Water quality criteria for an acidifying ocean: Challenges and opportunities for improvement

Stephen B. Weisberg a, *, Nina Bednarsek b, Richard A. Feely c, Francis Chan d, Alexandria B. Boehm e, Martha Sutula a, Jennifer L. Ruesink f, Burke Hales g, John L. Largier h, Jan A. Newton i

a Southern California Coastal Water Research Project Authority, 3535 Harbor Blvd., Costa Mesa, CA 92626, USA
b University of Washington, School of Marine and Environmental Affairs, 3707 Brooklyn Avenue NE, Seattle, WA 98105, USA
c National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115, USA
d Oregon State University, Department of Integrative Biology, 3029 Cordley Hall, Corvallis, OR 97331, USA
e Stanford University, Department of Civil & Environmental Engineering, 473 Via Ortega, Stanford, CA 94305, USA
f Department of Biology, University of Washington, Box 35100, Seattle, WA 98195, USA
g Oregon State University, College of Earth, Ocean, and Atmospheric Sciences, 104 CEOS Administration Building, Corvallis, OR 97331, USA
h University of California Davis, Bodega Marine Laboratory, PO Box 247, Bodega Bay, CA 94923, USA
i University of Washington, Applied Physics Laboratory, Washington Ocean Acidification Center, Seattle, WA 98105, USA

Article info
Article history:
Received 25 October 2015
Received in revised form 14 March 2016
Accepted 21 March 2016
Available online 13 April 2016

Keywords:
Water quality criteria
Acidiﬁcation
303(d)
Pteropods

ABSTRACT
Acidiﬁcation has sparked discussion about whether regulatory agencies should place coastal waters on the Clean Water Act 303(d) impaired water bodies list. Here we describe scientiﬁc challenges in assessing impairment with existing data, exploring use of both pH and biological criteria. Application of pH criteria is challenging because present coastal pH levels fall within the allowable criteria range, but the existing criteria allow for pH levels that are known to cause extensive biological damage. Moreover, some states express their water quality criteria as change from natural conditions, but the spatio-temporal distribution and quality of existing coastal pH data are insufﬁcient to deﬁne natural condition. Biological criteria require that waters be of sufﬁcient quality to support resident biological communities and are relevant because a number of biological communities have declined over the last several decades. However, the scientiﬁc challenge is differentiating those declines from natural population cycles and positively associating them with acidiﬁcation-related water quality stress. We present two case studies, one for pteropods and one for oysters, which illustrate the opportunities, challenges and uncertainties associated with implementing biological criteria. The biggest challenge associated with these biological assessments is lack of co-location between long-term biological and chemical monitoring, which inhibits the ability to connect biological response with an acidiﬁcation stressor. Developing new, ecologically relevant water quality criteria for acidiﬁcation and augmenting coastal water monitoring at spatio-temporal scales appropriate to those criteria would enhance opportunities for effective use of water quality regulations.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction
The ocean has absorbed more than a quarter of the carbon dioxide (CO2) emissions released into the atmosphere during the last century by burning of fossil fuels, deforestation and agricultural activities (Rhein et al., 2013). This oceanic uptake of anthropogenic CO2 lowers the pH and changes the chemical composition of seawater in a process referred to as ocean acidiﬁcation (OA) (Caldeira and Wickett, 2003; Feely et al., 2009). CO2 absorbed by seawater forms a weak acid (H2CO3), which dissociates to lessen pH, carbonate ion (CO32−) concentration, and calcium carbonate mineral saturation state, while increasing the partial pressure of CO2 (pCO2) and bicarbonate ion (HCO3−) concentration. As a result, the pH of open-ocean surface waters has decreased by about ~0.1 units and CO32− concentration has decreased about 16%. By the end of this century, surface ocean pH is expected to decline by another
0.3—0.4 units and carbonate concentration is expected to decline by 
-50% (Orr et al., 2005). Changes in the frequency of low aragonite (a
biomineral of calcium carbonate) saturation state events are also occurring (Harris et al., 2013; Hauri et al., 2013).

These acidification changes are already affecting biology in the oceans (Bednarsek et al., 2014a; Somero et al., 2016). The upwelling-dominated shoreline of the North American west coast is particularly vulnerable to OA (Feely et al., 2008, 2012; Barton et al., 2012, 2015; Hauri et al., 2013). Very nearshore waters of Oregon exhibit pH as low as 7.7 (Feely et al., 2008, 2012, 2015) due to the combined effects of atmospheric CO2 dissolution and upwelling low pH waters onshore. Bednarsek et al. (2014a) found OA hot spots along the entire coasts of Washington and Oregon, as well as northern California, where more than 50% of the upper water column in the very nearshore waters was under
saturated with respect to aragonite during the summer. Declines in oyster hatchery production in this region have been attributed to OA (Barton et al., 2012, 2015; Mabardy et al., 2015), leading to interest in identifying management levers for slowing the effects of OA (Kelly et al., 2011; Strong et al., 2014).

The US Clean Water Act (CWA) contains several mechanisms for protecting water resources, including discharge permits to ensure that water quality standards are met when a discharge is diluted in the ambient water body. However, cumulative effects of multiple permitted discharges and the presence of non-permitted sources can result in waters that violate water quality standards. When water bodies do not achieve their designated beneficial uses, they can be placed on the CWA 303(d) impaired water list. Subsequently, the Total Maximum Daily Load (TMDL) process can be initiated, whereby additional management practices or changes in permitting are employed to improve the water quality.

Given the potential for OA ecosystem effects, the Center for Biological Diversity (2013) petitioned the US Environmental Protection Agency (EPA) to list OA-impacted waters as impaired on the CWA 303(d) list. EPA responded by issuing a guidance memo recommending that states should include on their 303(d) lists those marine waters not meeting EPA-approved water quality standards for pH. The goal of this paper is to critically examine scientific issues underlying approaches to a 303(d) assessment for waters affected by OA. We describe the approaches, identify uncertainties associated with each, and indicate the science and data needs to lessen the uncertainties. While OA is a global problem, we focus on US waters where the CWA applies and on the US West Coast, as this region is particularly vulnerable to OA.

2. Water quality criteria

Water quality criteria form a quantitative basis for decisions on whether impairment exists. There are two primary types of criteria that potentially can be used to assess OA impacts: (1) numeric pH criteria and (2) narrative criteria to protect biological communities. In addition, antidegradation policies provide protections to waters that meet water quality standards but have declining water quality that is diminishing their beneficial uses. Antidegradation policy can theoretically be used for 303(d) water body listings, but there is no precedent for doing so, even for conventional pollutants. As such, the antidegradation approach is not considered further here.

2.1. pH criteria

Many states have pH criteria that are based on EPA’s Redbook recommendations, which give an acceptable range of pH for marine aquatic life of 6.5—8.5, but not more than 0.2 units outside the normally occurring range of natural variability, and a range of 6.5—9.0 for freshwater aquatic life (USEPA, 1976). Some states have modified the Redbook recommendations. Using the US West Coast as an example, Oregon does not include 0.2 unit excursions from natural conditions in their marine pH criteria, only a range of acceptable pH values (State of Oregon, 2014; Table 1). Washington adopted a narrower pH range than that in the Redbook and the range depends on the water body type under consideration (State of Washington, 2012; Table 1). California does not have a specific range of pH values for the marine criteria but instead considers whether pH deviates 0.2 units from natural conditions (State of California, 2004; Table 1). For estuarine waters, California has separate criteria for six regions of the state, which generally follow the range recommended in the Redbook but with minor modifications in each region. These modifications of the Redbook recommendations are representative of a range of modifications made by other coastal states.

States have also adopted assessment methods to determine whether or not a water quality standard is being met. For example, California requires that a minimum of five samples exceed the pH standards, but also employs a binomial statistical approach that could lead to a higher minimum. Oregon requires greater than 10 percent of the samples, and a minimum of two samples, to be outside of the appropriate criterion range. Washington requires a minimum of three samples, and at least 10 percent of values in a given year, not meet the criterion. Similar policies exist in other coastal states.

2.2. Aquatic life narrative criteria

States also have aquatic life criteria. These integrate beyond individual chemical perturbations and allow the health of the water to be assessed via the quality of the water’s biological communities (Table 2). These generally take the form of narrative criteria that are less quantitatively defined than are water quality criteria, but provide a mechanism for listing a water body as impaired via ocean acidification.

An individual state’s 303(d) listing policy defines how narrative criteria are interpreted. For example, California has an open-ended listing policy that allows development of narrative criteria evaluation guidelines if they are demonstrated to be applicable to and protective of the designated beneficial uses for that water body, linked to the pollutant under consideration, well-described, scientifically based, peer reviewed and identify a water quality range above which impacts occur and below which no or few impacts are predicted. Oregon has more prescriptive policies for using biological data to make 303(d) listing decisions, requiring use of invertebrate indices. Washington has prescriptive policies for use of invertebrates, but also has more open-ended policies that allow for use of other biological data.

3. Science challenges that arise in using a numerical pH standard for assessing compliance

Standards based solely on a percentage of samples outside of a designated range are straightforward to apply since they do not require determination of natural conditions. However, there are two challenges associated with application of this standard. The first is that the lower end of the acceptable pH range (6.5 in the EPA Redbook) is well below values known to adversely affect biological communities (Kroeker et al., 2013; Waldbusser and Salisbury, 2014; Somero et al., 2016). For example, no pH value less than 7.5 has been measured in more than five years of continuous monitoring at the Whiskey Creek oyster hatchery in Netarts Bay, even though acidification conditions there are poor enough to cause catastrophic shellfish hatchery failure (Barton et al., 2012, 2015).

The second challenge is that pH data quality can be poor for
are becoming widely accepted for open ocean and nearshore environments. The data of questionable or unknown quality. The federal court and quality assurance procedures used to collect historical data, were either not done or not done well. Moreover, the methodology recorded to the nearest 0.5 pH unit, and in other cases, calibrations were not done or not done well. Changes in normal ambient pH levels shall not exceed 0.2 units in waters with designated marine (MAR) or saline (SAL) beneficial uses nor 0.5 units within the range specified above in fresh waters with designated COLD or WARM beneficial uses.

San Francisco Regional Water Quality Control Plan. The pH shall not be depressed below 6.5 nor raised above 8.5. This encompasses the pH range usually found in waters within the basin. Controllable water quality factors shall not cause changes greater than 0.5 units in normal ambient pH levels.

Central Coast Water Quality Control Plan. For ocean waters including Monterey and Carmel Bays the pH value shall not be depressed below 7.0, nor raised above 8.5. For marine habitat, the pH value shall not be depressed below 7.0 or raised above 8.5. Changes in normal ambient pH levels shall not exceed 0.2 units.

Los Angeles Regional Water Quality Control Plan. Ambient pH levels shall not be changed more than 0.2 units from natural conditions as a result of waste discharges. Levels shall not be changed more than 0.2 units. The pH of bay or estuary waters shall not be raised above 8.6 or depressed below 7.0 as a result of controllable water quality factors; ambient pH levels shall not be changed more than 0.2 units.

Santa Ana Regional Water Quality Control Plan. The pH of bay or estuary waters shall not be raised above 8.6 or depressed below 7.0 as a result of controllable water quality factors; ambient pH levels shall not be changed more than 0.2 units.

San Diego Regional Water Quality Control Plan. For ocean waters, the pH value shall not be changed at any time more than 0.2 pH units from that which occurs naturally. In bays and estuaries the pH shall not be depressed below 7.0 nor raised above 9.0. Changes in normal ambient pH levels shall not exceed 0.2 units in waters with designated marine (MAR), or estuarine (EST), or saline (SAL) beneficial uses.

Table 1
The pH water quality criteria in the three US West Coast states.

<table>
<thead>
<tr>
<th>State of Washington</th>
<th>For water bodies classified as exceptional waters: pH must be within the range of 7.0–8.5 with a human-caused variation within the above range of less than 0.2 units. For water bodies classified as excellent or good waters: pH must be within the range of 7.0–8.5 with a human-caused variation within the above range of less than 0.5 units. For water bodies classified as fair waters: pH must be within the range of 6.5–9.0 with a human-caused variation within the above range of less than 0.5 units.</th>
</tr>
</thead>
<tbody>
<tr>
<td>State of Oregon</td>
<td>For marine waters: pH must fall between 7.0 and 8.5 For estuarine waters: pH must fall between 6.5 and 8.5</td>
</tr>
<tr>
<td>State of California</td>
<td>California Ocean Plan. The pH shall not be changed at any time more than 0.2 units from that which occurs naturally. North Coast Regional Water Quality Control Plan. Coastal waters pH shall not be changed at any time more than 0.2 units from that which occurs naturally. Changes in normal ambient pH levels shall not exceed 0.2 units in waters with designated marine (MAR) or saline (SAL) beneficial uses nor 0.5 units within the range specified above in fresh waters with designated COLD or WARM beneficial uses. San Francisco Regional Water Quality Control Plan. The pH shall not be depressed below 6.5 nor raised above 8.5. This encompasses the pH range usually found in waters within the basin. Controllable water quality factors shall not cause changes greater than 0.5 units in normal ambient pH levels. Central Coast Water Quality Control Plan. For ocean waters including Monterey and Carmel Bays the pH value shall not be depressed below 7.0, nor raised above 8.5. For marine habitat, the pH value shall not be depressed below 7.0 or raised above 8.5. Changes in normal ambient pH levels shall not exceed 0.2 units. Los Angeles Regional Water Quality Control Plan. Ambient pH levels shall not be changed more than 0.2 units from natural conditions as a result of waste discharges. Levels shall not be changed more than 0.2 units. The pH of bay or estuary waters shall not be raised above 8.6 or depressed below 7.0 as a result of controllable water quality factors; ambient pH levels shall not be changed more than 0.2 units. Santa Ana Regional Water Quality Control Plan. The pH of bay or estuary waters shall not be raised above 8.6 or depressed below 7.0 as a result of controllable water quality factors; ambient pH levels shall not be changed more than 0.2 units. San Diego Regional Water Quality Control Plan. For ocean waters, the pH value shall not be changed at any time more than 0.2 pH units from that which occurs naturally. In bays and estuaries the pH shall not be depressed below 7.0 nor raised above 9.0. Changes in normal ambient pH levels shall not exceed 0.2 units in waters with designated marine (MAR), or estuarine (EST), or saline (SAL) beneficial uses.</td>
</tr>
</tbody>
</table>

Table 2
Narrative biological quality criteria in the three US West Coast states.

<table>
<thead>
<tr>
<th>State</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oregon</td>
<td>Waters of the state must be of sufficient quality to support aquatic species without detrimental changes in the resident biological communities. OAR 340-041-0011</td>
</tr>
<tr>
<td>Washington</td>
<td>Toxic, radioactive, or deleterious material concentrations must be below those which have the potential, either singularly or cumulatively, to adversely affect characteristic water uses, cause acute or chronic conditions to the most sensitive biota dependent upon those waters, or adversely affect public health [WAC 173-201A-260(2)(a)]</td>
</tr>
<tr>
<td>California</td>
<td>Marine communities, including vertebrate, invertebrate, algae and plant species, shall not be degraded. Degradation shall be determined by comparison of the waste field and reference site(s) for characteristic species diversity, population density, contamination, growth anomalies, debility, or supplanting of normal species by undesirable plant and animal species. Degradation occurs if there are significant differences in any of three major biotic groups, namely, demersal fish, benthic invertebrates, or attached algae. Other groups may be evaluated where benthic species are not affected, or are not the only ones affected. California Ocean Plan, 2015 [II.B.E.1]</td>
</tr>
</tbody>
</table>

assessing small changes and trends. Most historical monitoring pH data have been collected with potentiometric glass electrodes, which the manufacturers claim to have precision of only ±0.2 pH units. Sensor drift can lead to measurement inaccuracies of greater than 1.0 pH units. Historical pH data have been in some cases only recorded to the nearest 0.5 pH unit, and in other cases, calibrations were either not done or not done well. Moreover, the methodology and quality assurance procedures used to collect historical data, particularly in estuaries, have not been well documented, making the data of questionable or unknown quality. The federal court recognized these data quality deficiencies in historical monitoring data when denying the Center for Biological Diversity lawsuit requesting EPA to reconsider Washington State’s decision about not using some of the historical pH data in a 303(d) assessment for OA (Center for Biological Diversity vs US EPA, 2015).

There are newer instruments and measurement techniques that provide opportunity to reduce this level of pH uncertainty. One of these is spectrophotometry, which can provide repeatable measurements within 0.001 pH units (Clayton and Byrne, 1993). These methods are laborious and costly for capturing high-frequency changes in pH, though spectrophotometric techniques have been recently adapted into automated systems for higher-frequency analysis (Martz et al., 2003; Carter et al., 2013). Another option is ion-sensitive field effect transistor (ISFET) pH sensors, which have been found to be stable and accurate for monitoring fine-scale changes in pH in open ocean environments (Martz et al. 2010) and are becoming widely accepted for open ocean and nearshore monitoring of high-frequency variability in pH (Hofmann et al., 2011; Yu et al., 2011). The management community needs to encourage development of new sensors and facilitate their adoption into regulatory-required monitoring when implementation becomes scientifically routine, as it is now becoming for ISFET sensors.

Another challenge is that pH can vary with depth, particularly in waters with density stratification and high productivity. In estuaries like Puget Sound and San Francisco Bay, lower pH values are generally found in the deeper water (Feely et al., 2010, 2012). Current pH criteria do not specify where in the water column one should make measurements and some estuarine monitoring programs, particularly continuous monitoring stations, focus on surface or near-surface sampling, which are least likely to fail pH criteria.

4. Science challenges that arise in using deviation from natural conditions to assess compliance

Standards for some states include requirements to demonstrate that pH differs from natural conditions (Table 1), but it is unclear whether the 0.2 pH unit difference defined in most standards refers to change attributable to local, regional or global inputs. The approach typically used in the 303(d) process, and which is followed in this manuscript, is to separate the impairment assessment from the TMDL solution. As such, the question is whether the water body is impaired relative to its anticipated state without any anthropogenic stressor, regardless of the spatial scale of that stressor. The determination of whether the influences of local...
sources are sufficiently large to warrant TMDL management actions would then be determined in a subsequent source attribution step.

However, the technical challenge of establishing what constitutes natural conditions remains. One approach used in 303(d) assessment for other parameters is spatial contrast, in which conditions in a reference water body are compared with those from a potentially polluted water body. Identification of reference areas is problematic for OA, though, as effects are experienced across entire regions. No coastal systems will be unaffected by global CO2 emissions and systems will likely differ naturally in the amount of dissolved inorganic carbon (DIC) present (Feely et al., 2008, 2010; 2012).

Alternatively, reference condition can be defined temporally, but this requires data sets having sufficiently long temporal history to allow anthropogenic perturbations to be distinguished from natural interannual/interdecadal fluctuations in ocean physics and chemistry. There are few high-quality data sets that have such temporal duration, particularly in coastal/estuarine waters where the natural variation is higher than in the open ocean, owing to the effects of freshwater inputs, primary production/respiration, and local atmospheric inputs (Feely et al., 2010; Cai et al., 2011; Frieder et al., 2012). Fluctuations in pH experienced over the course of a tidal cycle or a day can greatly exceed the magnitude of changes expected from anthropogenic changes and thus present a challenge in using temporal records to define ‘natural conditions.’

The Monterey Bay Aquarium Research Institute’s M1 mooring station represents the longest continuously sustained, quality-controlled OA monitoring effort in the California Current System and the best opportunity for assessment of compliance with marine pH criteria using a temporal comparison to define natural conditions. Two decades of observations indicate increasing pCO2 and decreasing pH trends that are consistent with both atmospheric CO2 increase (Fig. 1), and long-term changes in ocean carbonate chemistry from the Hawaii Ocean Time-Series data (Dore et al., 2009). However, the pH changes reflected in this data set over that time period are only about 0.1 pH unit, half of the 0.2 pH unit reduction from natural described by the EPA Redbook. Recent modeling studies suggest that the carbonate chemistry changes in the California Current will likely accelerate over the next few decades such that the 0.2 pH unit change will be exceeded before mid-century (Gruber et al., 2012; Hauri et al., 2013). The Monterey Bay M1 mooring data set will allow empirical documentation of that change. This data series also illustrates how a shorter temporal duration could provide a misleading trend.

The availability of other high-quality data for assessing changes from conditions in the past is poor. Review of historical pH measurements archived in the NOAA National Oceanographic Data Center’s World Ocean Database indicates that the data collected prior to the 1990s are typically not well documented and their metadata are incomplete, making these data of unknown quality and therefore inappropriate for 303(d) assessments. New time-series efforts are now underway. For example, along the West Coast, researchers are initiating the collection of high-quality data in California (CCE1, CCE2, M2), Oregon (Newport line), and Washington (Cape Elizabeth, Cha-Ba, Grays Harbor line, Dabob Bay, Twanoh) (Alin et al., 2015). However, these measurements are presently too short to delineate the changing acidification signal. Both the states of Washington and Oregon legislatures have recently begun investments in estuarine and nearshore acidification monitoring that will pay future dividends for such assessments. Moreover, the information gained in these local systems will have national benefits, increasing our understanding of natural pH fluctuations in coastal systems.

A third alternative is using models to define natural conditions by modeling OA chemistry in waters under a scenario with zero global emissions and zero local contributions. In the California Current, anthropogenic DIC varies over a relatively narrow range for surface waters that are in equilibration with today’s atmosphere, anthropogenic DIC will be on the order of 50–70 μmol kg–1 (Harris et al., 2013). For the deepest open-ocean waters that have been isolated from the atmosphere for >3 centuries, this anthropogenic burden will tend toward zero. Developing natural relationships between DIC and other acidification parameters provides a means for calculating what OA parameter values would be in the absence of anthropogenic CO2 emissions. This approach has been applied to assess changes in the depths of aragonite saturation horizons (Feely et al., 2008; Bednarsek et al., 2014a) as well as changes in the frequency distribution of corrosive conditions for fixed stations (Harris et al., 2013). However, this type of budgeting model alone is too conceptual to support an impairment decision because it is not possible to validate the model. Mechanistic coupled physical-biogeochemical models provide an alternative approach because these models can be validated, yielding estimates of uncertainty in predictions for managers to use in

**Fig. 1.** A) pCO2; and B) pH at the MBARI M1 Time-series Station in Monterey Bay (after Borges, 2011).
making their determination. Although more challenging to construct, mechanistic models allow for evaluation of how pH would differ if selected nutrient or carbon inputs were reduced (Howard et al., 2014). Such quantitative evaluations directly address whether anthropogenic input alters pH by more than 0.2 pH units. Such models are presently being developed (e.g., Sutula et al., 2014; Roberts et al. 2015). Further advances in biogeochemical process rates, boundary conditions and model validation would enhance their applicability to 303(d) assessments.

5. Science challenges that arise in using aquatic life use criteria to assess compliance

Narrative aquatic life criteria provide a means for assessing ecosystem condition via the health of their biological communities. A number of studies have shown marine species are sensitive to changes in seawater carbonate chemistry (Somero et al., 2016). Adverse effects on shell development are directly linked to changing carbonate ion availability and are the most well documented (Waldbusser et al., 2015b; Bednarsek et al., 2014a,b; Busch et al., 2014). Pteropods (pelagic snails), in particular, appear quite sensitive to the effects of OA, as detailed in the case study below.

Metabolic rates of the Humboldt squid have been shown to decrease when exposed to increasing pCO₂ (Rosa and Seibel, 2008).

Further, studies have focused on fish, but rising pCO₂ has been shown to affect fish through multiple pathways ranging from reduced metabolic rates, altered otolith growth, and impaired neurosensory functions and behavior (Somero et al., 2016).

There are several important considerations involved in applying aquatic life criteria to make a 303(d) impairment determination. The first is to demonstrate population-level effects on native biota. Ideally, evidence would be obtained from biological monitoring data showing that species abundance is declining, or that the size structure of the species is changing. Such data may be available in the form of catch data for economically important fisheries. For example, oyster recruitment in Willapa Bay has been monitored for decades and shows a decline over the last 30 years (see case study below). Biological monitoring data for organisms that are not typically fished are rarer. In cases where monitoring data to assess declining abundance or changing size structure are unavailable, decreasing fecundity or survival could serve as a surrogate for likely population-level effects, though long-term records for even those kinds of responses are rare.

EPA has determined that evidence for population declines as it relates to the aquatic life criteria must be for in situ natural populations, a position that has been upheld in federal court (Center for Biological Diversity vs US EPA, 2015). EPA’s rationale is that laboratory studies do not sufficiently mimic the fluctuations and multiparameter stresses of real world exposure conditions and thus cannot account for potential organism adaptation or acclimation. Similarly, EPA agreed with Oregon’s determination that reduced shellfish hatchery success is inadequate for demonstrating water quality insufficiency since the narrative criteria focus on naturally occurring biota.

A second consideration in making an OA-related listing decision is connecting the observed population-level effects to an OA stressor, which can be a parameter other than the pH criteria. This can be accomplished by correlative analysis of field data if OA chemistry data were collected simultaneously with the biotic population data or biological responses to OA stressors were linked using laboratory studies. Using the first approach is challenging, since there are few data sets where biological and OA chemistry measurements have been collected simultaneously over a large enough range of conditions to support a correlative analysis. Moreover, correlation is not proof of causation, as other natural cycles, such as changes in temperature, oxygen or rainfall may covary with carbonate chemistry. As such, the majority of studies indicating species sensitivity to OA have been conducted using laboratory experiments. Whereas EPA concluded that laboratory studies alone are insufficient to establish population-level effects, laboratory study data are an excellent tool for stressor attribution. Ideally, a combination of laboratory and field studies will provide evidence that population-level effects are likely caused by changing carbonate chemistry attributable to OA, with the pteropod case study below a good example of how both can be used together for this purpose.

A third consideration is to show that the affected species is ecologically important, which is somewhat vague since all species in a water body are linked to other species through the food web. However, studies that document OA effects on natural biota that are economically important (e.g., crab), key species in the food web (e.g., pteropods) or threatened species (e.g., abalone) will be useful to managers in determining whether a water body fails the aquatic life use assessment.

Two case studies below describe situations where a biological species appears to be experiencing stress in OA-impacted waters. The case studies illustrate the opportunities, challenges, data gaps and uncertainties associated with implementing biological criteria.

5.1. Pteropods in the California Current

Pteropods are pelagic sea snails with aragonite shells that are highly abundant on continental shelves, such as the California Current Ecosystem (CCE) (Bednarsek et al., 2012). Numerous laboratory studies have shown negative effects of OA on vital biological processes, such as shell growth, integrity and dissolution; metazoan scope; swimming ability and survival (Comeau et al., 2009; Lischka et al., 2011; Bednarsek et al., 2012, 2014b; Seibel et al., 2012; Busch et al., 2014). A principal challenge in using pteropods for assessing compliance with the aquatic life use criteria is connecting laboratory studies to population effects occurring in the ambient environment.

There are several lines of evidence, encompassing multiple levels of biological organization from physiological, individual and population level (Fig. 2; Table 3), suggesting that pteropods from the natural environment are already adversely affected by OA. Synthesized results in Fig. 2 illustrate a biological condition gradient relating these biological responses with carbonate chemistry. In supersaturated conditions (Ωar > 1), pteropods are healthy, while at near-saturation (Ωar = 1), pteropod biological conditions begin to deteriorate and indicate early warning of stress. In undersaturated conditions (Ωar < 1), pteropods experience major damage.

At the physiological and individual level, the evidence of negative OA effects in the natural environment is strong. First, field measurements indicate that the percent of Limacina helicina individuals affected by shell dissolution closely corresponds to carbonate chemistry conditions (Bednarsek et al., 2014a). Along the West Coast, more than 50% of pteropod individuals already show some extent of shell dissolution due to OA, and by 2050, >70% of individuals are predicted to be affected (Bednarsek et al., 2014a). While calcification can partially offset dissolution in supersaturated conditions, this slows or discontinues in near-saturated or undersaturated conditions (Bednarsek et al., 2014b; Fig. 2), impairing integrity of the shells. These changes coincide with reduced swimming abilities, individuals being more prone to infection and higher predator pressure (Bednarsek et al., 2014a). Second, calcification is impaired or completely absent in pteropod’s living in near- or undersaturated conditions (Bednarsek et al., 2014a). The time-line for calcification effects is similar to dissolution effects,
occurring on time scales of a few weeks and intensify during upwelling events (Fig. 2). The changes on the physiological level result in reduced growth rates leading to bioenergetic and fecundity impairment, and reduced survival (Table 3).

The thresholds for physiological processes, such as calcification and dissolution, correspond with those for individual survival probability (Fig. 2). The changes in survival in these experiments can be attributed to previous exposure to undersaturated conditions in their natural environment and demonstrate loss of resilience. These changes on the individual level are precursors to long-term population effects, such as abundance, size-frequency distribution, reproductive integrity and food-web alterations (Bednaršek et al., 2014a).

At the next level of biological organization (Table 3), there have been changes in pteropod vertical distribution correlated with aragonite saturation state (Bednaršek and Ohman, 2015). The behavior of some pteropod species indicates an avoidance of undersaturated waters and a tendency of occupying shallower, supersaturated waters. While changes in vertical distribution are not full evidence of a population effect, they can potentially alter food-web dynamics and trophic-level relationships (Table 3), potentially affecting the vertical distribution of pteropod immediate predators, as well as pelagic and demersal fish (Lalli and Gilmer, 1989).

Making the extension to changes in population size is more tenuous because of the incomplete understanding of natural cycles and limited length of the temporal record, often confounded with additional environmental drivers. There is some evidence for declining Limacina helicina abundance observable in the northern coastal CCE subregions (Mackas and Galbraith, 2012). The decline was most pronounced in the summer and autumn months, closely corresponding with the timing of the most corrosive conditions in the nearshore areas. Long-term trends (1990–2010) show significant decline along the continental margin as well as nearshore regions where saturation state has decreased substantially. This contrasts with a lack of population decline in offshore regions (Mackas and Galbraith, 2012) where pteropod habitat was more suitable (Bednaršek et al., 2014a).

The assessment process involves establishing a relationship between the affected biological population and the health of the broader biological community. For pteropods, a close link has been established with higher trophic levels. Pteropods have been identified as an important food source for a variety of fish species, like cod, mackerel, sablefish, clupeids, gadoids, and salmon, contributing up to 50–60% of fish food diet (Emmett et al., 1986; Brodeur

<table>
<thead>
<tr>
<th>Shell dissolution</th>
<th>Species richness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcification</td>
<td>Vertical distribution</td>
</tr>
<tr>
<td>Respiration</td>
<td>Size and age distribution</td>
</tr>
<tr>
<td>Survival impact</td>
<td>Bioenergetics Parameters</td>
</tr>
</tbody>
</table>

**Table 3**

Representative bioindicators measured at three major levels of biological organizations. Each of the three levels represent increasing levels of biological organization, decreasing level of response sensitivity and increasing ecological relevance.

<table>
<thead>
<tr>
<th>Physiological/Individual (Rapid response)</th>
<th>Behavior (Weekly to monthly response)</th>
<th>Population (Ecologically relevant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shell dissolution</td>
<td>Species richness</td>
<td>Abundance</td>
</tr>
<tr>
<td>Calcification</td>
<td>Vertical distribution</td>
<td>Size and age distribution</td>
</tr>
<tr>
<td>Respiration</td>
<td></td>
<td>Loss of Resilience</td>
</tr>
<tr>
<td>Survival impact</td>
<td></td>
<td>Bioenergetics Parameters</td>
</tr>
</tbody>
</table>
and Peirce, 1990; Aydin et al., 2005; Armstrong et al., 2008). A strong correlation has also been demonstrated between pteropod abundance and pink salmon survival (Doubleday and Hopcroft, 2015).

Another challenge is that supporting data for 303(d) determinations must be collected within state waters. Pteropods span nearshore and offshore regions, with only a small portion of their population occurring within the 3–mile state jurisdiction for impaired waters decisions. Half of the pteropod findings presented here derive from research cruises outside of state waters, with the other half in the coastal nearshore waters. From a scientific perspective, these data represent a comprehensive description of a population that likely extends beyond the state border. However, from a regulatory perspective, documentation of effects in offshore federal waters may not apply to the same population found within state waters.

5.2. Willapa Bay Pacific oyster case study

Pacific oysters (Crassostrea gigas) have been the dominant aquaculture product from Willapa Bay, Washington, since their 1928 introduction from Japan, following commercial over-exploitation of the native oyster (Kincaid, 1968; Ruesink et al., 2006). Commercial interests established a long-term record of the density of newly-settled Pacific oysters (Dumbauld et al., 2011), with quantitative monitoring initiated in 1942 (Chapman and Esvedt, 1943) and a continuous qualitative record since 1936 (Fig. 3). Numbers of settling oysters over the last decade have been low, and Willapa Bay waters are near or below carbonate chemistry thresholds for larval development (Burke Hales, pers. comm.).

However, there are three challenges in using this record of population decline for making a 303(d) impairment determination using the aquatic life use criteria. The first challenge is a legal one. While the Pacific oyster was introduced Willapa Bay nearly 100 years ago and naturally reproduces, it might still be considered a non-native species and state standards focus impairment decisions on indigenous species (Washington Administrative Code 173–201A).

The second challenge is differentiating a trend in Pacific oyster recruitment from natural variability. Over the past three decades, the frequency of commercial sets has declined (logistic regression, N = 30, P = 0.03), but there is no significant trend when examining the entire 79-year time series (P = 0.48, Fig. 3).

The third difficulty is attributing causality to recruitment variability. Recent high-quality chemical data show that Willapa Bay waters are often close to carbonate-saturation-state levels that negatively affect newly-spawned larvae in laboratory experiments (Barton et al., 2012, 2015; Waldbusser and Salisbury, 2014; Waldbusser et al., 2015b). Willapa Bay is located along an upwelling coast and is affected by shoaling of the carbonate saturation horizon (Feely et al., 2008), with pH fluctuations at the mouth of Willapa Bay mirroring offshore upwelling (Ruesink et al., 2015). However, oyster recruitment occurs in the upper estuary where water residence time exceeds a month (Banás et al., 2007). The key driver of carbonate saturation state is the ratio of alkalinity to total dissolved carbon (Alk:DIC), followed by the total abundance of these parameters. The former is strongly related to salinity, while the latter is also linked to the local and distal metabolism signals that drive consumption and release of DIC. Willapa Bay is subject to several water sources that have unfavorable carbonate chemistry—freshwater sources, whether from precipitation, local rivers or the Columbia River plume are all low in Alk:DIC, as is the highly saline upwelled source water (Feely et al., 2008).

Thus, in the absence of net photosynthetic manipulation of the Alk:DIC ratio, either by in-bay or adjacent coastal-ocean net community productivity, Willapa Bay waters will be persistently near or below identified carbonate saturation thresholds (Waldbusser et al. 2015b). Carbonate saturation conditions improve in Willapa Bay through summer oceanic upwelling; bay and coastal ocean productivity increase, and bay salinity (as well as Alk and DIC) rises during this low-inflow season. Because the sensitivity of Pacific oyster larvae is to concentration of the carbonate ion rather than pH or CO2 concentration (Barton et al., 2012, 2015; Waldbusser et al. 2015a, b), and pH itself is a poor predictor of carbonate-mineral saturation, other high-quality chemical data are necessary to define environmental exposure for the oysters. Surrogates that are available as potential indicators of carbonate chemistry (e.g., river flow and upwelling) do not exhibit a simple linkage with carbonate-mineral saturation state in the key recruitment areas of the Bay. Further, the pH measurement itself is plagued by historical data quality issues that have only been resolved by the oceanographic community in the last 1–2 decades and have not been satisfactorily addressed in the estuarine setting. Therefore, historical pH data, which are available from oceanographic cruises in Willapa Bay in the 1950s/60s, are not sufficiently accurate to reconstruct carbonate chemistry experienced by organisms at the time—and thus unsuitable for demonstrating the environmental cause of population change.

Attribution of recruitment declines to acidification is also confounded by the presence of other factors, such as thermal or hypoxic stressors. Temperature-sensitivity is marked in the reproductive biology of Pacific oysters, given spawning at 18 °C (Bourlès et al., 2009) and accelerated larval development and improved metamorphosis over the range 17–27 °C (Rico-Villa et al., 2009; Ben Khaled et al., 2010). Typical summer water temperatures in Willapa Bay are at the low end of this range (17–19 °C); recruitment has been more reliable in two other west coast regions where temperature typically exceeds 19 °C (Hood Canal in Washington and Pendrell Sound in British Columbia). Hales (pers. comm.) found that the modern-day thermal optima are poorly synchronized with modern-day carbonate-mineral saturation optima, and that this asynchrony is worsened by rising atmospheric CO2.

Biological interactions can also serve to confound interpretation, as competition from Manila clams (Ruditapes philippinarum) could be suppressing Pacific oyster recruitment. Manila clam production has increased by an order of magnitude in Willapa Bay over the last several decades (Ruesink et al., 2013). Larvae of this introduced species were rarely reported in historic samples but now frequently exceed 1000 per 44 L, co-occurring with Pacific oysters (Ruesink et al., 2013) and possibly increasing competition for food.

This case study emphasizes the importance for having both long-term biological records and co-location of high-quality

---

**Fig. 3.** Number of annual sets (oyster larvae that settle on substrate) for Pacific oysters in Willapa Bay, Washington, by decade (following Dumbauld et al., 2011).
chemical data. At one level, a decline in Pacific oysters accompanied by low carbonate-mineral saturation state conditions provides circumstantial evidence for making an acidification-related impairment assessment in Willapa Bay. However, when examined in context of fluctuations in the longer historical record, with the absence of chemical data that are truly co-located and of sufficient temporal duration to establish a correlative causal relationship, the case for making an impairment assessment becomes less clear. One option for addressing causality would be to expand the investigations to assess larval survival of cohorts spawned at times or places with different carbonate saturation states and placing those events into the context of laboratory (Waldbusser et al., 2015a, b) and hatchery studies (Barton et al., 2012).

6. Discussion

There are two primary approaches for applying water quality criteria toward making 303(d) assessment for waters impacted by OA: 1) assessments based on pH, and 2) assessments of aquatic life effects. Significant scientific and technical challenges arise in implementing these approaches, which could be improved through adoption of more scientifically appropriate chemical criteria and improved monitoring on which to base both biological and chemical assessments.

6.1. Contemporaneous monitoring of biological and chemical data would improve use of narrative criteria for aquatic life assessments

In both case studies, attribution of population-level OA effects was found to be hampered by scarcity of acidification-relevant stressor data co-located with the long-term biological monitoring programs; something that has been called for by a number of acidification monitoring planning efforts (Newton et al., 2014; Alin et al., 2015; McLaughlin et al., 2015). The few biological monitoring programs simultaneously measuring chemical and biological parameters often do so with older glass-electrode potentiometric probes and would benefit from upgrades to newer, more accurate technology (Feely et al., 2008, 2010; Byrne, 2014; Martz et al., 2015) and use of standardized best practices for measurement of carbonate chemistry parameters (Dickson et al., 2007; Bresnahan et al., 2014). Moreover, several important long-term monitoring programs are conducted primarily outside of state waters and would be more relevant to 303(d) decisions if they were extended closer to shore. While there are financial challenges associated with upgrading and maintaining long-term monitoring, such monitoring is needed to formulate a coherent management response to OA impacts on marine resources, of which 303(d) determinations are a component. Strengthening monitoring programs should include measurement of OA variables beyond pH, address depth variation and ensure a temporal resolution that enables assessment of the mechanisms driving variation.

6.2. Changes to the existing pH criteria are necessary if they are to reflect conditions that protect aquatic life

While basing water quality assessments on biological responses is a good path forward, it should not be relied on entirely because detection of population-level effects may occur too late for corrective actions for some species or food webs. However, the existing pH criteria do not reflect the current state of the science; water bodies unfit for aquatic life may still meet the pH criteria. The existing criteria allow a deviation of 0.2 pH units from natural conditions and a lower limit of 6.5, which only appear relevant to protect against local effects caused by discharge of acidic water from a specific effluent outfall, for which they were originally developed. However, these criteria are scientifically inappropriate for assessing a regional effect mediated by global-scale processes (Byrne et al., 2010). Numerous studies have shown diverse biological effects routinely manifest at pH levels well above pH of 7.5, an order of magnitude higher (since the pH scale is logarithmic) than the existing criteria of 6.5 (Barton et al., 2012; Waldbusser and Salisbury, 2014; Bednarsek et al., 2014a,b; Somero et al., 2016). Moreover, the requirement to demonstrate a 0.2 pH unit deviation from natural is impractical, as the reference condition cannot be established spatially when the entire region is undergoing change and is difficult to accomplish temporally because there are few long-term data sets collected with enough precision and accuracy to capture that level of change.

6.3. New chemical criteria are needed and should be expanded to include other acidification parameters

pH is only one of several possible parameters for describing the aquatic carbon system, and it is unclear that pH is the measure most relevant to biological response (Somero et al., 2016). Adding alternative carbon system parameters should also be considered when developing new criteria. Aragonite saturation state is one such candidate, having been found to be more biologically relevant than pH for a number of calcifier groups (Doney et al., 2012; Waldbusser et al., 2015a; Breitburg et al., 2015). There is growing scientific evidence, particularly from the Netarts Bay oyster survival studies (Barton et al., 2015) and from the pteropod case study above, to begin establishing both chronic and acute thresholds for this parameter. In addition, carbonate system parameters such as pCO2 have been found to be biologically relevant for biota that do not have shells (Kroeker et al., 2013; Waldbusser et al. 2015b). The challenge to both scientists and managers will be to assess not only which of those carbon chemistry parameters are sufficiently important that they have the potential to affect population success and should elevate to be part of a chemical criteria, but also how to craft criteria so as to be implementable with existing measurement technology. In the immediate future, a scientific workshop is needed to identify appropriate biologically relevant indicators and thresholds to assess OA and short-term research needs for informing criteria and thresholds.

6.4. Development of ecologically relevant criteria will support management of acidification

Criteria are used by water quality managers to make regulatory decisions, but their value extends well beyond that. The parameters included in criteria, and the associated thresholds, become part of a shared benchmark that drive discussions about appropriate management. Criteria provide context for interpreting monitoring trends and provide perspective for assessing the benefits of potential management actions when examining the output from biogeochemical models. Their benefits also extend beyond the water quality management community. An example of this is the use of dissolved oxygen as an endpoint to protect survival and reproduction of fisheries, developed by EPA-supported research in the 1970s and adopted by federal and state water quality agencies in the 1980s. These criteria are used almost universally as biologically relevant endpoints in non-regulatory, estuarine and marine natural resource and habitat assessments (Bricker et al., 2003; Best et al., 2007; Zaldívar et al., 2008). Criteria can also be used as interpretable thresholds for communicating with the public about OA and its effects on local waters.
6.5. While refining the criteria and improving monitoring will enhance the ability for making 303(d) assessments, a TMDL process may prove challenging

A 303(d) listing usually leads to a TMDL process. The first step of a TMDL would be to quantify the various nutrient and carbon loads that contribute to the OA signal. However, our ability to complete this first step is lacking as the modeling necessary to achieve this, and the underlying data to support such models, vary considerably (Boehm et al., 2015). Existing coastal-scale models require validation and further refinement, downscaling and bringing their spatial domains closer to shore at temporal and spatial scales relevant to management decisions. They also need to be better linked with ecosystem effects through both empirical and numerical ecosystem models. Local-scale models are also important. Investments are needed in organizing a modeling forum to summarize the status of existing data and models that would support those needs (Sutula et al., 2014). This would be a wise investment in acidification management regardless of whether it is initiated through a TMDL process, because these models can also serve to aid in marine spatial planning linked to fisheries and reserve management, as well as mariculture applications, by identifying areas that are most/least susceptible to acidification.

Global atmospheric inputs will be the main contributor to OA in most settings, particularly for oceanic sites. The Clean Water Act does not provide specific authority to limit atmospheric inputs, but there are precedents for atmospheric reductions being identified in TMDLs. For example, the nutrient Chesapeake Bay TMDL quantified the amount of nutrients coming from air sources and identified programs that could be engaged to reduce these inputs (Linker et al., 2013). Additionally, TMDLs have been developed for water bodies impaired by mercury and acid rain, which are caused primarily by atmospheric inputs.

6.6. OA requires a cooperative initiative where multiple jurisdictions work collaboratively toward regional scientific tools and assessments

OA and the organisms it affects do not respect state boundaries, so actions may need to be regional or national. This could start with a multi-state regional assessment of water body impairment with respect to OA, but also extend to development of regional management models. Nutrient management in Chesapeake Bay has been handled regionally and provides an example of how this might be done. Large portions of the inputs that were causing hypoxia flowed from watersheds in states upstream of the Bay (Beegle, 2013). This was compounded by atmospheric inputs from states in the Midwest, which contributed one-third of total nitrogen to the Bay (Fischer and Oppenheimer, 1991). In response, and in part because there was a strong scientific foundation for quantifying relative contributions among sources through physical/biogeochemical modeling, an innovative watershed agreement was forged to accelerate the pace of restoration and align federal directives with state and local goals (Linker et al., 2013). This solution is more challenging for acidification because the atmospheric sources are generated from an even larger geographic area that could include other nations, but similar leadership to address regional/national sources will likely be required to address acidification.

Acknowledgments

We thank Terry Fleming for his assistance as we prepared the manuscript. We also thank Miyoko Sakashita, Brian Rappoli, Kelly Susewind, Catherine Kuhlman, Liz Chornesky, Ruth Chemerys, Sara Shen and Mindy Roberts for their helpful suggestions on the draft manuscript. This is PMEL Contribution 4396.

References


