity, an examination of the congressional record describing the formulation of the Act implies that naturalness and integrity were redundant, and that integrity captured naturalness. The idea of naturalness as a key part of biological integrity appears in both the House and Senate Committee on Public Works deliberations on the 1972 Clean Water Act (U.S. Senate 1971, U.S. House of Representatives 1972, Davis and Simon 1995). Others have implied naturalness as a benchmark by which to judge the effects of human activities on aquatic (and other) ecosystems. For example, Frey (1977:128; paraphrased by Karr and Dudley [1981]) have described biological integrity, in part, as a "community of organisms having a species composition, diversity and functional organization comparable to those of *natural habitats* within a region" (emphasis added). The National Research Council (1992) evaluated the status of the national effort to restore aquatic ecosystems; in their report, the concept of naturalness is repeatedly emphasized as the goal for restoration, e.g., "restoration is defined as the return of an ecosystem to a close approximation of its condition prior to disturbance" (NRC 1992:2). In a later report proposing ecological indicators for the nation, the National Research Council (2000) suggested the loss of native species diversity as an indicator reflecting human impact.

Within the European Union (and agreed to by several unaffiliated European countries), the recent Water Framework Directive, European Union legislation, (Directive 2000/60/EC of the European Parliament and the Council of 23 October 2000, establishing a framework for Community action in the field of water policy, *available online*)[6] is a significant effort aimed at

[6] ⟨http://europa.eu.int/comm/environment/water/water-framework/index_en.html⟩

preserving and restoring the biodiversity of inland waters, wetlands, and coastal areas. While previous statutes were focused on curbing emissions and monitoring based on chemical indicators, the Water Framework Directive (WFD) focuses on catchment planning and management, and views aquatic ecosystems not as isolated entities but as larger interconnected ecosystems. A key feature of the WFD is its focus on detecting ecological change (i.e., degradation and recovery) and determining what human-generated pressures (or stressors) are acting as drivers of change—these foci imply a heavy reliance on ecological monitoring and biological assessment. One challenging aspect of WFD implementation is the need to establish benchmarks, or reference conditions, to be used in setting water-quality class boundaries and departures from ecological expectations that may be caused by anthropogenic stress.

The WFD (see footnote 7: page 2) defines reference condition in terms of "no or minimal anthropogenic stress" and satisfying the following criteria: (1) reflecting totally, or nearly, undisturbed conditions for hydromorphological elements, general physicochemical elements, and biological-quality elements; (2) having concentrations of specific synthetic pollutants close to zero or below the limit of detection of the most advanced analytical techniques in general use; and (3) exhibiting concentrations of specific nonsynthetic pollutants within the range normally associated with background levels (see footnote 7). According to the WFD, reference conditions are to be linked to stream typologies, and the population of reference sites should represent, as well as possible, the full range of conditions that are expected to occur naturally within the stream type. A reference condition can be either spatially based (the condition of existing sites meeting the above criteria), based on modeling, or developed using a combination of these, and will be used to classify all water bodies into ecological-quality classes. These quality classes, in turn, govern whether specific sites require restoration or remedial action to bring their ecological condition up to an acceptable level.

In 1994 all Australian governments met through the Council of Australian Governments and agreed to an innovative Water Reform Framework for Australia (ANZECC and ARMCANZ 2000). It was agreed that a strategic framework was required to recognize the unique characteristics of Australia's water resources and their vital contribution to the economic, social, and environmental health of Australia. Major reforms were needed to ensure that the trend towards degradation was reversed and that Australia's water resources were used sustainably in the long term. The National River Health Program (Davies 2000) supported the environmental components of the Water Reform Framework. It modified the definitions of biotic integrity mentioned above and described "river health" as: "The ability of the aquatic ecosystem to support and maintain key ecological processes and a community of organisms with a species composition, diversity, and functional organisation as comparable as possible to that of *undisturbed habitats* within the region" [emphasis added][7] (Simpson and Norris 2000).

Difficulties in implementing legislative mandates.— These legislative mandates are consistent in expressing a need to characterize a biological condition that would occur in the absence of human impacts, i.e., a natural state. We face a practical dilemma in describing a reference condition for this state because it is difficult to locate sites (i.e., sampling sites whose data are used to estimate reference condition) that represent the undisturbed state. Often the sampling sites used in estimating reference condition are chosen because they are considered by local experts to be "the best of what's left"— in many regions of the world, these sites bear only passing resemblance to the natural condition that we might think of as reference condition, because the entire population of possible sites has been degraded by widespread human use of the landscape.

As a result, we often have a "reference condition" description that departs from a natural state by, usually, some unknown amount. A secondary result is the evolution of multiple definitions of the term reference condition, and multiple methods used to estimate it. Each has merit and each has some historical precedent. It is not our intention to dictate which of these definitions is appropriate for use in setting regulatory standards, or other official uses under the general heading of environmental management. Instead, our intent is to establish some clarity by proposing specific terms to describe several related concepts.

Defining Reference Condition

To interpret many of the ecological indicators being measured by monitoring programs throughout the world, indicator or index scores for each sampled site need to be evaluated against some expectation, or reference condition. Most commonly, expectations are defined by a range of indicator or index scores—in this sense, "reference condition" describes a distribution (Fig. 1) rather than a single absolute value. The range of values (for any given index or metric) results from sampling error and natural variability, both in time and in space. At any point in time, a set of sites, all in undisturbed condition, will exhibit a range of biological attributes. In addition, single sites in a natural state will vary over time, due to the influences of climate and natural disturbance. Thus the distribution illustrated in Fig. 1 represents the temporal and spatial variability that is inherent in any measure chosen to represent the natural state of ecological systems.

Once this distribution is described, parameters in that distribution can be selected as criteria for classifying the condition of individual sites, and ultimately assessing

[7] ⟨http://www.deh.gov.au/water/rivers/nrhp/⟩

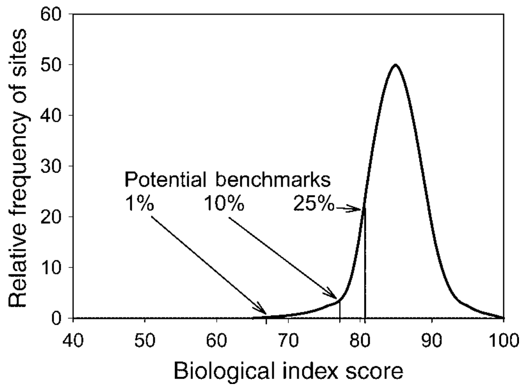

FIG. 1. The range of ecological condition (represented through a hypothetical biological index) found under reference conditions describes a distribution of values rather than an absolute value. The range of this distribution results from natural variability both in time and in space, and can be used to set criteria for assessing or defining classes of ecological condition.

sites of interest. The selection of these criteria can be as much a political decision as a scientific one (e.g., ecological break points), and we will not in this paper try to justify any particular procedure for setting standards or criteria. We do, however, emphasize that the method(s) used to measure and describe the reference distribution can have a profound effect on the final thresholds chosen, by influencing values in the tail(s) of the distribution. Accepting more human-caused disturbance in the distribution has the effect of expanding the distribution and changing the benchmark values. Being more restrictive in the level of acceptable disturbance has the effect of constraining the distribution, and increasing the criteria. Because the quantification of a reference distribution influences the final outcome of biological assessments to such a strong degree, we feel it is important to be consistent in the use of terms relative to the reference-condition concept.

Several of the existing, varied meanings and uses of the term "reference condition" are described below, along with our proposed terms and definitions. We propose that the original concept of reference condition be preserved in a modified form of the term: "reference condition for biological integrity" or RC(BI). We incorporate the biological-integrity modifier to be consistent with the objectives of the U.S. Clean Water Act and legislative intention reflected therein. Designating this phrase for this specific purpose recognizes the need for a "reference condition" term reserved for "naturalness" or "biological integrity" even though we might only approximate it in most parts of the world because of the pervasiveness of human disturbances. For reference condition other than with respect to biotic integrity, we prefer that the following specific terms be used to promote clarity and specificity.

Minimally disturbed condition (MDC)

This term describes the condition of streams in the absence of significant human disturbance, and we submit that MDC is our best approximation or estimate of biotic integrity. We recognize that finding sampling sites that are truly undisturbed by the global influence of human activity is not possible. Therefore, for practical purposes, we incorporate the concept of minimal disturbance (e.g., condition in the presence of atmospheric contaminants well below the threshold for effects, but nonetheless present). One important aspect of MDC is the recognition that some natural variability in indicators will always occur, and this needs to be taken into account when describing MDC. Once established, the distribution created by a group of sites in MDC (e.g., Fig. 1) will vary little over time. Long-term climatic, geologic, and ecological fluctuations will inevitably change the characteristics of individual sites within this distribution, but the range of MDC should be nearly invariant, and its distribution can serve as a nearly invariant anchor by which to judge current condition.

Historical condition (HC)

This term describes the condition of streams at some point in their history. It may be an accurate estimator of true RC(BI) if the historical point chosen is before the start of any human disturbance. However, many other historical reference points are possible (e.g., pre-industrial, pre-Columbian). Two examples of HC in current use are as follows.

1) *Pre-intensive agriculture*. For the EU Water Framework Directive, "reference condition is a state in the present or in the past corresponding to very low pressure, without the effects of major industrialization, urbanization and intensification of agriculture, and with only very minor modification of physicochemistry, hydromorphology and biology" (Wallin et al. 2003:36). This corresponds to "pre-intensive agriculture or impacts compatible with pressures pre-dating any recent land-use intensification" and "pressures pre-dating any recent intensification in airborne inputs that could lead to water acidification" (Wallin et al. 2003:52). This description of HC implies no fixed date, but rather defines a specific stage in the development of human use of the landscape. This state is considered to have been reached ca. 1850 in Great Britain, but may have been reached as early as the 17th century in Germany (Wallin et al. 2003).

2) *Presettlement*. In North America, a historical period that includes the impact of indigenous peoples, but excludes the impacts of European immigrants, has been suggested to define HC (Hughes et al. 1998). Broadly described as "pre-Columbian" this benchmark again does not define a specific historical period, but instead varies according to the westward migration of settlers. It may describe a period as early as the 18th century in the northeastern United States, but as late as

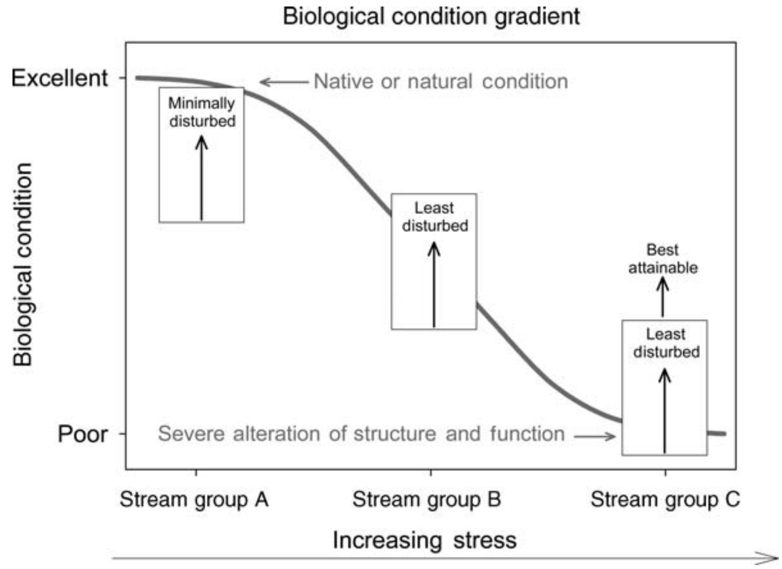

FIG. 2. The differing levels of human disturbance of the landscape in different ecological regions (or in different stream types or stream sizes) create a situation where the least-disturbed sites remaining in each region describe very different definitions of "reference condition." (The best attainable condition will never be better than the minimally disturbed condition or worse than the least disturbed condition but may be equivalent to either depending on the level of human disturbance in a region.) Here, Stream Group B is distinguished from Stream Group C in that the level of degradation is greater for C, and that a reasonable goal or "reference condition" for C might be a condition that does not presently exist, but could be achieved with reasonable management (illustrated as "best attainable"). In contrast, the condition at the least disturbed sites for Group B might be a reasonable goal or reference condition for these streams.

the early 20th century in the western United States and Canada.

3) In more recently settled countries such as Australia, HC is usually defined as pre-European (pre-1750) occupation, although it is recognized that native Australian occupation (ca. 40,000 years ago) also may have caused significant changes to the environment (Norris and Thoms 1999).

Least disturbed condition (LDC)

Least disturbed condition is found in conjunction with the best available physical, chemical, and biological habitat conditions given today's state of the landscape. It is ideally described by evaluating data collected at sites selected according to a set of explicit criteria defining what is "best" (or least disturbed by human activities; (see Bailey et al. 2004:Chapter 3, Hughes et al. 1986, Hughes 1995). These criteria will vary from region to region, and are developed iteratively with the goal of establishing the least amount of ambient human disturbance (e.g., $<1\%$ agricultural land use, $<3\%$ agriculture, $<20\%$ agriculture, etc.) in the region under study. The specifics of these criteria will vary across ecological regions, as the characteristics of the landscape, and human use of the landscape, vary. Because the condition of the environment changes over time, as either degradation or restoration proceeds, LDC may vary with time. As the ecological condition of the very best available sites changes through time, so will our measure of LDC. A desirable feature of distinguishing MDC from LDC is the potential for estimating how different the best current conditions are from RC(BI), and to facilitate comparison of ecological condition among regions.

Best attainable condition (BAC)

Best attainable condition is equivalent to the expected ecological condition of least-disturbed sites if the best possible management practices were in use for some period of time. Sites in BAC would be places where the impact on biota of inevitable land use is minimized. This is a somewhat theoretical condition predicted by the convergence of management goals, best available technology, prevailing use of the landscape, and public commitment to achieving environmental goals. The upper and lower limits on BAC are set by the definitions of MDC and LDC respectively (Fig. 2). Best attainable condition will never be "better" than MDC, nor "worse" than LDC, but may be equivalent to either, depending on the prevailing level of human disturbance in a region. As is the case with LDC, BAC is not invariant, because all of the factors influencing it (e.g., available technology, public commitment) will vary over time.

The dilemma created by multiple definitions of reference condition is illustrated in Fig. 2, for three hypothetical groups of streams with differing levels of human disturbance. One might think of the groups as: (1) headwater streams in an Alaskan wilderness area (Group A); (2) low-order streams in a well-managed Scandinavian forest (Group B); or (3) small streams in a heavily agricultural region in the Midwestern United States or Europe, or in Australia's Murray-Darling Basin (Group C). In a relatively pristine group of streams (Group A—where there may be some human disturbance, even though they are located in wilderness areas), the best streams would all be in MDC—all variability in their condition would be due primarily to natural causes. In this case, reference sites chosen by any number of methods would be very accurate estimators of true RC(BI)—reference condition for biological integrity. As the level of human disturbance increases (say, in a neighboring region subject to more human activity), the least-disturbed streams no longer represent minimally disturbed conditions. Reference sites chosen as "the best of what's left" would meet the LDC definition, but not the MDC; they may also represent the best attainable condition if it is impossible to mitigate the effects of human activity any more than they already are in the least-disturbed sites. At a more extreme level of disturbance (lower right box in Fig. 2), the least-disturbed sites represent neither the MDC nor the BAC. The level of disturbance in this hypothetical group of streams is high enough that it might be relatively easy to improve the ecological condition of the least-disturbed sites through the implementation of better management practices. The hypothetical example in Fig. 2 illustrates how using reference sites that represent a variety of conditions relative to biotic integrity may imply very different definitions of reference condition, and result in quite different ecological assessments.

Methods of Estimating Reference Condition

The reference-site approach

By far the most common approach for estimating the various reference states is to quantify the biological condition at a set of sites that are either minimally or least disturbed by human activity. This approach is widely known as the "reference-site approach" (Hughes 1995, Bailey et al. 2004), and is a scientifically sound method for setting expectations, provided that the form of reference condition that the reference sites represent is clearly defined.

Describing the condition at minimally disturbed sites.— Although the human footprint is pervasive across the landscape, in some ecological regions there might be places that have escaped all but the broadest-scale human disturbances (e.g., escaped all but minimal pollutant exposure derived from long-range transport of atmospheric pollutants). In these regions, characterizing sites that meet an agreed-upon set of minimal-disturbance criteria can provide the data needed for describing minimally disturbed condition, MDC. Protected areas (e.g., designated wilderness areas), forested landscapes with remnants of late-stage/old-growth watersheds; landscapes that have substantially recovered from past disturbances (re-growth of mature forests in areas that were earlier logged; grazed landscapes that are now protected) can contain stream sites the condition of which approximates reference condition for biological integrity, RC(BI), at least for some assemblages (e.g., periphyton, macroinvertebrates).

Describing the condition of least-disturbed sites.— Regardless of the extent of human disturbance in a region, some watersheds/stream sites will have less human disturbance than others and these yield the best existing condition. The challenge is to find and document the least-disturbed sites in a region or class, and to describe their biological characteristics quantitatively. Ideally, the least-disturbed sites would be sites in which biota were exposed to the lowest stressor regime. However, obtaining a complete picture of stressor exposure is usually impractical and expensive. As a result, sites are ranked by an evaluation of a variety of indicators of potential stressors. These indicators may include a combination of those measured at the scale of stream reaches and catchments (e.g., Bryce et al. 1999). A preferred approach is to establish a set of criteria that, in total, describe the characteristics of sites in a region that are the least exposed to stressors. An example of the criteria approach is given in Gerritsen et al. (1993) for the Ridge and Valley ecoregion of the Mid-Atlantic states of the United States. To locate the least-disturbed sites for a large stream monitoring effort, a series of general guidelines were developed, describing land use (largely forested), livestock (no cattle or sheep directly in the streams), riparian vegetation (present), in-stream substrate (minimal siltation) and water quality (un-acidified) characteristics in the least-disturbed portions of the region.

Following these general guidelines, local resource managers were asked to recommend candidate least disturbed condition (LDC) sites for sampling. A total of 31 sites were suggested and sampled in 1993, in conjunction with a probability sampling of streams in the region (U.S. Environmental Protection Agency 2000*a*). The large sample size used in the probability sample ($\sim$500 stream reaches) allowed researchers to describe the range of conditions across the population of streams in the region. Researchers also developed screening criteria to extract the least-disturbed sites from all of the sampled sites (Table 1), and applied these criteria to both the probability and hand-selected sites. Sites that passed all of the listed criteria were identified as least disturbed. Some example results for a macroinvertebrate index of biotic integrity (MBII; Klemm et al. 2003) show how the selection of least-disturbed sites can be evaluated against the condition of the entire population of streams in the region, inferred from the probability sites (Fig. 3). The probability sites exhibit a range of MBII scores from near zero to near 100 (Fig. 3a).

The LDC sites identified by this filtering process, applied to both the probability and hand-selected sites, exhibited IBI scores that closely matched expectations— values were skewed toward the upper end of the

TABLE 1. Variables used to screen the probability sites and filtering criteria used to produce a list of least-disturbed-condition (LDC) sites (based on Waite et al. [2000]).

Variable	Range of values in probability sample	Criterion for inclusion in LDC list
Acid neutralizing capacity	−2000 to 5600 μeq/L	>50 μeq/L
Sulfate	6 to 34 000 μeq/L	<400 μeq/L
Total P	0 to 700 μg/L	<20 μg/L
Total N	45 to 22 000 μg/L	<750 μg/L
Chloride	7 to 2000 μeq/L	<100 μeq/L
Rapid bioassessment habitat score	2 to 20	>15

distribution. Interestingly, roughly 20% of the Mid-Atlantic stream population was estimated to be in LDC based on this exercise (Fig. 3b); the proportion of hand-selected sites that met these criteria was also ~20% (Fig. 3c).

The Mid-Atlantic study used in this example had the luxury of being able to test the results of various methods for determining LDC against the distribution for the entire population of streams. This study allows us to state with confidence that good estimates of LDC can result from an approach that focuses on well-defined criteria (e.g., Table 1). One possible conclusion from the comparison in Fig. 3 is that it is more important (from a reference-condition viewpoint) to focus on adherence to good criteria than on any particular site-selection approach.

Chemical, physical, or biological criteria?—The preferred approach for estimating either MDC or LDC is to use a set of criteria for site selection that, with a few exceptions noted, exclude data on resident biota. Our concern is one of circularity—the structure of the biotic assemblage itself should not be used to classify sites as either reference or non-reference, because we want to avoid any preconceived notions about the structure and composition of biotic assemblages at a "typical" reference site (Bailey et al. 2004). Perhaps most important, the goal of the reference-site approach is to define and describe the amount of natural variability present at sites in the absence of human disturbance, and we cannot know a priori how much variation is typical of any given assemblage among a population of reference sites. We therefore argue strongly against the selection of MDC or LDC sites based on judgment of which biota *ought* to be present. Instead, we urge the use of independent criteria in the selection of sites. Once the set of sites is selected by application of the independent criteria (e.g., through one of the methods discussed here), then an evaluation of the resident biota defines the distribution of MDC or LDC for the region through the derivation of appropriate biological indices.

Some exceptions to the non-use of biological data for reference-site selection may be warranted. For example, the presence of specific stressors, such as toxic compounds, are often best evaluated based on the presence of anomalies or lesions on fish. In an effort to exclude sites contaminated by toxics from a reference distribution, it may be more cost effective to use the organisms themselves to gauge the presence of toxics than to measure toxics directly. In areas where the presence of nonnative species is considered a stress, we may wish to focus on those reference sites where the assemblages are made up only of natives. It is important to note, however, that in both of these exceptions biological data are being used to exclude a form of stress that cannot be

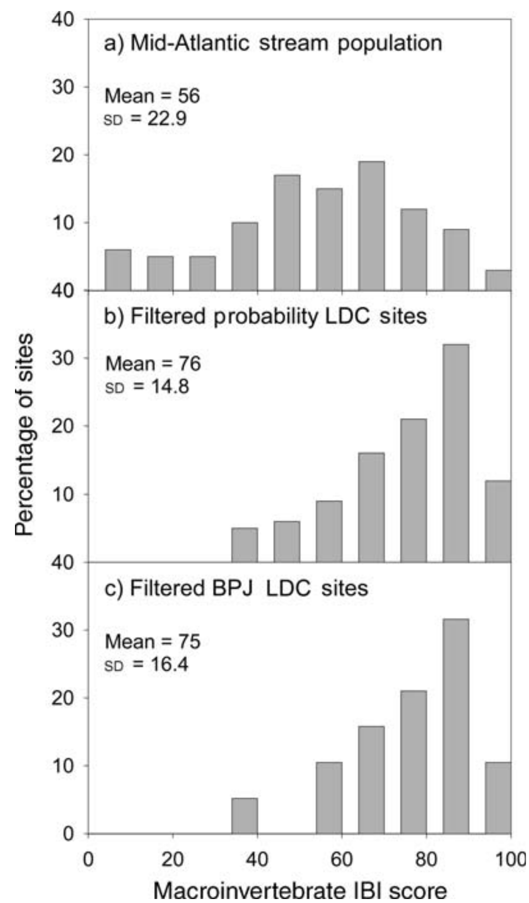

FIG. 3. Macroinvertebrate index of biotic integrity (IBI) scores (see *Methods of estimating reference conditions: Reference-site approach: Describing the condition of least disturbed sites*). (a) A regional population of streams in the Mid-Atlantic region of the United States. (b) A set of least disturbed condition (LDC) sites identified a posteriori by filtering the probability data shown in panel (a) (Waite et al. 2000). (c) A set of LDC sites identified a posteriori by filtering hand-selected, best-professional-judgment (BPJ) sites.

deduced (or would be too expensive to deduce) from independent data.

Other estimation methods

Best professional judgment (BPJ).—Experienced aquatic biologists, with perhaps decades of experience sampling and examining physical, chemical, and biological attributes across wide ranges of severity and types of human disturbance, develop an empirical understanding of condition in the absence of significant human disturbance. Although not always quantifiable in an explicit manner, professional judgment can provide valid insights into both MDC and LDC. Desirable features of the use of BPJ would be a sound basis (or justification) in ecological theory; the inclusion of "replication" by others with comparable experience; documentation of the "rules" by which the experts develop MDC or LDC; and description (to the extent feasible) of how the expert came to his or her conclusions. For example, in the Mid-Atlantic Highlands, the U.S. EPA used a verbal description of current riparian LDC: (1) a multi-storied corridor of woody vegetation; (2) canopies that are closed (or nearly closed); and (3) riparian areas free of visible human disturbance (trash, roads, fences, etc.). This vision of how riparian LDC appears was then translated to quantitative measures of riparian habitat (i.e., presence/absence of multiple canopy layers throughout stream reach, percentage of canopy cover, and riparian disturbance) for use in an ecological assessment (U.S. Environmental Protection Agency 2000*a*).

Interpreting historical condition.—In some cases it is feasible to examine records from earlier times, whether these records are derived from stream samples (e.g., early fish surveys) and museum collections, journals, and other records written by early explorers, land survey notes, or early photographs (e.g., Hughes et al. 1998). Careful perusal and interpretation of these types of records can sometimes provide insight into conditions that existed before extensive human disturbance. Historical condition can also sometimes be inferred by measuring current-day indicators that maintain a record of the past—e.g., historical lake conditions can often be inferred from the composition of diatoms and other deposited planktonic organisms in lake sediments. Pollen profiles in deposited sediments can also be used to interpret watershed vegetation cover.

Extrapolating from empirical models.—In situations without minimally disturbed sites, empirical models derived from associations between biological indicators and human-disturbance gradients can be extrapolated to infer conditions in the absence of human disturbance (e.g., Karr and Chu 1999). The models might be simple univariate plots of the indicator scores against a measure of disturbance; MDC would be inferred as the *y*-intercept (if indicator scores are plotted on the *y*-axis and disturbance scores on the *x*-axis). These plots are often wedge shaped, in which case, inference might be derived by extrapolating the upper bound of the wedge to the *y*-axis. The models might be multiple regressions, in which case, MDC would be inferred by setting the coefficients for the human-disturbance variables to zero. Certainly caution is advised if extrapolations are to be used because estimates occur outside the range of the model "calibration."

Ambient distributions.—By making some assumptions about current data, LDC can be estimated through various interpretations of the range of index or metric values currently observed in a region. For example, in its guidance on developing nutrient criteria, the U.S. EPA Office of Water suggests using the 5th or 25th percentile value of current nutrient concentrations in a regional population as a criterion to separate acceptable from unacceptable values (in the absence of reliable estimates of reference condition, or dose–response relationships for biota affected by nutrient enrichment; U.S. Environmental Protection Agency 2000*b*). Obvious limitations of this approach are that: (1) it requires an a priori decision about what proportion of the regional population is considered to be in LDC (Are 5% of streams in acceptable condition [leading to use of the 5th percentile value], or is 25% a more reasonable value?); (2) it assumes that higher index scores represent better conditions, rather than just different environments (e.g., smaller vs. larger streams) that require separate estimates of expectations; and (3) it is dependent on the distribution of sites sampled relative to the range of the indicator.

Another variant of the ambient-distribution approach is commonly used to score species-richness metrics, as used in indices of biotic integrity (IBIs; Karr 1981). The approach involves an assumption that, for any specific stream type or stream size, LDC is represented by the highest species richness (Fig. 4). Often this assumption is used to develop maximum species richness (MSR) lines, as illustrated in Fig. 4. The MSR line would be used to set expectations ("reference conditions") for each sampled stream, with the expectation equal to the maximum species richness for a site of equal catchment area. In practice, the assumption that the highest number of taxa is found in the least-disturbed sites is usually not tested. In Fig. 4b, we highlight the least-disturbed sites in the Mid-Atlantic Highlands data set (McCormick et al. 2001). An obvious concern results when the MSR line describes the condition of sites that are not in LDC, as in this example. Several potential explanations for this pattern are possible, including the intermediate-disturbance hypothesis (e.g., Connell 1978), but regardless of mechanism, caution is urged when making conclusions about the location of least-disturbed sites within a current-day distribution of biotic index or metric scores.

In the absence of a verifiable set of minimally disturbed or least-disturbed sites, integrating several approaches may lead to firmer, more defensible conclusions about MDC and LDC, particularly if the

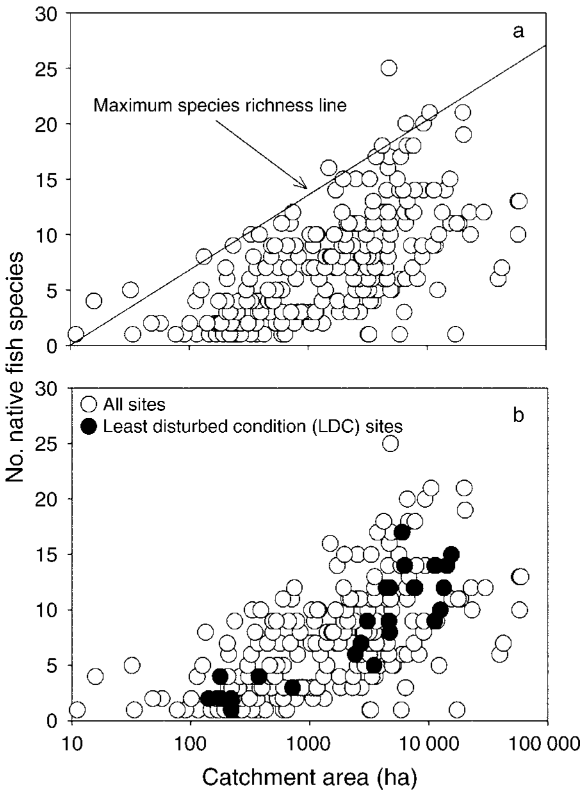

FIG. 4. Illustration of the use of ambient data to estimate least disturbed conditions: (a) maximum species richness (MSR) line applied to a taxonomic-richness metric (native-fish species richness) for the Mid-Atlantic Highlands that varies with catchment (or stream) size; (b) the same relationship, with values for least disturbed condition (LDC) sites shown as solid circles.

conclusions derived from the different approaches are consistent.

CONCLUSIONS

We propose a conceptual framework for a better understanding of the reference-condition concept and its use for stream bioassessments to help clarify the many current uses of the concept. Various legislative mandates specify "naturalness" as a goal, to provide a benchmark against which to judge effects of human activity on stream biota. Many definitions of the term "reference condition" are consistent with the legislative mandates, however the term has also been used to refer to multiple, and often confusing and contradictory concepts. We propose that the term "reference condition for biological integrity," or RC(BI), be reserved as a definitive benchmark to capture the original intent of efforts to maintain and/or restore biological condition to some state of naturalness (including the Clean Water Act in the United States, the Water Framework Directive in the European Union, and the Water Reform Framework in Australia). We also suggest that other terms can be used to specify different states. The term "minimally disturbed condition" (MDC) refers to the biological state at stream sites that show only slight signs of human disturbances, whereas the term "least disturbed condition" (LDC) refers to the biological state at stream sites that exhibit the lowest signs of human disturbance in an area with extensive human disturbance. The term "best attainable condition" (BAC) can be used to specify a state that is better than any in existence in a heavily modified region, but differs from either MDC or RC(BI) because those states might not be achievable. The term "historical condition" (HC) refers to a state interpreted from historical records (journals, early surveys, early photographs), or from remains (e.g., pollen or diatoms in lake sediments) that can be used to reconstruct condition in times of lower human disturbance. We urge practitioners to be more specific in their use of the generic phrase "reference condition." Referring specifically to RC(BI) when appropriate, and stating clearly which practical definition of reference condition is being used to set expectations, will go a long way toward making biological assessments more comparable.

ACKNOWLEDGMENTS

The authors acknowledge the useful comments from William Swietlik, Robert Hughes, Thom Whittier, and Sandy Bryce, as well as countless discussions with other colleagues who helped improve the quality of this paper. Two anonymous reviews also contributed substantially to the final characteristics of the paper. C. P. Hawkins was supported by EPA STAR grant R-82863701.

The information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

LITERATURE CITED

ANZECC and ARMCANZ [Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand]. 2000. Australian and New Zealand guidelines for fresh and marine water quality. Volume 2. Aquatic ecosystems—rationale and background information. Australian and New Zeland Environment and Conservation Council (ANZEEC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ), Australian Government, Department of the Environment and Heritage, Canberra, Australia.

Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology **39**:765–774.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Kluwer Academic Publishers, New York, New York, USA.

Bryce, S. A., D. P. Larsen, R. M. Hughes, and P. R. Kaufmann. 1999. Assessing relative risks to aquatic ecosystems: A Mid-Appalachian case study. Journal of the American Water Resources Association **35**:23–36.

Connell, J. H. 1978. Diversity in tropical rain forests and coral reefs. Science **199**:1302–1310.

Davies, P. E. 2000. Development of a national river bioassessment system (AUSRIVAS) in Australia. Pages 113–128 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Davies, S. P., and S. K. Jackson. 2006. The biological-condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications **16**: 1251–1266.

Davis, W. S., and T. P. Simon, editors. 1995. Biological assessment and criteria: tools for water resource planning and decision making. CRC Press, Boca Raton, Florida, USA.

Frey, D. G. 1977. Biological integrity of water—a historical approach. Pages 127–140 *in* R. K. Ballentine and L. J. Guarria, editors. The integrity of water. Proceedings of a symposium. U.S. Environmental Protection Agency. Washington, D.C., USA.

Gerritsen, J., J. Green, and R. Preston. 1993. Establishment of regional reference conditions for stream biological assessment and watershed management. Pages 797–801 *in* Watershed '95, a national conference on watershed management, Alexandria, Virginia, USA. EPA 840-R-94-002. U.S. Environmental Protection Agency, Washington, D.C., USA.

Hughes, R. M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages Chapter 4. Pages 31–47 *in* W. Davis and T. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making for rivers and streams. Lewis, Boca Raton, Florida, USA.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, T. M. Kincaid, L. Reynolds, and D. P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences **55**: 1618–1631.

Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management **10**:629–635.

Jackson, S., and W. Davis. 1995. Meeting the goal of biological integrity in water-resource programs of the U.S. Environmental Protection Agency. Journal of the North American Benthological Society **13**:592–597.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries **6**(6):21–27.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C., USA.

Karr, J. R., and D. R. Dudley. 1981. Ecological perspective on water quality goals. Environmental Management **5**:55–68.

Klemm, D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and W. T. Thoeny. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. Environmental Management **31**:656–669.

McCormick, F. H., R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands region. Transactions of the American Fisheries Society **130**: 857–877.

NRC [National Research Council]. 1992. Restoration of aquatic ecosystems: science, technology and public policy. National Academy Press, Washington, D.C., USA.

NRC [National Research Council]. 2000. Ecological indicators for the nation. National Academy Press, Washington, D.C., USA.

Norris, R. H., and M. Thoms. 1999. What is river health? Freshwater Biology **41**:197–210.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society **16**: 833–852.

Reynoldson, T. B., and J. F. Wright. 2000. The reference condition: problems and solutions. Pages 293–303 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of fresh waters: RIVPACS and other Techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Simpson, J. C., and R. H. Norris. 2000. Biological assessment of water quality: development of AUSRIVAS models and outputs. Pages 125–142 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Steedman, R. J. 1994. Ecosystem health as a management goal. Journal of the North American Benthological Society **13**: 605–610.

USEPA [U.S. Environmental Protection Agency]. 2000*a*. Mid-Atlantic Highlands streams assessment. EPA/903/R-00/015. U.S. Environmental Protection Agency, Region 3, Philadelphia, Pennsylvania, USA.

USEPA [U.S. Environmental Protection Agency]. 2000*b*. Nutrient criteria technical guidance manual: rivers and streams. EPA-822-B-00–002. U.S. EPA Office of Water, Washington, D.C., USA.

U.S. Senate, Public Works Committee. 1971. Report of the Senate Public Works Committee, United States Senate, together with supplemental views to accompany S. 2770. Federal Water Pollution Control Act Amendments of 1971, 92nd Congress, 1st session. Senate Report 92-414, U.S. Government Printing Office, Washington, D.C., USA.

U.S. House of Representatives, Committee on Public Works. 1972. Report of the Committee on Public Works, United States House of Representatives, with additional and supplemental views. H.R. 11896 to amend the Federal Water Pollution Control Act, 92nd Congress, 2nd session. House Report 92–911, U.S. Government Printing Office, Washington, D.C., USA.

Waite, I. R., A. Herlihy, D. P. Larsen, and D. J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. Journal of the North American Benthological Society **19**:429–441.

Wallin, M., T. Wiederholm, and R. K. Johnson. 2003. Final guidance on establishing reference conditions and ecological status class boundaries for inland surface waters. EU Common Implementation Strategy (CIS) for the Water Framework Directive.