Estimating the probability of exceeding elevated pH values critical to fish populations in a hypereutrophic lake

Jacob Kann and Val H. Smith

Abstract: Current eutrophication models typically are used to predict seasonal mean conditions. However, the risk of summer fish kills in hypereutrophic lakes is likely to be more closely dependent on periodic extreme events, such as potentially lethal peaks in pH driven by algal photosynthesis. In hypereutrophic Upper Klamath and Agency lakes, Oregon, peak summertime pH values frequently exceed critical levels that can reduce fish growth and survival (pH > 9.50, a likely sublethal tolerance limit for two resident endangered fish species). We developed two empirical models, one parametric and one nonparametric, that predict the likelihood of exceeding user-defined critical values of pH from concentrations of chlorophyll *a* in these lakes. Separate models were derived to incorporate seasonal dynamics and differences between the two lakes, and the behavior of these models was tested under four different critical pH scenarios. Both parametric and nonparametric models performed similarly, suggesting that management efforts to reduce chlorophyll *a* in these lakes from 200 to 100 $\mu g \cdot L^{-1}$ should decrease the probability of exceeding pH 9.5 by 45%. We suggest that this general approach potentially can be applied to the management of fish populations in other hypereutrophic lakes as well.

Résumé : Les modèles d'eutrophisation actuels sont généralement utilisés pour prévoir les conditions saisonnières moyennes. Par ailleurs, le risque d'hécatombes de poissons dans les lacs hypereutrophes en été est sans doute plus dépendant d'événements périodiques extrêmes (comme les pics de pH dus à la photosynthèse des algues et pouvant s'avérer mortels). Dans les lacs hypereutrophes Upper Klamath et Agency de l'Orégon, les valeurs pics du pH en été excèdent fréquemment les valeurs critiques pouvant réduire la croissance des poissons et nuire à leur survie (pH > 9,50, une valeur limite sans doute sublétale pour deux espèces de poisson de ces lacs). Nous avons élaboré deux modèles empiriques, l'un paramétrique et l'autre non paramétrique, afin de prévoir la probabilité de dépasser les valeurs de pH critiques définies en fonction de l'utilisateur, à partir des concentrations de chlorophylle *a* notées dans ces lacs. Des modèles distincts ont été obtenus afin d'inclure les dynamiques saisonnières et les écarts entre les deux lacs, et le comportement des modèles a été testé dans le cadre de quatre scénarios différents de pH critique. Les modèles paramétriques et non paramétriques ont donné des résultats semblables, ce qui porte à croire que les mesures de gestion visant à réduire les concentrations de chlorophylle *a* de ces lacs de 200 à 100 $\mu g \cdot L^{-1}$ devraient donner lieu à une diminution de 45% de la probabilité que le pH soit supérieur à 9,5. Nous croyons que cette démarche générale pourrait être appliquée à la gestion des populations de poisson d'autres lacs hypereutrophes.

[Traduit par la Rédaction]

Introduction

Summer algal blooms are a consistent feature of eutrophic and hypereutrophic lakes worldwide (Smith 1998). These algal blooms dramatically reduce water clarity and frequently cause taste, odor, and filtration problems in water supplies (Cooke et al. 1993). In many such lakes, poor water quality causes fish kills that can be costly (American Fisheries Society 1992) and can result in a complete restructuring of the fish community (Vanni et al. 1990). The development of an objective plan that minimizes eutrophication-related risks to fish populations in lakes requires the use of a quantitative framework that (*i*) relates fish populations to key water quality variables of concern and (*ii*) subsequently relates these key variables to manageable watershed or lake characteristics.

Quantitative approaches to the effects of water quality on fish have previously been proposed by Barica (1975) and Mericas and Malone (1984), who developed water chemistry based fish kill models. Unfortunately, the utility of these models is limited because historical records of fish kills are typically poor due to inconsistent reporting, rapid predation by avian and fish predators, settling of dead fish to benthic areas, and the difficulty in observing mortality in larval and juvenile fish. Furthermore, models based solely on fish kill probability do not account for deleterious sublethal or chronic population-level effects. An alternative approach presented here is based on (*i*) documented physiological tolerance limits for fish and (*ii*) the probability of exceeding these critical limits.

Accomplishment of (*ii*) above requires a modification of

Received January 15, 1999. Accepted August 5, 1999.
J14975

J. Kann.[1] Aquatic Ecosystem Sciences, 232 Nutley Street, Ashland, OR 97520, U.S.A.
V.H. Smith. Department of Ecology and Evolutionary Biology, Haworth Hall, University of Kansas, Lawrence, KS 66045, U.S.A.

[1]Author to whom all correspondence should be addressed. e-mail: jacobkann@aol.com

traditional approaches that utilize regression-based models to relate seasonal means of a key water quality variable to its putative controlling factors (e.g., Dillon and Rigler 1974). The use of seasonal means may not be adequate when evaluating the impact of lethal or sublethal stressors, for which the frequency of extreme events is likely to be a better predictor of fish growth and survival than a simple seasonal mean. In the case of dissolved oxygen stress, Oregon Department of Environmental Quality (1994) suggested that repeated exposure to concentrations near acute levels or to chronic stress levels poses serious risks to the aquatic community. Predicting the frequency of exceeding critical tolerance limits thus may provide a valuable alternative approach to assessing eutrophication-related risk to fish populations.

Upper Klamath and Agency lakes, a system of connected lakes located in south-central Oregon, are an example of ecosystems whose fish populations have been strongly impacted by eutrophication. The shortnose sucker (*Chasmistes brevirostris*) and Lost River sucker (*Deltistes luxatus*) (Catostomidae) are endemic to the Upper Klamath River Basin but are now near extinction. Both species were listed as endangered, and water quality degradation resulting from algal blooms has been identified as a probable major factor in their declines (Williams 1988). Both species are predominantly adfluvial (spawning takes place in tributaries with rapid emigration of larvae to the lake upon hatching) and are long-lived (>25 years) and depend heavily on satisfactory lake water quality for long-term population viability. Adverse effects of eutrophication on native fishes in Upper Klamath Lake include summer fish kills (Scoppettone and Vinyard 1991) as well as a pronounced horizontal redistribution of the fish populations (Hazel 1969). Massive blooms of *Aphanizomenon flos-aquae* now occur nearly continuously from June through October (Kann 1998), causing dramatic variations in dissolved oxygen concentrations and pH. Summertime pH values typically exceed 9.50, a likely sublethal limit for the two resident fish species, and periodically exceed 10. These water quality conditions are of great concern because pH values >9.55 caused a loss of equilibrium in native *Chasmistes brevirostris* juveniles (Falter and Cech 1991). Similarly, short-term acute lethality tests (96-h LC_{50}s) indicate that mortality of juvenile *Deltistes luxatus* occurs when pH values >9.94 are reached (Saiki et al. 1999). Elevated pH is of further concern to fish populations as the frequency of sublethal and lethal concentrations of un-ionized ammonia increases with increasing pH in Upper Klamath and Agency lakes.

Based on the above evidence, pH is considered to be an important factor regulating the growth and survival of resident fish populations in Klamath and Agency lakes and is thus a major focus of water quality concern. We present here two empirical models, one parametric and one nonparametric, that can be used to predict the frequency or likelihood of exceeding critical values of photosynthetically elevated pH from chlorophyll *a* (Chl) in these lakes, and that we suggest can be used to help guide the future management of the resident fish populations in Klamath and Agency lakes. The first model uses a linear regression approach to predict the probability of exceeding defined physiological limits for fish growth or survival. The second model is based on a nonparametric cross-tabulation procedure that has been used successfully in other aspects of eutrophication modeling by Heiskary and Walker (1988) and Walker and Havens (1995).

Materials and methods

Data collection and analysis

Limnological measurements were made in Upper Klamath and Agency lakes from 1990 to 1996 as part of a long-term eutrophication monitoring program. Samples were taken biweekly or occasionally weekly during June–September and at monthly intervals (except when prevented by ice cover) for the remainder of the year; in general, 6–10 sites were sampled on each date. Depth profiles of pH, temperature, and oxygen were measured using a Hydrolab Surveyor multiparameter probe. Because the lake is polymictic and only undergoes weak and intermittent stratification, a depth-integrated water sample was taken on each date by combining a minimum of three replicate hauls of a weighted 5-cm-diameter plastic tube at each site. A portion of this sample was then analyzed for Chl and phaeophytin (Phae) (Nusch 1980).

The first 5 years of data (1990–1994) were used in model development, and the last 2 years (1995 and 1996) were used in model verification. Very strict care was taken to evaluate the data for conformity to the statistical requirements of regression analysis. The assumptions of linear regression were evaluated using residual and normal probability plots, autocorrelation of residuals, estimates of skewness and kurtosis, and Kolmogorov–Smirnov tests of fit (Kleinbaum et al. 1988; Wilkinson 1996). Chl data were transformed to their common logarithms to improve linearity and to reduce heteroscedasticity.

Regression-based probability modeling

Linear regression analyses were performed between lake-wide mean pH and Chl values using two different data subsets: (*i*) June–September lake-wide means and (*ii*) June-only lake-wide means. Only data from open-water sites in Upper Klamath Lake were used to compute lake-wide means; a separate regression analysis was performed for Agency Lake because the shallower Agency Lake typically exhibits more extreme pH and Chl fluctuations as well as differences in bloom timing (Kann 1998).

Regression modeling was also used to develop probabilistic models. When the assumptions of linear regression are met, one can assume that the distribution of the dependent variable at fixed values of the independent variable is normally distributed (Kleinbaum et al. 1988). The model's predictions can then be re-expressed as the probability of exceeding a defined value of the dependent variable (in this case, a defined critical concentration of pH). This was accomplished here by (*i*) predicting the mean value of pH (pH_{pred}) at specified Chl values from the linear regression equations, (*ii*) computing the standard error of the estimate ($S_{y|x}$, as estimated from the square root of the error or residual mean sum of squares; Kleinbaum et al. 1988), (*iii*) normalizing a chosen critical pH value (pH_{crit}, based on bioassay data) to its standard score (Z; cf. Zar 1984) using the following equation:

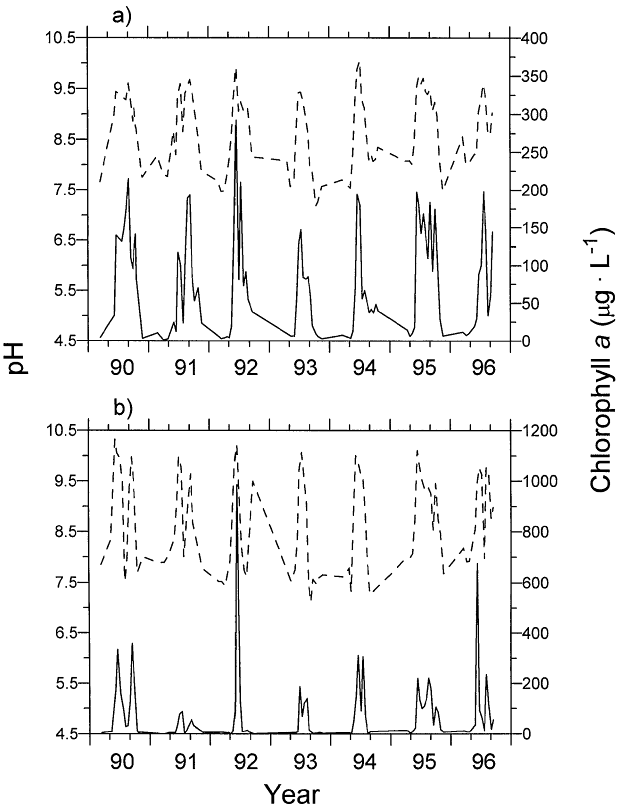
$$(1) \qquad Z = \frac{pH_{crit} - pH_{pred}}{S_{y|x}}$$

and then (*iv*) computing the proportion (expressed as a percentage) of the normal distribution that lies beyond the computed Z value:

$$(2) \qquad \text{Probability of pH} > pH_{crit} = 100(1 - \text{zcf}(Z))$$

where zcf denotes the standard normal cumulative distribution function (Wilkinson 1996). This procedure allowed us to calculate the probability of obtaining values greater than pH_{crit} for the observed range of Chl data.

Fig. 1. Seasonal trends in lake-wide mean pH (broken line) and Chl (solid line) in (*a*) Upper Klamath Lake and (*b*) Agency Lake, 1990–1996.



Nonparametric cross-tabulation probability modeling

The same subset of pH and Chl measurements used to develop the above models was used in development of a nonparametric model. However, instead of using lake-wide means, this analysis was performed on simultaneous paired pH and Chl measurements from individual lake sampling sites and dates. Following Heiskary and Walker (1988) and Walker and Havens (1995), the data first were ordered by ascending Chl concentration, divided into a chosen number of data intervals, and the median Chl for each interval computed. The frequency of pH observations within each interval that exceeded the chosen pH_{crit} value was computed, and this computed exceedence frequency (expressed here as a percentage of the total observations within each interval) was then plotted against the median Chl value for each interval.

In order to quantitatively describe the relationship expressed by this plot, and to make the model more useful for predictive purposes, a logistic equation was used to describe the relationship between exceedence frequency of chosen values of pH_{crit} and Chl. Equation parameters for the logistic model were estimated using Systat nonlinear regression procedures (Wilkinson 1996) using the following equation:

$$\text{(3)} \qquad \text{Frequency of pH} > pH_{crit} = \frac{pH_{\infty}}{1 + e^{(\beta_0 - (\beta_1 \text{Chl}))}}$$

where pH_{∞} is the asymptotic limit for pH_{crit}, β_0 is a location parameter for the curve, β_1 is a slope parameter, and Chl is the median Chl value of an interval. The fit was evaluated using residual plots, standard errors of the parameter estimates, and R^2 values.

Establishment of pH_{crit}

An important step in the development of both of the models outlined above is the choice of critical values of the dependent variable (in this case, pH_{crit}). Although excursions of pH beyond these critical limits only suggest a strong likelihood that a negative fish response will occur, avoidance of high-pH water masses would be a minimal behavioral response. As a result, stressed fish populations are often restricted to limited areas of tolerable water quality where food limitation and crowding may lead to negative effects on fish success, as reflected in reduced survivorship, growth, and reproduction (Beitinger 1990).

Bioassays using the two endangered fishes of the Upper Klamath system have revealed large interspecific and life stage dependent variability in short-term 96-h LC_{50} tests (Saiki et al. 1999). However, the most conservative value from these bioassays (pH = 9.94) was the lower 95% confidence limit for juvenile *Deltistes luxatus* (Saiki et al. 1999). When coupled with the pH 9.55 loss of equilibrium pH threshold observed by Falter and Cech (1991), these data suggest that pH 9.50 is a likely protective level that would encompass sublethal effects as well. These values are consistent with other studies that concluded that pH values between 9 and 10 are harmful to many fish species and that pH values >10.0 are typically lethal (e.g., Alabaster and Lloyd 1980).

In addition, behavioral, ontogenetic, and sublethal factors, along with interactions with other environmental factors such as ammonia and dissolved oxygen availability, can reduce fish success at critical pH levels lower than those derived from acute lethality tests (e.g., Bergerhouse 1992). Moreover, the fraction of total ammonia that remains in the toxic un-ionized form increases from 28 to 55% between pH 9.0 and pH 9.5 at 20°C. It was therefore decided to apply four separate values for pH_{crit} (9.00, 9.50, 9.75, and 10.00) in the analyses in order to investigate the models' behavior under a range of differing critical pH values.

Results and discussion

Simple linear regression analyses

A strong degree of covariation was evident between pH and Chl concentrations in Upper Klamath and Agency lakes when paired pH and Chl measurements from all sites and dates were plotted (Fig. 1). However, because regression statistics are biased when correlations exist between observations that are proximate in time or space, an evaluation of serial and spatial autocorrelation statistics is necessary when working with multiyear and or multisite data from a particular lake (Reckhow and Chapra 1983). In the case of Upper Klamath and Agency lakes, these measures indicated that utilizing paired pH and Chl measurements from all sites and dates was inappropriate (data not shown). It was therefore decided to develop the regression-based models utilizing only biweekly lake-wide means. Moreover, because the timing and magnitude of pH and Chl values in Agency Lake differed from those at the Upper Klamath Lake sites, the computation of lake-wide means was best achieved by performing separate analyses for the two lakes. Further analysis indicated that due to seasonal variance and the presence of outliers and extreme values in Chl (Kann 1998), these data were generally lognormally distributed so that the geometric mean provided the best estimate of lake-wide central tendency. When the goal of regression models is to make predictive or probabilistic statements regarding future conditions, it is necessary to both accurately reflect central tendency when combining spatial or temporal data for model input and ensure that the basic assumptions of linear regression

Table 1. Summary statistics for regressions of pH on Chl in Upper Klamath and Agency lakes.

	Parameter		Regression		
Model	β_0	β_1	$S_{y\mid x}$	r^2	n
Lake-wide means, June–September dates	6.396	1.398	0.317	0.72	43
Lake-wide means, June dates only	6.809	1.260	0.165	0.95	11
Agency Lake, June–September dates	6.859	1.265	0.533	0.73	37

Note: The regression model is $pH_{pred} = \beta_0 + \beta_1(\log \text{Chl})$; all regressions and parameter estimates were significant at $P <$ 0.0001. β_0 is the intercept, β_1 is the slope, $S_{y\mid x}$ is the standard error of the estimate, and r^2 is the coefficient of determination.

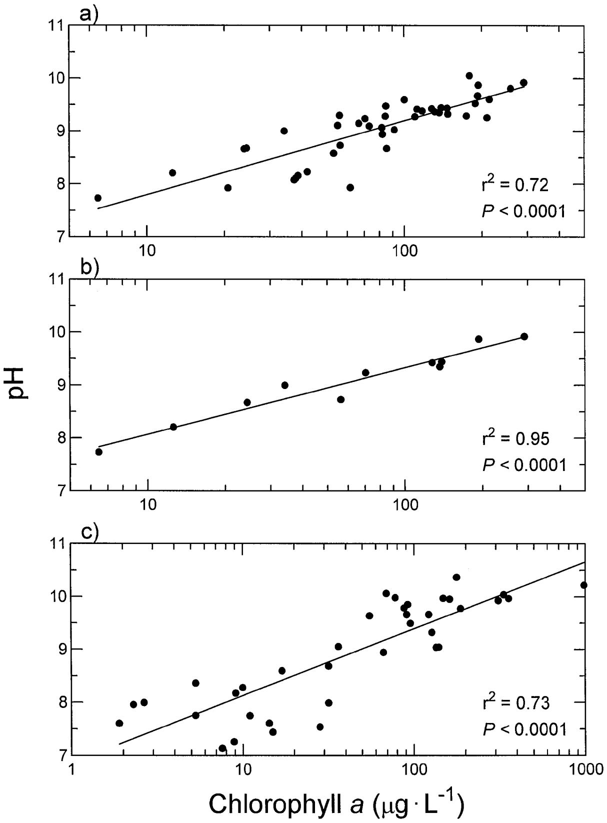


Fig. 2. Lake-wide mean pH versus Chl (log scale) during (*a*) June–September and (*b*) June only for Upper Klamath Lake and during (*c*) June–September for Agency Lake, 1990–1994. Regression lines are computed from parameters in Table 1.

are not violated. For example, all estimates of uncertainty and probability partially rely on the computation of $S_{y\mid x}$, which could be underestimated considerably due to sample size (n) inflation when autocorrelation is significant.

Despite the controlling effect of time of day that the sample was collected, temperature, photoinhibition, self-shading, solar irradiance (intensity and duration), and physiological state of the bloom on photosynthesis (e.g., Reynolds 1984), and therefore on subsequent elevated pH, Chl was still found to be a strong predictor of pH for all linear regression models (Table 1; Fig. 2). Seventy-two and 73% of the observed variation in pH during June–September could be explained by Chl for the lake-wide Upper Klamath Lake and Agency Lake models, respectively. A third model developed from June lake-wide means explained 95% of the observed variability in pH; despite a lower sample size (n = 11), there was considerably less error in the model estimate ($S_{y\mid x}$ = 0.165) for this model, when compared with the two models developed from June–September data ($S_{y\mid x}$ = 0.317). We conclude that this difference likely results from the fact that the June–September period encompasses a wide range of bloom conditions, including rapid growth and bloom collapse, and the photosynthetic elevation of pH produced per unit Chl would be expected to vary strongly with these changing physiological conditions. In contrast, June is a period of rapid algal growth and would not be affected by variability in photosynthetic performance associated with bloom collapses or senescence.

Because the concentration of Phae increases when Chl degradation occurs, it was hypothesized that the ratio of Phae to Chl (Phae/Chl) would provide a relative index of algal physiological state and could be used to account for some of the variability in pH unexplained by the simple regression models in Table 1. The addition of Phae/Chl ratio to the lake-wide June–September model produced a new predictive equation of the form

$$\text{(4)} \qquad pH_{pred} = 6.800 + 1.277(\log \text{Chl}) - 0.615(\text{Phae/Chl}); \quad R^2 = 0.79,\ S_{y\mid x} = 0.282,\ P < 0.0001.$$

This new model explained an additional 7% of the variability in pH, and $S_{y\mid x}$ was reduced by about 11% when compared with the June–September lake-wide regression model using Chl alone. As would be expected if a high Phae/Chl ratio provides evidence for reduced photosynthetic performance, the coefficient for this term is negative, indicating a reduction in predicted pH at a given concentration of Chl as the Phae/Chl ratio increases. Moreover, the intercept and slope values of this multiple regression model are nearly identical to those of the June-only bivariate model (Table 1), indicating that when differences in algal physiological state are accounted for, pH_{pred} is similar for both June-only and June–September conditions.

Probability of exceedence analyses

The primary goal of this paper was to develop new probabilistic models that can be used to predict the frequency of exceeding critical values of water quality in these lakes, and that can be used to help guide the future management of the resident fish populations in Klamath and Agency lakes. The availability of objective pH_{crit} values based on established physiological limits for pH tolerance by fish (see Materials and methods), along with the observed relationships between algal biomass and pH in Upper Klamath and Agency lakes (Fig. 2), provided the empirical basis for a probabilistic approach to the modeling of fish success. New regression-based and cross-tabulation probability models for pH in these hypereutrophic lakes are presented below.

Fig. 3. Probability of exceeding four critical pH levels computed from the regression-based probability models developed for Upper Klamath and Agency lakes. Probability curves are computed from eq. 5 with input parameters from Table 1 (model subset descriptions for *a*, *b*, and *c* are the same as for Fig. 2). Symbols represent observed exceedence frequencies from the independent model validation data set and are plotted against mean Chl (see text for detailed explanation).

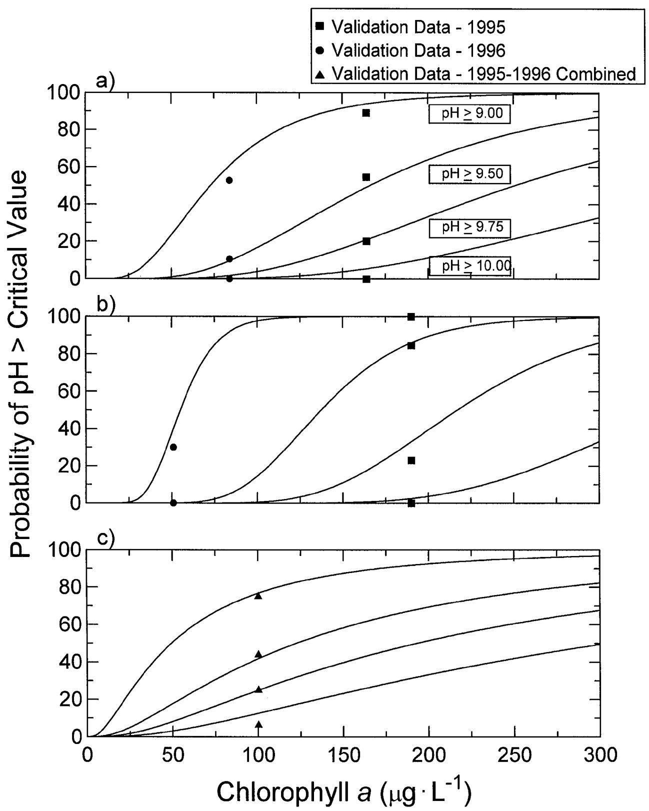
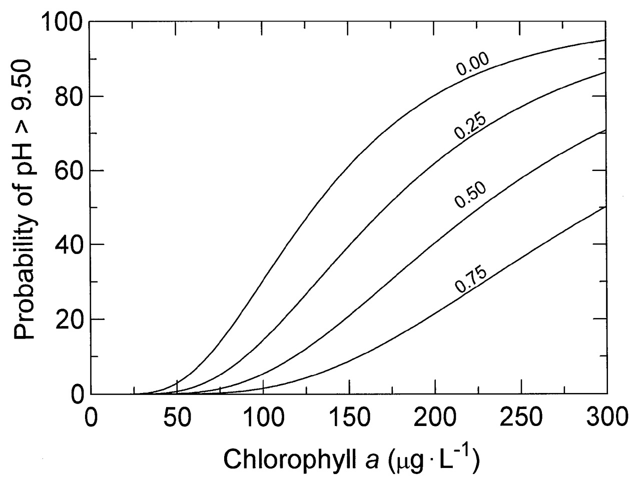


Fig. 4. Probability of exceeding pH = 9.50 at four constant levels of the Phae/Chl ratio (Phae/Chl = 0.00, 0.25, 0.50, 0.75) using the regression-based probability models developed for Upper Klamath and Agency lakes. Probability curves are computed by substituting eq. 4 for pH_{pred} in eq. 5.



Regression-based probability model

By combining the regression equations from Table 1 with eqs. 1 and 2, the probability of exceeding the four defined pH_{crit} values was computed from the following equation:

$$\text{(5)}\qquad \text{Probability of pH} > pH_{crit} = 100(1-\text{zcf})\left(\frac{pH_{crit}-(\beta_0+\beta_1(\log \text{Chl}))}{S_{y|x}}\right)$$

where $\beta_0 + \beta_1(\log \text{Chl})$ equals pH_{pred} and the portion of the equation within the large parentheses equals Z. The statistical basis for this model stems from the fact that by meeting the assumptions of linear regression (hence the strict attention paid to evaluating these assumptions), one can assume that pH_i is normally distributed with mean $\beta_0 + \beta_1\text{Chl}_i$ (where i = 1,2,..., n) and constant variance $S^2_{y|x}$ (Kleinbaum et al. 1988). Thus, for any value of Chl, a mean pH and its standard error can be computed, and from this method, any proportion of a normal distribution that lies beyond a defined critical pH value can be computed based on the normal probability density function.

When applied to the range of Chl data for each of the three models in Table 1, eq. 5 produces a series of probability curves for each chosen pH_{crit} value (see lines in Fig. 3). For example, Fig. 3*a* shows that for the lake-wide mean June–September model, the probability of exceeding pH_{crit} > 9.50 ($P(pH_{crit} > 9.50)$) is 60% at Chl = 200 $\mu g \cdot L^{-1}$. In contrast, $P(pH_{crit} > 9.50)$ is only 18% at Chl = 100 $\mu g \cdot L^{-1}$. For these same two Chl values, $P(pH_{crit} > 9.50)$ was 90 and 18%, respectively, using the June model (Fig. 3*b*); for the Agency Lake model, $P(pH_{crit} > 9.50)$ was 70 and 40%, respectively (Fig. 3*c*). Given the healthier physiological state of the June blooms, it is not surprising that a higher probability of exceedence is predicted per unit Chl for this model.

The apparent effects of algal physiological state on the predicted probability of exceedence for Upper Klamath Lake during June–September can be illustrated by holding the Phae/Chl ratio constant in eq. 4 and computing a series of probability curves for each Phae/Chl ratio by substituting eq. 4 for pH_{pred} in eq. 5 (Fig. 4). This plot clearly shows reduced probability of exceeding pH_{crit} as the Phae/Chl ratio increases. It is interesting to note that Phae/Chl = 0 produces results similar to the June model (Fig. 3*b*), while Phae/Chl = 0.25 produces similar results to the lake-wide mean June–September model (Fig. 3*a*).

These data also suggest that the predicted exceedence values per unit Chl are higher in Agency Lake than in Upper Klamath Lake. Because all of these empirical models are based on integrated water column Chl samples (which would capture algal cells both inside and outside the photic zone), we conclude that photosynthetic pH elevation per unit Chl may be higher in Agency Lake because its shallower mixing depth should increase the amount of time that algal cells are exposed to optimal light conditions.

Nonparametric cross-tabulation probability model

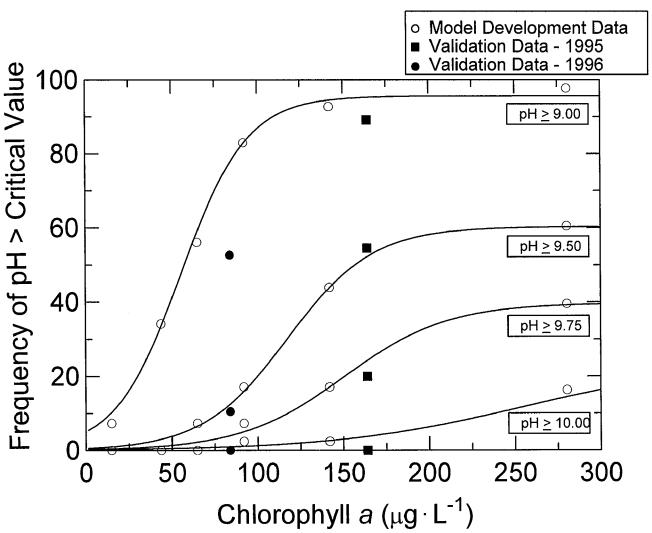
Similar probability of exceedence curves were generated using a nonparametric cross-tabulation procedure (see Materials and methods). This procedure requires no assumptions about the shape or functional form of the pH–Chl relationship (cf. Walker and Havens 1995); one simply computes exceedence frequencies for pH_{crit} values within the desig-

Table 2. Parameter estimates (±SE) and summary statistics for logistic probability models (eq. 3) of exceeding four different proposed values of pH_{crit} in Upper Klamath Lake.

pH_{crit}	pH_{∞}	β_0	β_1	$S_{y\|x}$	R^2
9.00	95.6±2.185	2.878±0.274	0.052±0.005	3.001	0.99
9.50	60.3±1.901	4.694±0.400	0.040±0.004	1.899	0.99
9.75	39.7±1.970	4.970±0.923	0.033±0.007	1.879	0.99
10.00	18.0[a]	5.143±0.662	0.024±0.003	0.818	0.99

[a]Value fixed based on plot.

Fig. 5. Probability of exceeding four critical pH levels computed from the nonparametric cross-tabulation probability models developed for Upper Klamath and Agency lakes. Frequency of exceedence curves are computed from eq. 3 with input parameters from Table 2. Open circles are the observed frequencies computed within six ordered and paired pH and Chl intervals and plotted on the median Chl of each interval). Solid symbols represent observed exceedence frequencies from the independent model validation data set and are plotted against mean Chl for the June–September time period.



nated intervals of ordered and paired pH/Chl values. To maximize the number of observations, and to maximize the amount of spatial and temporal variance accounted for, only one model was developed that used all paired measurements from the Upper Klamath Lake subset (n = 248). This data set was divided into six intervals, with the number of samples within an interval (n_{int}) equal to 41 for intervals 1–5 and 43 for interval 6. The frequency of pH values that exceeded each of the defined pH_{crit} values within each interval was then plotted versus the median Chl value for the interval (open circles in Fig. 5). Because these six data points appeared to approximate a logistic function, the logistic equation (eq. 3; Table 2) was used to describe the apparent relationship between exceedence frequency and Chl (lines in Fig. 5). These models provided an excellent fit to the observed frequencies within the six intervals (open circles in Fig. 5). Moreover, the predictions of these models were very similar to the June–September regression-based model; the frequency of exceeding pH_{crit} = 9.5 was 59% at Chl = 200 $\mu g \cdot L^{-1}$ and 20% at Chl = 100 $\mu g \cdot L^{-1}$.

Despite the lack of parametric assumptions in development of this modeling approach, an assessment of the model sensitivity to both n_{int} and the use of an interval median (or any chosen measure of interval central tendency) to construct the exceedence plot is necessary to ensure an accurate description of the data. When n_{int} is small, one or two values within an interval can greatly alter the computed frequencies. For example, if n_{int} = 10, then for every pH value that is below or exceeds pH_{crit}, there is a 10% decrease or increase in the computed frequency within the interval. Applying this concept to model development, it is apparent that a large data set is required and that the number of intervals that the data set is divided into cannot be so great as to reduce n_{int} to an unacceptable level. However, the number of samples within an interval must be balanced against the need to have enough intervals to describe the shape of the relationship. For example, a data set could be divided into two intervals to ensure high n_{int}, but this would not allow the functional shape of the relationship to be described. When exceedence frequencies are computed for one interval (the entire data set, n_{int} = 248), two intervals (n_{int} = 124), and so on through six intervals (Fig. 6), one can evaluate model sensitivity to interval choice.

Because the plotted values (open circles in Fig. 5) are based on median Chl values of each interval, each point actually represents a range of Chl values, and the described functional shape could vary if the placement of these values along the abscissa changed due to an inappropriate description of the Chl central tendency within intervals. An evaluation of model sensitivity to this can be gained from Fig. 6, whereby each set of intervals fewer than six produced completely different median Chl values from the six-interval model, yet the frequencies associated with them were generally well described by the logistic functions. To evaluate the entire range of Chl, exceedence frequencies for the pH_{crit} value 9.50 were computed based on moving intervals (n_{int} = 41, the same as that for six intervals) of each successive set of 41 samples (1–41, 2–42, 3–43,..., 208–248) and plotted on the median Chl of each moving interval (Fig. 7). As with Fig. 6, exceedence frequencies for Chl values falling between the six independent (nonoverlapping) intervals were also relatively well described by the logistic model.

The excellent fit of the logistic equations to the observed data, and the fact that model sensitivity to both the number of intervals chosen and the use of median Chl values for intervals was minimal, provide the basis for use of the logistic functions to predict future conditions.

Model validation

While the above models provide excellent descriptions of the observed data, their predictive ability must be validated with independently derived data. Walker and Havens (1995) proposed using the binomial distribution to compute stan-

Fig. 6. Sensitivity of the frequency of exceedence models to interval choice and use of medians to represent interval central tendency. Frequency of exceedence curves are the same as those for Fig. 5 and are overlain on observed exceedence frequencies for one (entire data set, n_{int} = 248), two (n_{int} = 124), three (n_{int} = 83), four (n_{int} = 62), five (n_{int} = 50), and six intervals (n_{int} = 41), which are plotted on their respective interval medians. The symbols denote each of the four pH_{crit} models (●, pH 9.00; ▲, pH 9.50; ■, pH 9.75; ◆, pH 10.00), and the symbol labels (numbers) refer to the number of intervals. In this way, the label "1" shows up once for each model, the label "2" twice, etc., through to the label "6", which shows up six times.

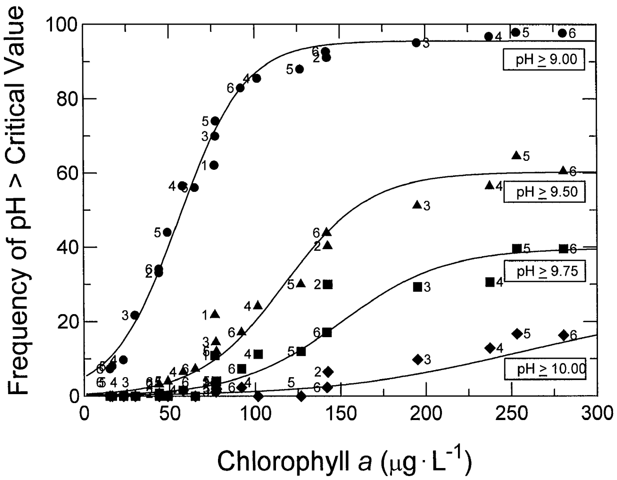
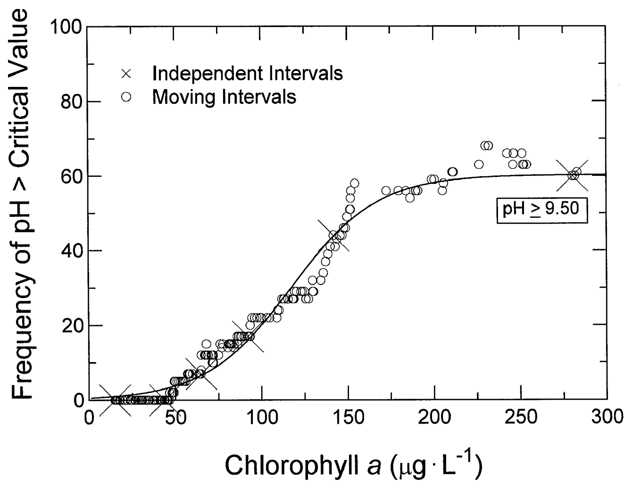


Fig. 7. Sensitivity of the modeled exceedence frequencies for pH_{crit} = 9.50 (line, same as that from Fig. 5) to the use of medians to represent interval central tendency for the entire range of Chl. Exceedence frequencies (○) were computed based on moving intervals (n_{int} = 41, same as that for the original six intervals) of each successive set of 41 samples (1–41, 2–42, 3–43,..., 208–248) and plotted on the median Chl of each moving interval. This generates frequencies for six nonoverlapping or independent intervals (×), which are the same as the original six intervals from Fig. 5.



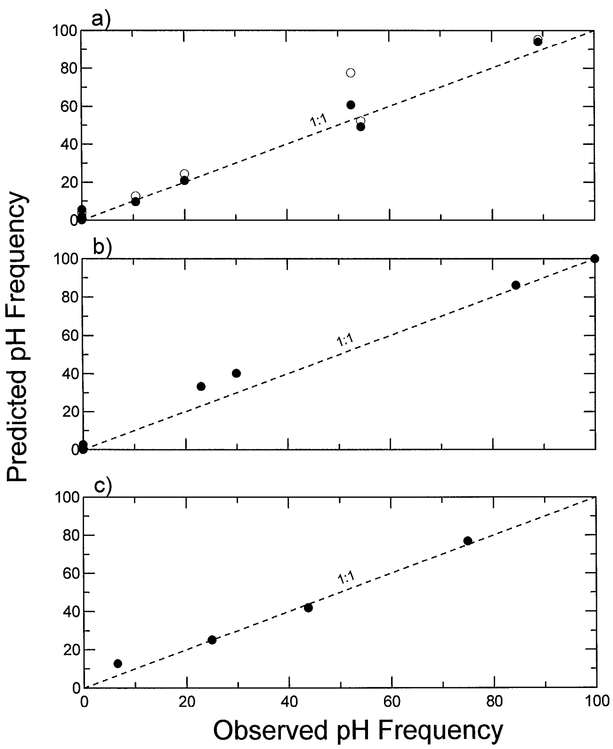
dard errors of measured frequencies within intervals for the cross-tabulation model. However, this procedure only provides an estimate of error in the descriptive model and does not serve to validate the predictive capability of the models, nor does it describe prediction uncertainty. Likewise, the nonlinear and linear regression statistics (Tables 1 and 2) are estimates of error in model fit to the observed data and also do not describe prediction uncertainty. With the multiyear database in Upper Klamath and Agency lakes, all available data were not needed to formulate the model, and the predictive ability of the models could be verified from independently derived data.

For each year of the Upper Klamath Lake 1995–1996 validation data set, the frequency of exceedence of pH_{crit} was computed as the number of observations (expressed as a percentage of the total observations) at all sites and dates within the June–September and June-only time periods that exceeded the chosen pH_{crit} values. The Agency Lake model is based on only one site; therefore, data from both years were combined to ensure an adequate sample size for validation. Exceedence frequencies were paired with the mean Chl for each of the time periods and overlain on the parametric- and nonparametric-based probability plots (Figs. 3 and 5). With the exception of the nonparametric model for pH_{crit} = 9.00, which overpredicted exceedence frequency by about 25% in 1996 (Fig. 5), both plots qualitatively indicate good prediction by all models. A quantitative measure of the predictive ability of the models was assessed by regressing predicted values from the models with observed values from the model validation data set (Fig. 8). The predictive ability of all models was highly significant ($r^2 > 0.95$, $P < 0.01$), and the predicted regression lines were not significantly different from the 1:1 relationship between observed and predicted values ($\beta_0 = 0$, $\beta_1 = 1$). After Keller (1989), a quantitative measure of the predictive ability between the regression-based and nonparametric models was assessed through F tests on the residual variability from the regression between observed and predicted values. This test showed no difference ($P > 0.10$) between the predictive ability of the two models. Despite a nearly twofold difference in June–September lake-wide mean Chl between 1995 (164 $\mu g \cdot L^{-1}$) and 1996 (84 $\mu g \cdot L^{-1}$), predicted exceedence frequencies for all pH_{crit} values were remarkably accurate for both years. Seasonal or time period mean Chl thus provides a good indication of the expected probability of exceeding various pH_{crit} values and therefore provides a good indication of the pH conditions experienced by fish both spatially and temporally at given seasonal mean algal biomass. These models clearly show that the frequency and severity of photosynthetically elevated pH events are greatly decreased as the concentration of Chl in the water column is reduced.

Practical implications

Probabilistic approaches have been used very successfully in the assessment of lake trophic state (Organization for Economic Cooperation and Development 1982; Reckhow and Chapra 1983). For example, they have been used in the trophic state classification of South African lakes (Walmsley 1984) and in the development of phosphorus criteria for Minnesota lakes (Heiskary and Walker 1988). In addition, they have been used to relate algal bloom frequency to concentrations of Chl (Havens 1994), to relate Chl to concentrations of total phosphorus (Walker and Havens 1995), and to

Fig. 8. Relationship between predicted exceedence frequency from the regression-based (●) and nonparametric (○) probability models and the observed frequency from the independent model validation data set (model subset descriptions for *a*, *b*, and *c* are the same as for Fig. 2; the broken line is the 1:1 relationship between the observed and predicted values). The regressions between the predicted and observed values were significant for all models ($r^2 > 0.95$, $P < 0.01$), and the predicted regression lines were not significantly different from the 1:1 relationship between observed and predicted values ($\beta_0 = 0$, $\beta_1 = 1$).



relate summer fish kill probability to total phosphorus (Mericas and Malone 1984).

We have adopted a probabilistic approach in this research because the proper management of fisheries in water systems requires not only a knowledge of the environmental conditions that result in fish kills, but also an understanding of both the lethal and sublethal responses of fish to changes in water quality. In the analysis presented here, we have recognized the importance of considering both responses. We have shown a clear relationship between algal biomass and photosynthetically elevated pH and have used parametric and nonparametric methods to evaluate the probability of exceeding critical limits for fish at varying concentrations of Chl in Upper Klamath and Agency lakes, Oregon.

In this paper, we have been strongly influenced by the work of Mericas and Malone (1984), who proposed that their probabilistic fish kill model could be linked to phosphorus loading models and that such a combined eutrophication modeling framework could be used in the objective assessment of lake management alternatives designed to minimize the risk of summer fish kills. We similarly suggest that the empirical models illustrated here (Figs. 3 and 5) can be used to help provide guidelines for the reductions in Chl that will be necessary to improve fisheries in Upper Klamath and Agency lakes. In large shallow hypereutrophic lakes with high external and internal nutrient loading, such as Upper Klamath Lake, complete reversals of eutrophication are typically not technically or economically feasible. However, Chl in general and blue-green algae in particular show substantial reductions in response to reduction in total phosphorus loading in such lakes, even when total phosphorus concentrations remain in the hypereutrophic range (>100 $\mu g \cdot L^{-1}$) after restoration (Jeppesen et al. 1991; Seip et al. 1992).

In Upper Klamath Lake, our analysis suggests that nearly a 45% reduction (72% for the June model) in the probability of exceeding pH 9.5 could be realized by reducing Chl from 200 to 100 $\mu g \cdot m^{-3}$. When linked with models that predict water column Chl concentrations from phosphorus loading, the combined models can be used as the basis for development of an objective framework to restore endangered fish populations in the Upper Klamath Lake system.

However, we emphasize that the models presented here are specific only to Upper Klamath and Agency lakes and that attempts to apply them directly to lakes of differing fertility, alkalinity, mean depth, and algal assemblages should be done with extreme caution. Rather, we suggest that lake-specific relationships between algal biomass, pH, and fish success should be developed for use in water bodies elsewhere, using approaches similar to those used here. This probability-based approach incorporates the inherent spatial and temporal variance in the independent and dependent variables and reexpresses it as the probability or frequency of stressful conditions likely to be experienced by fish. Once validated, these probability-based models permit an objective definition of the degree of water quality impairment and provide a quantitative lake management tool.

However, we also suggest that the general risk assessment approach presented here (relating a key quality parameter of concern → physiological tolerance limits → probability of exceedence of critical tolerance limits) can be more broadly applied in lake management and restoration. This procedure could equally well be applied to modeling the behavior of dissolved oxygen, whereby concentrations of Chl were linked to the frequency with which oxygen concentrations in the water column exceed known tolerance limits for fish or for other animals. Moreover, the biota in most surface waters are now being subjected to a wide variety of stressors, including toxic contaminants (Laws 1993). We suggest that the probabilistic methods used here to relate fish success, pH, and concentrations of Chl may potentially be of value in modeling the harmful effects of these other environmental stressors as well.

Acknowledgments

We particularly thank Kenneth Knight and Don Gentry for their assistance with field collections and laboratory analysis. We are also grateful to Larry Dunsmoor and Craig Bienz for suggestions and editorial comments. In addition to the very helpful comments of all anonymous reviewers, discussions with Drs. William Walker, Eugene Welch, and Kenneth Reckhow helped solidify ideas. Additional field assistance was provided by Edward Case, Bill Ehinger, Richard Barbiero, and Mark Osburn. Sharon Campbell, Mark Buettner, and the Klamath Tribes were instrumental in providing continued support for data collection through U.S. Bureau of Reclamation – Klamath Tribe Cooperative Agreements.

References

Alabaster, J.S., and Lloyd, R. 1980. Water quality criteria for freshwater fish. Butterworths, London, U.K.

American Fisheries Society. 1992. Investigation and valuation of fish kills. Spec. Publ. 24, American Fisheries Society, Bethesda, Md.

Barica, J. 1975. Summerkill risk in prairie ponds and possibilities of its prediction. J. Fish. Res. Board Can. **32**: 1283–1288.

Beitinger, T.L. 1990. Behavioral reactions for the assessment of stress in fishes. J. Gt. Lakes Res. **16**: 494–528.

Bergerhouse, D.L. 1992. Lethal effects of elevated pH and ammonia on early life stages of walleye. N. Am. J. Fish. Manage. **12**: 356–366.

Cooke, G.D., Welch, E.B., Peterson, S., and Newroth, P. 1993. Restoration and management of lakes and reservoirs. Lewis Publishers, Boca Raton, Fla.

Dillon, P.J., and Rigler, F.H. 1974. The phosphorus–chlorophyll relationship in lakes. Limnol. Oceanogr. **19**: 767–773.

Falter, M.A., and Cech, J. 1991. Maximum pH tolerance of three Klamath Basin fishes. Copeia, **4**: 1109–1111.

Havens, K.E. 1994. Relationships of annual chlorophyll *a* means, maxima and bloom frequencies and intensities in a shallow eutrophic lake (Lake Okeechobee, Florida, U.S.A.). Lake Reservoir Manage. **10**: 133–138.

Hazel, C.R. 1969. Limnology of Upper Klamath Lake, Oregon, with emphasis on benthos. Doctoral dissertation, Oregon State University, Corvallis, Oreg.

Heiskary, S.A., and Walker, W.W. 1988. Developing phosphorus criteria for Minnesota lakes. Lake Reservoir Manage. **4**: 1–9.

Jeppesen, E., Kristensen, P., Jensen, J.P., Sondergaard, M., Mortensen, E., and Lauridsen, T. 1991. Recovery resilience following a reduction in external phosphorus loading of shallow, eutrophic Danish lakes: duration, regulating factors and methods for overcoming resilience. Mem. Ist. Ital. Idrobiol. **48**: 127–148.

Kann, J. 1998. Ecology and water quality dynamics of a shallow hypereutrophic lake dominated by cyanobacteria (*Aphaniomenon flos-aquae*). Doctoral dissertation, Curriculum in Ecology, University of North Carolina, Chapel Hill, N.C.

Keller, A.A. 1989. Modeling the effects of temperature, light, and nutrients on primary productivity: an empirical and a mechanistic approach compared. Limnol. Oceanogr. **34**: 82–95.

Kleinbaum, D.G., Kupper, L.L., and Muller, K.E. 1988. Applied regression analysis and other multivariable methods. PWS-Kent Publishing, Boston, Mass.

Laws, E.A. 1993. Aquatic pollution. 3rd. ed. Wiley, New York.

Mericas, C., and Malone, R.F. 1984. A phosphorus-based fish kill response function for use with stochastic lake models. N. Am. J. Fish. Manage. **4**: 556–565.

Nusch, E.A. 1980. Comparison of different methods for chlorophyll and phaeopigment determination. Arch. Hydrobiol. Beih. **14**: 14–36.

Oregon Department of Environmental Quality. 1994. Issue paper: dissolved oxygen. 1992–1994 Water Quality Standards Review, Dissolved Oxygen Standard Subcommittee. Oregon Department of Environmental Quality, Water Quality Division, Portland, Oreg.

Organization for Economic Cooperation and Development. 1982. Eutrophication of waters: monitoring, assessment and control. Organization for Economic and Cooperative Development, Paris.

Reckhow, K.H., and Chapra, S.C. 1983. Engineering approaches for lake management. Vol. 1. Data analysis and empirical modeling. Ann Arbor Science, Butterworth Publishers, Woburn, Mass.

Reynolds, C.S. 1984. The ecology of freshwater phytoplankton. Cambridge University Press, New York.

Saiki, M.K., Monda, D.P., and Bellerud, B.L. 1999. Lethal levels of selected water quality variables to larval and juvenile Lost River and shortnose suckers. Environ. Pollut. **105**: 37–44.

Scoppettone, G.G., and Vinyard, G. 1991. Life history and management of four endangered lacustrine suckers. *In* Battle against extinction: native fish management in the American West. *Edited by* W.L. Minkchley and J.E. Deacon. University of Arizona Press, Tucson, Ariz. pp. 359–377.

Seip, K.L., Sas, H., and Vermij, S. 1992. Nutrient–chlorophyll trajectories across trophic gradients. Aquat. Sci. **54**: 58–76.

Smith, V.H. 1998. Cultural eutrophication of inland, estuarine, and coastal waters. *In* Successes, limitations and frontiers in ecosystem science. *Edited by* M.L. Pace and P.M. Groffman. Springer-Verlag, New York. pp. 7–49.

Vanni, M.J., Luecke, C., Kitchell, J.F., and Magnuson, J.J. 1990. Effects of planktivorous fish mass mortality on the plankton community of Lake Mendota Wisconsin: implications for biomanipulation. Hydrobiologia, **200**/**201**: 329–336.

Walker, W.W., Jr., and Havens, K.E. 1995. Relating algal bloom frequencies to phosphorus concentration in Lake Okeechobee. Lake Reservoir Manage. **11**: 77–83.

Walmsley, R.D. 1984. A chlorophyll *a* trophic status classification system for South African impoundments. J. Environ. Qual. **13**: 97–104.

Wilkinson, L. 1996. Systat 6.0 for Windows: statistics. SPSS Inc., Chicago, Ill.

Williams, J.E. 1988. Endangered and threatened wildlife and plants: determination of endangered status for the shortnose sucker and the Lost River sucker. Fed. Register, **53**(137): 27 130 – 27 134.

Zar, J.H. 1984. Biostatistical analysis. Prentice Hall, Englewood Cliffs, N.J.