Setting water quality objectives for the health of freshwater fish

J. M. Dunn[1]   J. Rong[2]   M. McKubre-Jordens[3]
B. Henderson[4]   R. Goudey[5]   R. S. Anderssen[6]

(Received 10 May 2012; revised 12 February 2013)

Abstract

An important responsibility of the Environment Protection Authority, Victoria, is to set objectives for levels of environmental contaminants. To support the development of environmental objectives for water quality, a need has been identified to understand the dual impacts of concentration and duration of a contaminant on biota in freshwater streams. For suspended solids contamination, information reported by Newcombe and Jensen [*North American Journal of Fisheries Management*, 16(4):693–727, 1996] study of freshwater fish and the daily suspended solids data from the United States Geological Survey stream monitoring network is utilised. The study group was requested to examine both the utility of the Newcombe and Jensen and the USA data, as well as the formulation of a procedure for use by the

Environment Protection Authority Victoria that takes concentration and duration of harmful episodes into account when assessing water quality. The extent to which the impact of a toxic event on fish health could be modelled deterministically was also considered. It was found that concentration and exposure duration were the main compounding factors on the severity of effects of suspended solids on freshwater fish. A protocol for assessing the cumulative effect on fish health and a simple deterministic model, based on the biology of gill harm and recovery, was proposed.

Contents

1   The study group's participants

The colleagues who contributed to the deliberations of the study group were Simon James, Gang Li, Jia Rong and Gleb Beliakuv from Deakin University Melbourne; Maarten McKubre–Jordens from the University of Canterbury, Christchurch, New Zealand; Burzin Bhavnagri from Swinburn University Melbourne; Mali Abdollahian and Jessica Dunn from the RMIT University Melbourne; Rob Goudey and Brendan O'Malley from EPA Victoria, Melbourne; Bob Anderssen and Brent Henderson from CSIRO Mathematics, Informatics and Statistics, Canberra.

2   Introduction

An important statutory role of the Environment Protection Authority (EPA) is to set objectives that specify environmental quality requirements for protection

of beneficial uses of the Victorian environment. EPA environmental objectives usually take the form of limiting values for specific environmental pollutants, and are often set as absolute limits for individual values of a pollutant, or as a population parameter for the current distribution of the concentration of the pollutant.

For example, the EPA sets objectives to safeguard stream biota from the effects of particular stream pollutants. For a target contaminant C, using the available science, a threshold concentration $\mathcal{T}_{\mathcal{C}}$ is set to define the boundary between acceptable and unacceptable levels for the contaminant. The goal is to protect the ecological health of the stream by initiating management actions when the threshold is exceeded. In order to check whether the situation is acceptable or unacceptable, it is necessary to compare some representative estimate of the current concentration $\mathcal{R}_{\mathcal{C}}$ with $\mathcal{T}_{\mathcal{C}}$. The threshold $\mathcal{T}_{\mathcal{C}}$ may be specified as an ideal statistical population parameter, such as a median or an extreme percentile. The concentration $\mathcal{R}_{\mathcal{C}}$ will then be the corresponding sample estimate derived from monitoring data. Because of the large number of Victorian water quality monitoring sites, and the costs of sampling and analysis, the sampling rate is usually constrained to twelve observations per year (that is, monthly site visits). Percentiles outside the range of the first and third quartiles have been found to be very unreliable if based on twelve observations, and so EPA often (depending on context) specifies objectives for the 75th percentile.

Currently, professional judgement and experience, along with historical information and published research, is utilized to set the threshold $\mathcal{T}_{\mathcal{C}}$, below which the concentration of the target contaminant is gauged to be acceptable.

2.1 The background

Currently, the setting by EPA Victoria of environmental objectives are based on the effects of exposure concentrations alone. Explicit use of the duration of the exposure is ignored. However, a low concentration over an extended
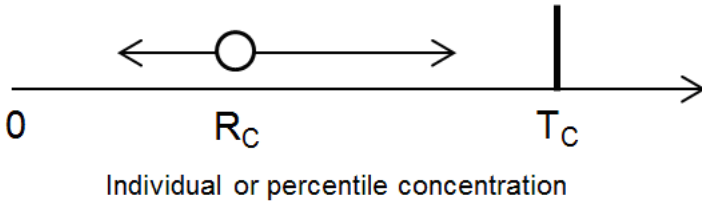
Individual or percentile concentration

Figure 1: A graphical illustration of the relationship between $\mathcal{T}_e$ and $\mathcal{R}_e$ for 'safe fish health' situations.

period can be as harmful as a high concentration over a short period. In addition, it is known that exposure concentrations and durations are jointly very important for determining the severity of biological effect. There is general agreement that the severity of the impact on fish health is influenced by the concentrations of suspended solids and the duration of exposure to those concentrations [11, 14, e.g.]. Effects of exposure to suspended solids can include, but are not limited to, behavioural changes, habitat degradation, reduced growth rates, and changes in fecundity and population densities.

Methods are needed for calculating objectives that more completely use both pieces of information. In many instances objectives are derived based on results of studies of biological effects of pollutants on organisms. Such studies determine the concentrations at which organisms show undesirable effects given exposure to the pollutant for a particular duration.

Of particular interest to EPA Victoria is the study by Newcombe and Jensen [10] which provides a promising method for determining suitable threshold objectives $\mathcal{T}_e$. Newcombe and Jensen [10], motivated by the data collection and dose-response work of Macdonald and Newcombe [8] and Newcombe [9], sought to assess the effect of suspended sediment on particular fish groups categorised by taxonomic group, natural history and life history phase. The effect was measured according to a scale of severity (SEV), which scores the ill effects of suspended sediments on fish on a scale from 0–14 and encompasses

four subcategories: nil effects, behavioural effects, sublethal effects, and lethal effects. This scale is mentioned in conjunction with and widely employed as a method for setting water quality objectives and exploring management scenarios [13, 16, 19, e.g.].

However, no known analysis of the scale of severity (SEV) exists. An approach to performing such an analysis is discussed and applied in Section 3.

2.2 The available information

The study group was provided with background information in order to investigate methods for deriving environmental objectives for limiting the harm to freshwater fish. The study group's investigation comprised a general proof of concept study based on the motivation of the Victorian EPA and the North American data. The specific information provided was the following.

1. Data published by Newcombe and Jensen [10] which summarised the biological effects of clean (no attached or independent toxins) suspended solids on freshwater fish from 80 laboratory studies. Included was Newcombe and Jensen's severity of effects scale (SEV) and regression equations relating SEV to the logarithms of exposure concentrations and durations.

2. Daily time series data of suspended solids concentrations in freshwater streams extracted from the United States Geological Survey (USGS) website. For the proof of concept presented in this article, we focus solely on data from the Sacramento at Freeport.

2.3 The focus for the study group's deliberations

In order to improve the associated decision making when setting water quality objectives, EPA Victoria asked the study group to explore whether

the information in the Newcombe and Jensen [10] data can contribute to and enhance this process. In addition, because the Newcombe and Jensen model was purely phenomenological (that is, a linear fit to log concentration and log duration) EPA Victoria requested consideration be given to the formulation of a model which took the appropriate fish biology into account.

In order to answer these questions, the following matters became the focus for investigation by the members of the study group.

1. Using modern dimension reduction and categorizaton tools, re-examine Newcombe and Jensen's conclusions about the structure of the information content in their data, and in particular whether the data partition naturally into two separate categories of 'sublethal' and 'lethal' effects.

2. Examine the relevance and appropriateness of the Newcombe and Jensen model of *"the impact of contamination on the survival of fish"*, and, in particular, the appropriateness of the 'Severity Index' (SEV) characterization of the impact, and how a time series of a contaminant might be assessed against both concentration and duration thresholds.

3. Model the impact of contamination on the gills of fish as a cumulative process, and explore the extent to which such a model represents a biological basis for improving on the Newcombe and Jensen model.

Other related matters were discussed informally, such as possible roles for survival analysis, multivariate statistical methodology, and data mining, which did not subsequently become major focusses of the deliberations.

2.4 The unifying thread

The analysis of the Newcombe and Jensen [10] data and the associated modelling of fish health, performed by the study group's participants, involved the following coordinated sequence of investigations.

(i). The application of dimension reduction, self-organizing mapping (SOM)

and clustering protocols to explore for structure in the data and review the Newcombe and Jensen categorization into 'sublethal' and 'lethal'. This is the focus of the discussion in Section 3. It established that a slightly modified categorization is more appropriate than that suggested by Newcombe and Jensen.

(ii). A statistical analysis of the data to formalize a measure of the risk to fish health using the new categorization coming from (i). This is the focus of Section 4. The analysis and proposed measure of the risk were used to derive a protocol for predicting the level of risk for various concentration and duration scenarios.

(iii). Modelling the accumulation of fish gill damage using ordinary differential equations to obtain an estimate for fish survival. This is the focus of Section 5. It established that, by assuming that the decreasing gill area was inversely proportional to severity, lifetime can be defined in terms of the time it takes for the gill area to decrease to a critical threshold.

(iv). The interrelationship between the results coming from (ii) and (iii) was utilized as validation for the separate results coming from the investigation. This is the focus of Section 6.

Each step gave new insight into the analysis and interpretation of the data. Together, they yielded a new classification for lethal, a statistical protocol for estimating the risk to fish health, an interpretation of the Newcombe and Jensen data from a mathematical modelling of gill damage perspective, and an overall validation of the interconnectedness of the separate investigations.

2.5   The organization of the article

This article reports on the conclusions resulting from an investigation of the above matters. It has been organized in the following manner. Items (i) and (ii) above are examined in Sections 3 and 4, where the goal was to explore whether there is information and structure in the Newcombe and

Jensen data yet to be identified and exploited. This includes an examination in Section 4 of the joint utilization of the concentration and duration data of Newcombe and Jensen in conjunction with the USGS sedimentation data to predict the risk to fish health. Item (iii) is examined in Section 5 using ordinary differential equation modelling of the harm to the gills. The results of these deliberations are brought together in Section 6 where it is shown that the ordinary differential equation modelling of (iii) yields validation for the protocol.

Several authors have quantified the effects of suspended solids on gill structure and function [1, 4], but mathematical models of how suspended sediment effect gill area have not been formulated and analysed. A simple model is proposed and analysed in Section 5.

3   Data segmentation, self-organizing mapping and classification

In this section, the Newcombe and Jensen [10] categorization into 'sublethal' and 'lethal' is reassessed using a combination of data segmentation, self-organizing mapping, clustering and classification to highlight the dominant structures in the data.

3.1   Background

In diverse areas including business management and food classification, segmentation refers to the process of identifying similar groups such as people [15] or products [3], where the groups are fairly homogeneous with respect to some specified objective. The value of performing segmentation analysis includes ascertaining appropriate groups for specific future analysis with respect to the application under consideration, or at least gaining an understanding about

how to differentiate between the groups. The need for in-depth knowledge of segments remains an essential element of understanding the characteristics of any collected data set. Various techniques have been used to perform segmentation, ranging from elementary percentiles to multivariate analysis methods such as clustering analysis, factor analysis and principle component analysis. From a methodological point of view, clustering algorithms, based on Ward's hierarchical clustering [17], are being utilized to identify groups in quite novel situations including travelers with similar profiles [18].

When the number of variables involved is large, a dimension reduction technique such as principal component analysis (PCA) or factor analysis may be used to reduce the dimensionality to a manageable size before applying a clustering algorithm. The goal, in part, in performing the dimension reduction, is to obtain an 'independent' set of representative variables to reduce the possibility of bias in the original data due to cross-correlations.

Recently, more salient segmentation methods have been based on Kohonen's Self-Organizing Map (SOM) procedure, which maps an n-dimensional input space to a lower dimensional region while maintaining the basic original topological structure in terms of the chosen similarity measure. In the various implementations of the SOM procedure, the data items that were close in the higher dimensional input space remain close in the reduced lower dimensional map. The two dimensional graphical representations, generated by applying a dimensional reduction followed by an appropriate implementation of Ward's methodology [17], provide easy-to-understand maps and helps the decision makers visualize the dominant connections between the various inputs. These strengths make the SOM an appropriate technique for not only identifying the key interconnection in a wide variety of applications such as international travel, but also in the survival dependence of fish on the concentration and duration of a toxic episode.

## 3.2   Self-organizing mapping: data segmentation and clustering

The clustering phase of the SOM is performed on the two dimensional data. Its role is the segmentation of the data into similarity groups. Software is available for the application of Ward's procedure [17].

The clustering results generated by SOM for the Newcombe and Jensen [10] data are shown in Figure 2. In this situation, the input data, being just the 'duration' and 'concentration' values, was two dimensional. Using the algorithmic procedure outlined above with the Euclidean distance as the measure of similarity, a partitioning into seven clusters was generated, as shown in Figure 2(a). The relevance and appropriateness of the clustering is validated by plotting, for each vector ν_i using an appropriate colour scale, the value of some appropriate defining property. The resulting colour plots for ln(duration) and ln(concentration), with blue and red corresponding to the minimum and maximum values, are shown in Figure 2(b) and Figure 2(c). Not unexpectedly, they show that the clustering of Figure 2 performs a clear partitioning of both ln(duration) and ln(concentration), which thereby validates its appropriateness. Furthermore, as highlighted in Figure 2(b), the property controlling the clustering in Figure 2 most strongly is duration. The dominant importance of duration is further validated in the deliberations below in the statistical analysis of the Newcombe and Jensen data and in the modelling of fish health in terms of proportional gill area compromised. This is consistent with intuition in that a small concentration over an extended period can be as lethal (or more) as a high concentration over a short period.

The corresponding plot for the SEV categories is given in Figure 2(d). Its patchwork colour pattern highlights explicitly that SEV depends on the combined effect of concentration and duration in a more variable manner. Nevertheless, the pattern in Figure 2(d) is consistent with the combined patterns in Figure 2(b) and Figure 2(c) in that the strong duration clusters in Figure 2(b) are combined with the strong concentration clusters Figure 2(c), whereas the clusters of weak duration and weak concentration are less affected.

(a) Clusters

(b) $\ln(D)$
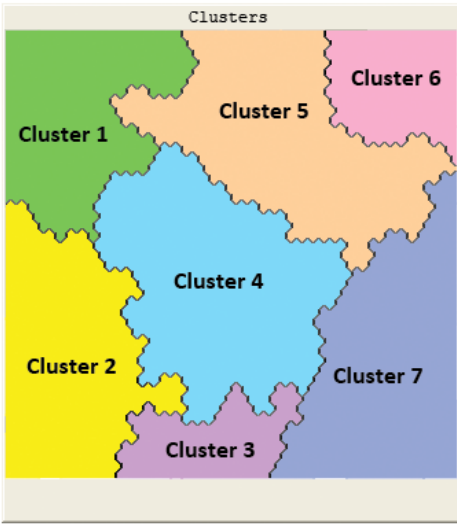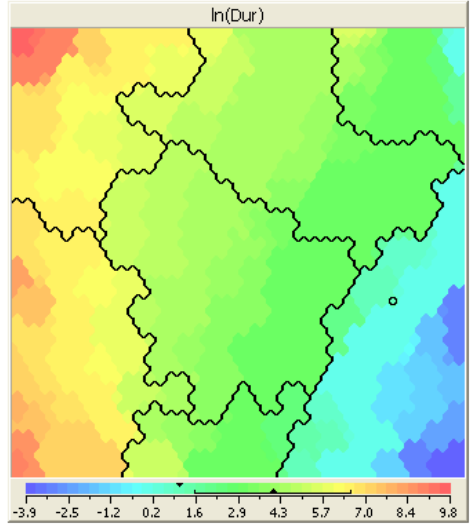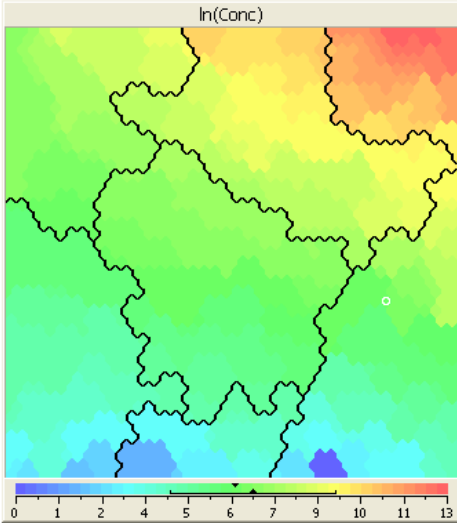
(c) $\ln(C)$

(d) SEV

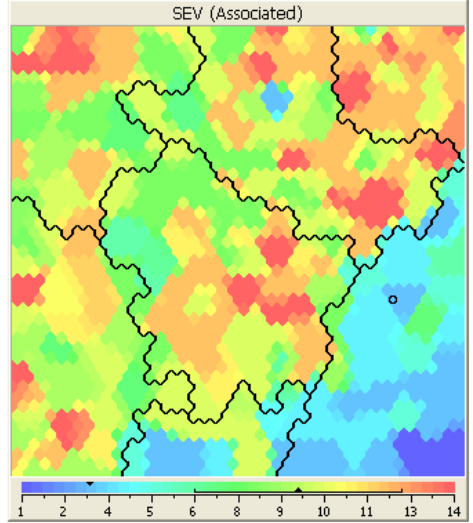Figure 2: The duration-concentration clustering: (a) the basic clustering; overplotting of (b) $\ln(D)$, (c) $\ln(C)$, (d) SEV.

Table 1: Clusters and their profiles.

Segment	Features	
Cluster 1	Long duration, Normal concentration,	High SEV
Cluster 2	Long duration, Low concentration,	High SEV
Cluster 3	Normal duration, Low concentration,	Low SEV
Cluster 4	Normal duration, Normal concentration,	High SEV
Cluster 5	Normal duration, High concentration,	High SEV
Cluster 6	Short/Normal duration, High concentration,	High SEV
Cluster 7	Short duration, Low Concentration,	Low SEV

Essentially, fish will have a high SEV categorization when either duration or concentration is high. A combination of normal duration and normal concentration can also give a high SEV categorization, whereas SEV is low when a short duration is applied even if the concentration level is reasonably high.

In addition, each cluster in Figure 2 has its own distinguished profile, the details of which are summarized in Table 1. For example, the fish in Cluster 1 and Cluster 2 contain most of the fish data records with the longer duration; the fish in Cluster 6 were subjected to the highest concentration; Cluster 3 and Cluster 7 group the records with both lowest concentration and shortest duration; Cluster 4 and Cluster 5 summarize the normal situations.

3.3 Assessing the sublethal and lethal categorization

The SEV categorization proposed by Newcombe and Jensen [10] was assessed by comparing it to an alternative classification or clustering approach based on k-Nearest Neighbours (k-NN) protocol [2] for which software is readily available (for example, WEKA [7]). Using the WEKA software, a k-NN classification for the Newcombe and Jensen data was performed. The resulting confusion matrix is given in Figure 3. It records the number of times that the k-NN

```
ix ===

   g  h  i  j  k  l  m  n    <-- classified as
   0  0  0  0  0  0  0  0 |   a = 1
   0  0  0  0  0  0  0  0 |   b = 2
   2  1  0  0  0  0  0  0 |   c = 3
   0  1  0  0  0  0  0  0 |   d = 4
   1  1  0  1  0  0  0  0 |   e = 5
   0  0  0  0  0  2  0  0 |   f = 6
   4  0  0  0  1  0  0  0 |   g = 7
   0 21  1  7  4  9  0  2 |   h = 8
   1  3 16  1  1  4  0  0 |   i = 9
   0  9  2 28  2  7  2  5 |   j = 10
   1  3  1  1  2  3  0  0 |   k = 11
   0 15  5  9  4 23  1  7 |   l = 12
   0  0  0  3  0  2  0  0 |   m = 13
   0  2  2  4  0 11  0  7 |   n = 14
```

Figure 3:   The confusion matrix corresponding to the Newcombe and Jensen [10] SEV categorization when compared with a k-NN categorization.

classification agrees with the Newcombe and Jensen SEV (which corresponds to the diagonal elements in the confusion matrix) and the number of time that the k-NN classification gives some other classification which does not agree with the SEV (which corresponds to the off-diagonal elements in the confusion matrix). As shown in Figure 3, the confusion matrix generates a different 'sublethal' and 'lethal' classification with the separation being SEV greater than or equal to eight, compared with seven as proposed by Newcombe and Jensen.

Because of the clear partitioning between 'sublethal' and 'lethal', highlighted in Figure 3, the subsequent analysis was based on setting SEV = 8 to be the threshold between 'sublethal' and 'lethal'.

4   Statistical analysis of the data and risk assessment protocol

A statistical analysis of the data was used to formalize a measure of the risk to fish health using the new (sublethal, lethal) categorization coming from the dimension reduction and SOM analysis. The resulting analysis and proposed measure of the risk were used to derive a protocol for predicting the level of risk for various concentration and duration scenarios.

4.1   Background

The severity index response was modelled within a regression framework (generalized additive models), using bivariate tensor splines to allow flexibility in the response surface. This model creates a smoother representation of the severity response and facilitates the interpretation. The response is examined for different fish groups and categorical assessments of health (lethal, sublethal). A key focus is on determining which combinations of concentration and duration are harmful to fish. This amounts to estimating a threshold in the concentration and duration space. Graphically this is represented as a line or curve on the response surface, above which there are substantial risks to health, and below which it is likely that there are no long lasting effects.

A form of statistical discriminant analysis determines the threshold that divides these two groups. The estimation of a concentration and duration threshold extends the current thresholds based solely on concentrations cur-

rently used by EPA Victoria and many other agencies focussed on water quality and ecosystem health within Australia and internationally. It also extends the response surface used by Newcombe and Jensen [10] from a plane to a more flexible bivariate spline surface that recognises the 0–14 severity score is not a linear scale.

4.2   The statistical analysis and definition of risk

4.2.1   The severity score data

The full set of severity scores of Newcombe and Jensen [10] for a large number of fish circumstances, as a function of contamination concentration and duration, is plotted in Figure 4. It shows that, for the variety of circumstances investigated, there is a clear increase in severity as each of duration and concentration is increased, with increased duration having the dominant effect. This increase is consistent with the findings in Section 2 and is utilized in motivating the deterministic modelling of gill harm in Sections 4 and 5. A generalized additive model, with bivariate tensor spline basis functions for concentration and duration, was fitted to this data and is plotted in Figure 5, the structure of which clearly highlights the critical role of duration.

The new categorization of 'sublethal' and 'lethal', with the threshold set at SEV $= 8$ is used in the subsequent analysis. From Figure 6, which plots the 'sublethal' and 'lethal' groups with separate plotting characters, it is clear that a small number of sublethal realizations are scattered within the lethal realizations. They correspond to measurements made on juvenile and adult fish and must correspond to quite healthy specimens. A generalized additive model consisting of a quadratic polynomial plus interaction terms in (natural log) concentration and duration was fitted and is plotted in Figure 6. This corresponds to a logistic regression given the binary response. The polynomial terms are less flexible than the tensor spline but we believe that the smoother structure captures the significant features of the data. Not unexpectedly,
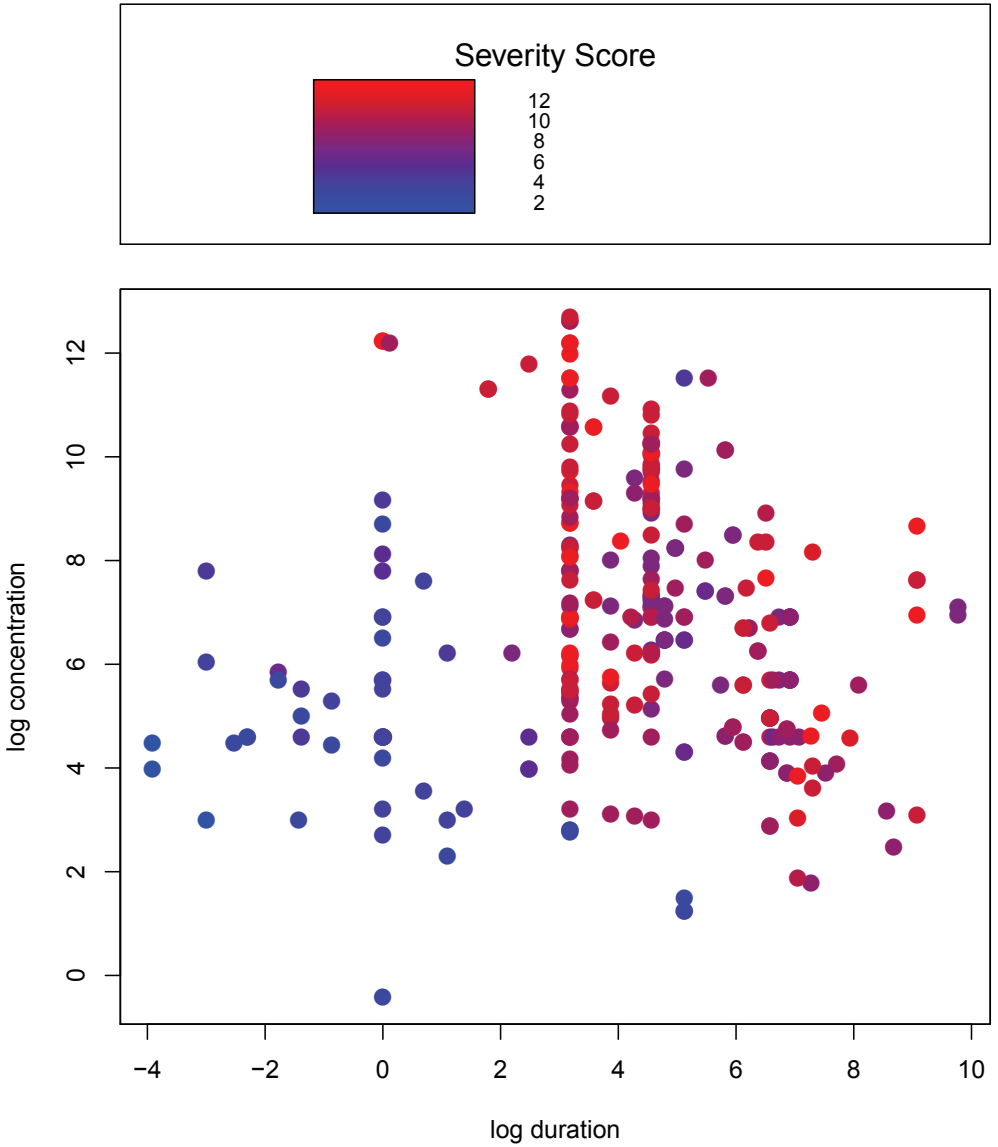
Figure 4: The severity score of Newcombe and Jensen [10] as a function of (natural) log concentration and log duration.
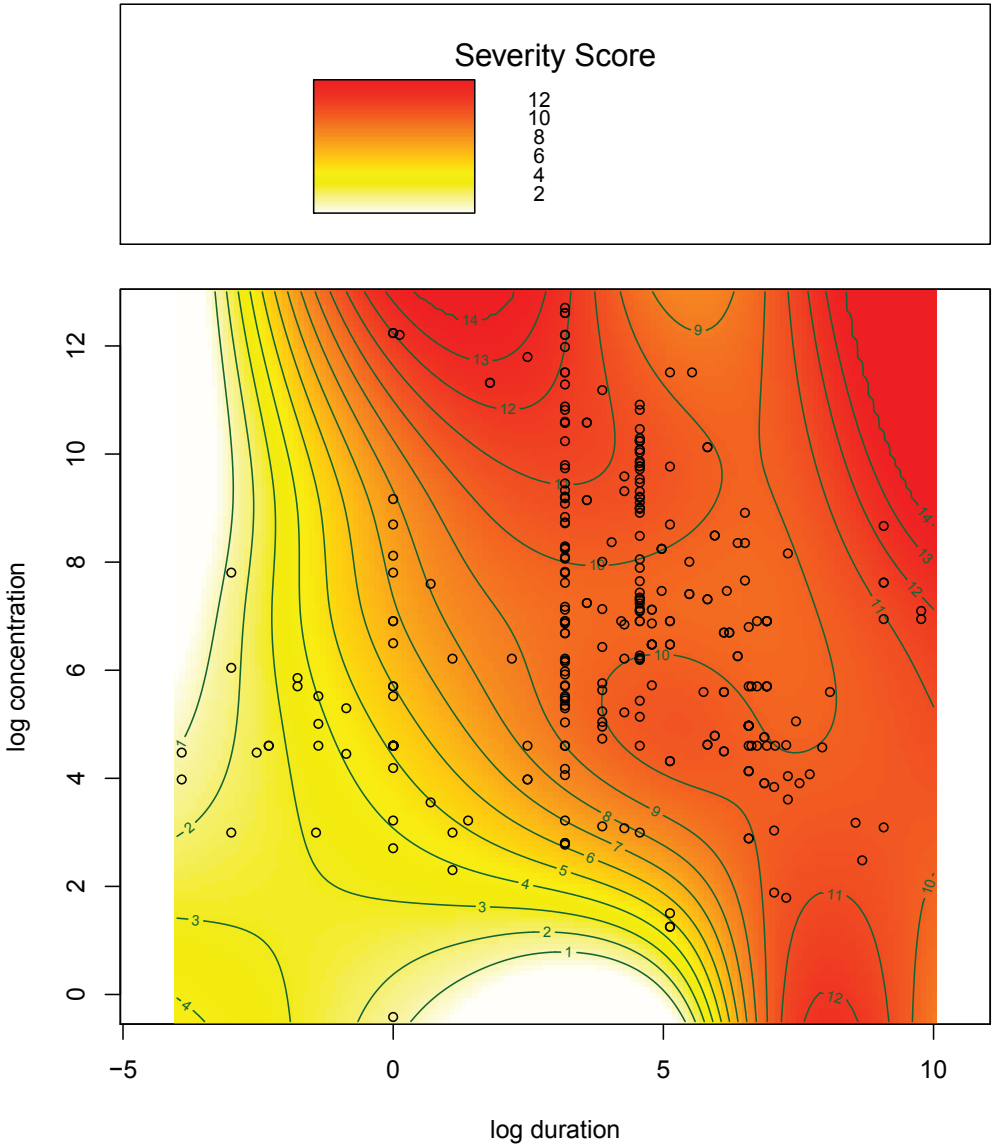
Figure 5: Predicted severity score from generalized additive model with bivariate tensor spline for (natural) log concentration and log duration.

this plot has a similar pattern to its counterpart in Figure 5. The model explains 46.1% of the deviance and can be viewed as having a reasonable predictive capacity. The scattering of some of the sub-lethal realizations among the lethal suggests that strong prediction will be difficult. Other covariates (for example, considering different fish types or stages) may help but this scattering may simply reflect natural variability. The *risk of a lethal effect* will be defined as the probability given by this model. As illustrated in Figure 6, the 0.5 probability contour represents an appropriate choice for the threshold between the sub-lethal and lethal effect of the contamination. Other choices may be used, particularly if it is more important to get classifications of one category (that is, lethal or sublethal) right more than the other.

4.3   The fish health risk assessment protocol

A representative example of river contamination that may compromise fish health is the United States Geological Survey (USGS) time series for daily suspended sediment concentrations, for the Sacramento River at Freeport between 1980 and 1988, and is plotted in Figure 7. Figure 8(a) plots the time series for a single year, 1985, and is a good example of the situation, where, for a reasonable amount of the time, the daily sediment levels are either acceptable or, if an unacceptable peak in the concentration occurs, it is of a short duration. Though the gills of the fish are briefly compromised under the brief duration situation, they recover quickly. Figure 8(b) highlights the maximum duration with a concentration of at least 50 mg/L.

Consequently, to assess fish health risk, the approach taken here is to consider the time series for the period of interest and examine maximum run lengths of concentrations exceeding a specific concentration value over a specific time period, here, taken to be a year. An illustration is given in Figure 8, where (a) plots the time series for 1985 and shows the maximum duration with a concentration at least 50 mg/L. A concentration and duration (maximum run length in days) pair is thereby generated. In this situation, a concentration
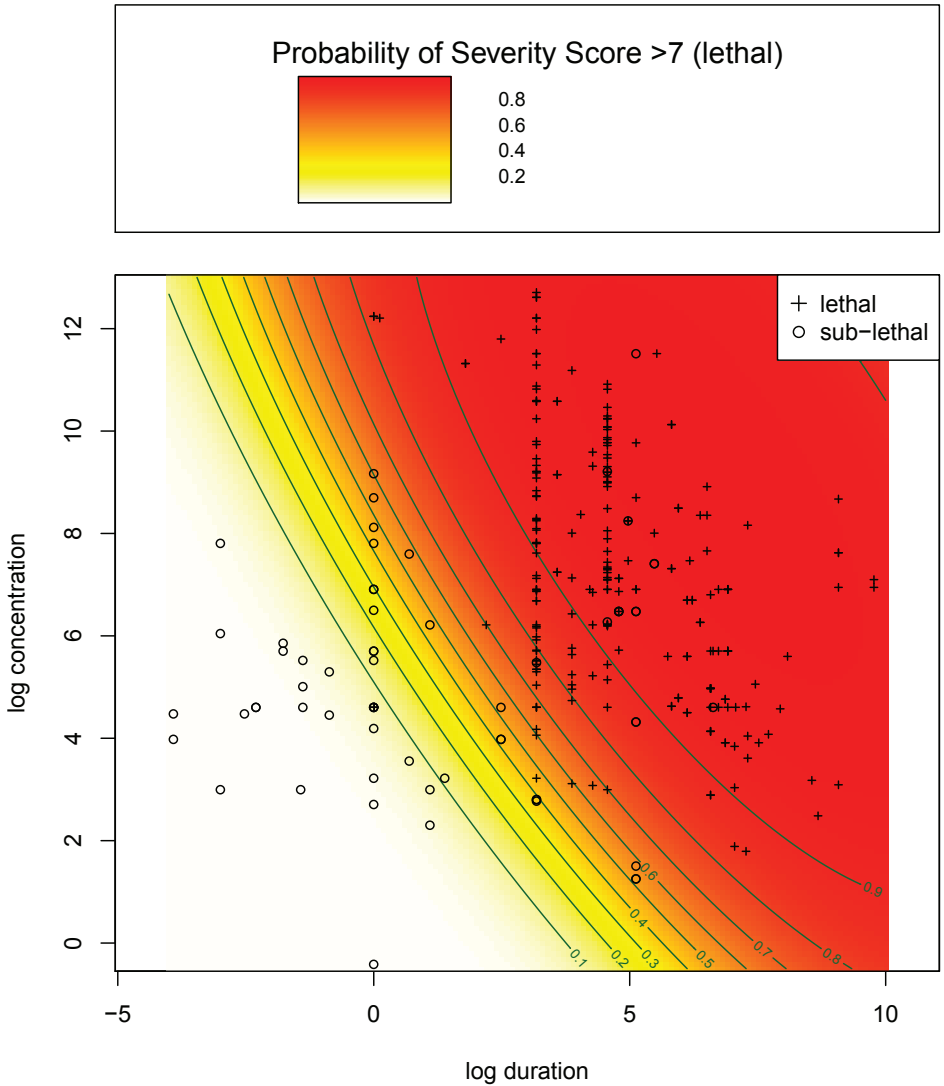
Figure 6: The probability of a lethal effect as determined by a generalized additive model, consisting of a quadratic polynomial plus interaction terms in (natural) log concentration and log duration.
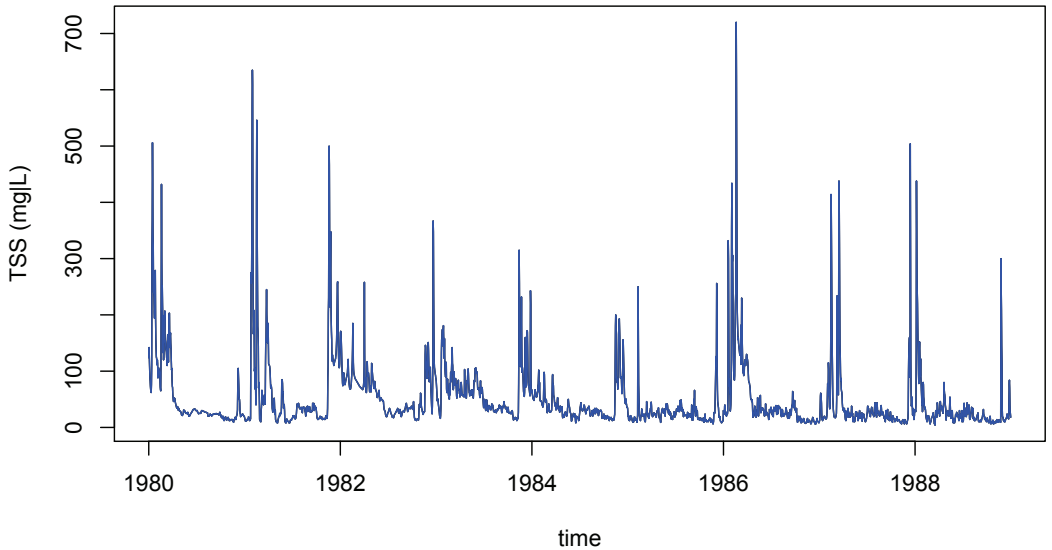
Figure 7: Suspended sediment for the Sacramento River at Freeport, 1980–1988.

of at least $50\,\mathrm{mg/L}$ has a maximum duration of nine days in 1985. This can be repeated for various choices of concentration values to create a set of concentration and duration pairs as illustrated in Figure 8(b) for concentrations ranging between 50 and $250\,\mathrm{mg/L}$. It gives a clear illustration of how the durations decrease as the concentrations increase, and how high concentrations are typically only sustained for short periods of time. To ascertain whether a specific pair corresponds to a lethal effect, it is considered in relation to the regression model and compared to the 0.5 probability contour. This is repeated for all pairs in a year and the maximum risk assessed as illustrated in Figure 9 where the maximum annual risk to fish health from concentration and duration is plotted for years 1980 to 1988.

Where the monitoring data is less frequent, a modelled or interpolated concentration series could be created first before considering the maximum run length, and the risk to fish health for that year. This may be important where
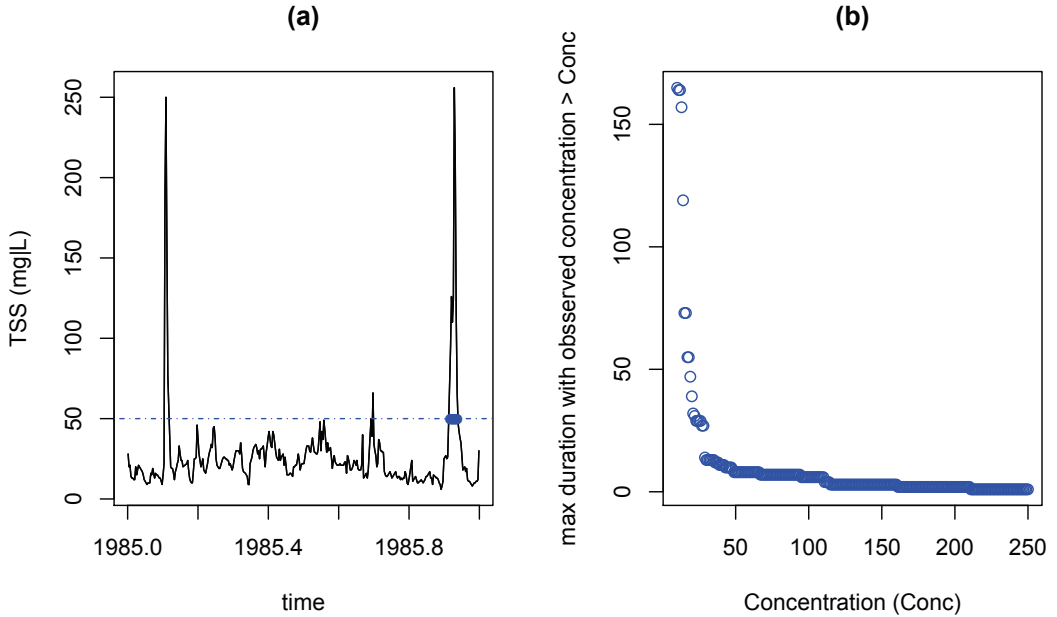
(a)

(b)

Figure 8: Illustration of maximum run length for total suspended solids greater than or equal to $50\,\mathrm{mg/L}$ for Sacramento River at Freeport during 1985.

records are collected weekly or based on event sampling. However, other challenges are introduced in that the modelling or interpolating must be seen to fairly represent the concentration peaks and durations.

This protocol is being described as a retrospective assessment where we may look back over the past year and assess the risk, and then put that in the context relative to other years to see if there have been changes or emerging trends in the risk to fish health. This framework also offers the potential to make an assessment of the risk in real time. If the concentration has been above some value $\mathcal{T}_{\mathcal{C}}$ for D days, then we can estimate the risk and decide whether a management response may be necessary.
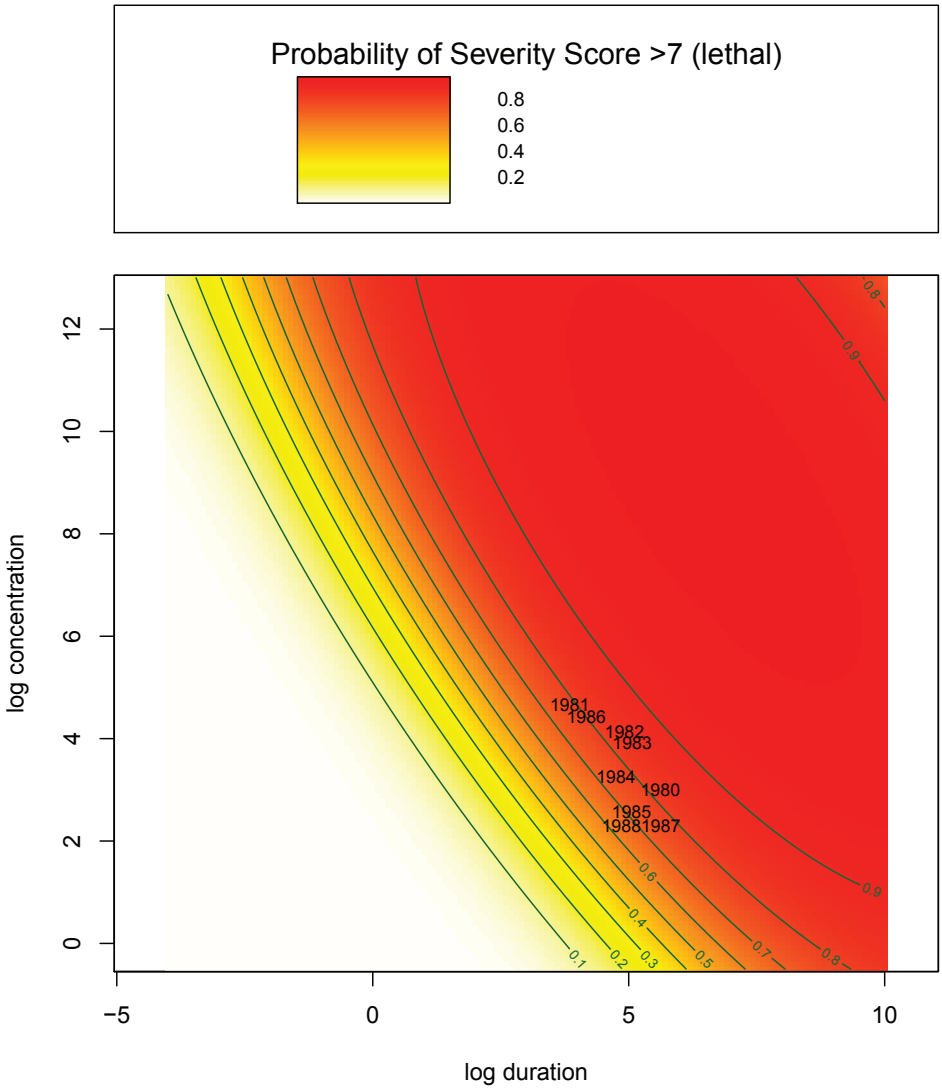
Figure 9: Maximum risk to fish health from joint consideration of concentration and duration for the Sacramento River at Freeport between 1980 and 1988 (natural logs).

5   Ordinary differential equation modelling of gill harm

The initial impact of a contaminant on fish health is via the effect that it has on the gills of a fish. Ordinary differential equation modelling was used to capture the essence of this process and thereby give a link back to the Newcombe and Jensen [10] data and the statistical analysis. This is achieved by assuming that the decreasing gill area is inversely proportional to SEV.

5.1   Background

An obvious effect of increased sediment loads is the harmful damage caused to fish gills. Changes to fish gill structure and function can cause effects which span the full range of SEV. In particular, in a review on the toxic effects on fish gills, Evans [5] (and citations therein) indicates that solute transportation interaction with the gill area causes gill malfunctions. Evans [5] suggests that it is this interaction with the gills that causes other physiological problems for fish. Consequently, fish health should be modelled on the basis of the effect of various contaminants on the gill.

Several authors quantified the effects of suspended solids on gill structure and function [1, 4], but mathematical models of how suspended sediment affect gill area have not been formulated and analysed. Here, a simple model is proposed and analysed.

It is known that

1. It is the gills of the fish that are affected by a contaminant when the concentration of the contamination exceeds a certain threshold $\mathcal{T}_{\mathbb{C}}$. This in turn directly reduces the efficiency of the gills to extract oxygen from

the water. Over an extended period the fish will develop a variety of complications which lead to death.

2. After a period where the contaminant has been above the threshold $\mathcal{T}_e$, when the contamination again drops below this threshold, the gills of the fish recover slowly.

Assumptions

1. *Recovery rate* is proportional to the area of damaged gill $[A_G - A(t)]$ where $A(t)$ is the functioning area of the gill and A_G is the maximum gill area. This reflects that the biology of an organism responds proportionally to the extent of a health challenge (for example, area of gill damage). For simplicity, areas of gill are assumed to be either functioning and available or non-functioning and unavailable, with harmed gill eventually returning to their pre-harmed state:

$$\text{Rate of recovery from abrasion} = \lambda_r[A_G - A(t)]$$

where $\lambda_r > 0$ is a positive constant.

2. *Harm rate* is proportional to the rate of random impact of suspended sediment particles with the functioning gill area. All gill area is exposed and is equally likely to be impacted, but already harmed gill tissue is assumed to suffer little additional impact damage:

$$\text{Rate of harm by abrasion} = \lambda_h A(t)C(t)$$

where $\lambda_h > 0$ is a positive constant and $C(t)$ is the concentration of suspended sediment at time t.

5.2   Model formulation and analytic solution

The total rate of change of functioning gill area is assumed to be the difference
between the recovery and harm rates:

$$A'(t) = \lambda_r[A_G - A(t)] - \lambda_h A(t) C(t), \quad A'(t) = \frac{dA(t)}{dt}, \tag{1}$$

which yields the required ordinary differential equation model. A solution
can be obtained by standard algebraic methods:

$$
\begin{aligned}
A(t) = {} & \left[\lambda_r A_G \int_0^t \exp\left(\int_0^\tau (\lambda_r + \lambda_h C(\zeta))\, d\zeta \right) d\tau + A(0) \right] \\
& \times \exp\left(-\int_0^t (\lambda_r + \lambda_h C(\zeta))\, d\zeta \right).
\end{aligned}
\tag{2}
$$

If $C(\tau) = \bar{C}$, with $\bar{C}$ a constant, then the analytic solution (2) simplifies to
become

$$\bar{A}(t) = \left[\frac{\lambda_r A_G}{k} \{ \exp(kt) - 1 \} + A(0) \right] \exp(-kt), \tag{3}$$

where $k = \lambda_r + \lambda_h \bar{C}$.

5.3   The dynamics of the basic model

The steady state solution, corresponding to $A'(t) = 0$, is

$$A^* = \frac{\lambda_r A_G}{\lambda_r + \lambda_h C^*}, \tag{4}$$

where $C^* = C(\infty)$. There are at least two ways in which this result can be
utilized.

1. For an assumed steady state value C^* for the contaminant, Equation (4)
   determines the corresponding steady state value A^* for the effective gill
   area and vice versa.

2. If, for all $t \geqslant 0$, it is assumed that $C(t) = \bar{C}$, a constant, then Equation (1) becomes an autonomous system, the stability of which can be analysed. Its solution $\bar{A}(t)$ is given in Equation (3) and is rearranged to the form

$$\bar{A}(t) = \frac{\lambda_r A_G}{k} + \left(A(0) - \frac{\lambda_r A_G}{k}\right) \exp(-kt), \quad k = \lambda_r + \lambda_h \bar{C}, \quad (5)$$

and thereby yields

$$\bar{A}^* = \bar{A}(\infty) = \frac{\lambda_r A_G}{\lambda_r + \lambda_h \bar{C}} = \frac{A_G}{1 + \lambda_h \bar{C}/\lambda_r}.$$

This establishes, as expected, the following properties for the steady-state value $\bar{A}^*$ for the functioning gill area,

(a) for fixed λ_r, $\bar{A}^*$ decreases as the value of either λ_h or $\bar{C}$ increases,

(b) for fixed λ_h and $\bar{C}$, the value of $\bar{A}^*$ increases as λ_r increases, and

(c) $\bar{A}^*$ takes the form of A^* of Equation (4), since $\bar{C}$, being constant, also corresponds to C^*.

For the situation where $C(t) = \bar{C}$, Equation (1) is rearranged to give

$$\bar{A}'(t) = k\left(\bar{A}^* - \bar{A}(t)\right) = \Theta[A(0)]\exp(-kt), \quad \Theta(\zeta) = (\lambda_r A_G - k\zeta), \quad (6)$$

where $\Theta(\zeta) = \Theta[A(0)]$ is a function of the initial state of the fish's gills at the time $t = 0$ when the fish is first subjected to the contamination at concentration $\bar{C}$. The sign of $\left(\bar{A}^* - \bar{A}(t)\right)$ is controlled by the sign of the initial state $\Theta[A(0)]$. Since $\zeta = A(0) \leqslant A_G$, it follows that

$$\Theta(\zeta) = \lambda_r A_G - k\zeta = \lambda_r(A_G - \zeta) - \lambda_h \bar{C}\zeta \geqslant -\lambda_h \bar{C}A_G.$$

For a healthy fish,

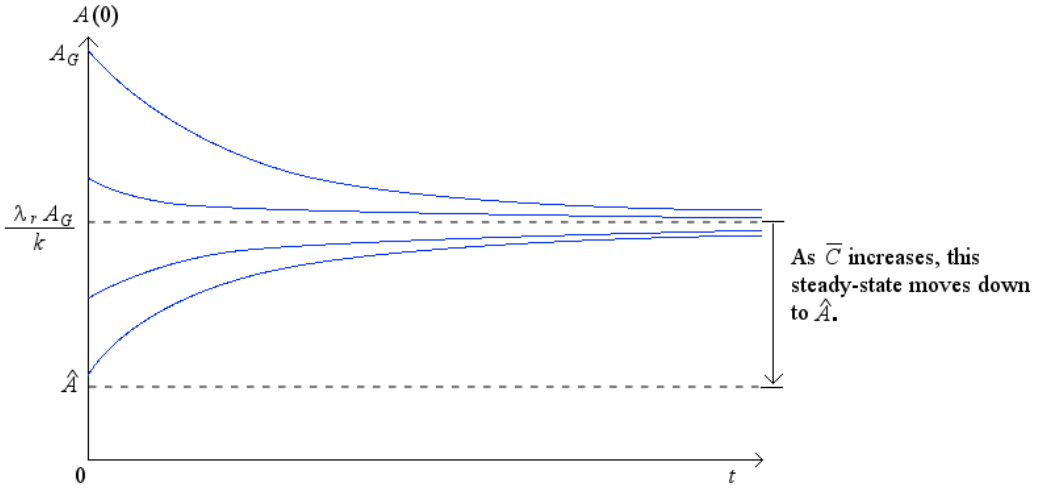$$A(0) = A_G, \quad \Theta(A_G) = -\lambda_h \bar{C}A_G.$$

Figure 10: Plot of $A(t)$ against time for several different choices of $A(0)$, with parameters $\lambda_r, \lambda_h, \bar{C}$ fixed.

It therefore follows from Equation (6) that

$$\bar{A}(0) = \bar{A}^* + \lambda_h \bar{C} A_G \,,$$

and hence

$$\bar{A}^* - \bar{A}(0) = -\lambda_h \bar{C} A_G \,.$$

Consequently, from Equation (5) and illustrated in Figure 10, with respect to a constant concentration $\bar{C}$, the functioning area of the gills of a healthy fish decays to the steady state value $\bar{A}^* = \lambda_r A_G / k$.

In this model, λ_r, λ_h and $\bar{C}$ determine the rate of decay k in the exponential term of Equation (6) independent of the initial health of the fish.

There is a sensible threshold $\hat{A} \ll A_G$ such that if $A(0) \leqslant \hat{A}$ then (for biological reasons) recovery is not possible and the fish dies. This threshold value is determined by fish biology and will depend on the species; it is

related to the critical oxygen tension that a fish must maintain to survive. Assuming $\bar{C}$ is nonzero, there are two possibilities.

$\lambda_r A_G/k \geqslant \hat{A}$ The level of $\bar{C}$ is such that the health of the fish's gills asymptotically approaches the value $\lambda_r A_G/k$.

$\lambda_r A_G/k < \hat{A}$ The level of $\bar{C}$ is such that the health of the fish's gills is compromised to the extent that the fish dies in a finite amount of time. This will occur when the value of $\bar{C}$ is such that $\lambda_r A_G/k \leqslant \hat{A}$ and, hence, when

$$\bar{C} > \frac{\lambda_r(A_G - \hat{A})}{\lambda_h \hat{A}} \,. \tag{7}$$

Equation (7) gives an estimate of what concentration of contaminant causes the fish to die of gill damage; and may thus be useful for setting safe levels.

For an unhealthy fish ($\hat{A} < A(0) < A_G$), there are three possibilities.

$A(0) > \lambda_r A_G/k$ The level of $\bar{C}$ is such that the health of the fish is further compromised with the functioning area decaying to the value $\lambda_r A_G/k$. This occurs when contamination ($\bar{C}$) changes from a low concentration to a higher value.

$A(0) = \lambda_r A_G/k$ Functioning gill area of the fish remains the same.

$A(0) < \lambda_r A_G/k$ The level of $\bar{C}$ is such that the health of the fish partly recovers with the functioning gill area asymptotically approaching $\lambda_r A_G/k$. This is the type of situation which will occur when the contamination environment changes from a high value of $\bar{C}$ to a lower value.

6   Linking the modelling with the laboratory (bioassay) experiments

In a standard bioassay, an initially healthy fish (with $A(0) = A_G$) is exposed to suspended sediment at a constant concentration level $C^\#$ until mortality occurs, or it is clear that the level of $C^\#$ is not sufficiently high to cause death. It follows from Equation (5) that the corresponding solution to (1) becomes

$$A(t) = A^\# + (A_G - A^\#) \exp\left(-(\lambda_r + \lambda_h C)t\right), \quad A^\# = \frac{\lambda_r A_G}{\lambda_r + \lambda_h C^\#}. \quad (8)$$

Mortality occurs if the oxygen tension in the gills of a fish falls and stays below a *critical oxygen tension*, P_C. In particular, Grigg [6, p.1] comments

> Below the critical oxygen tension the oxygen transport system is presumably unable to satisfy the demand of the tissues for oxygen and oxygen consumption decreases, eventually to a level below which further oxygen uptake ceases.

In a discussion of allometric scaling for fish gill area in relation to fish mass and oxygen requirements, Schmidt–Nielsen [12, p.114] concludes

> ... that the surface area of the fish gill is related to the requirements for oxygen supply and that the surface area is scaled to the body size with a relationship similar to that for oxygen consumption.

Extensive experimental evidence, which fits with intuition, has established that, to first order, the oxygen tension that a fish can maintain depends on the available gill area. It follows from the discussion above (Equation (5)) that, if $\hat{A}$ denotes the gill area at which the oxygen tension equals P_C, then the concentration of the contamination

$$\hat{C} = \frac{\lambda_r(A_G - \hat{A})}{\lambda_h \hat{A}} \quad (9)$$

defines the life-death threshold for fish survival.

Remark 1. A more comprehensive model would need to take account of the fact that a fish has the ability to adjust its effective gill area to changing circumstances.

Consequently, only if $C^{\#} > \hat{C}$ will the gill area of the fish be compromised to the point where death occurs in finite time $T^{\#} < \infty$. For such a $C^{\#}$, this will occur when

$$A(T^{\#}) = \hat{A} = A^{\#} + (A_G - A^{\#}) \exp\left[-(\lambda_r + \lambda_h C)T^{\#}\right]. \qquad (10)$$

The assumption that it is only the compromised gill area that is the cause of death now introduced into the analysis by setting $\hat{A} = \hat{\alpha} A_G$ into the last equation, where $\hat{\alpha}$ is a response characteristic of the fish. As a function of $T^{\#}$ and $C^{\#}$,

$$\hat{\alpha} = \frac{1}{A_G} \left\{A^{\#} + (A_G - A^{\#}) \exp\left[-(\lambda_r + \lambda_h C^{\#})T^{\#}\right]\right\}. \qquad (11)$$

The parameter $\hat{\alpha}$ denotes the proportion of gill area below which a fish is unable to maintain the critical oxygen tension level at or above P_C.

A simple rearrangement, taking account of the form of $A^{\#}$ then yields

$$T^{\#} = \frac{-1}{\lambda_r + \lambda_h C^{\#}} \ln\left[1 - (1 - \hat{\alpha})\left(1 + \frac{\lambda_r}{\lambda_h C^{\#}}\right)\right], \qquad (12)$$

the validity of which holds only if

$$C^{\#} > \frac{(1 - \hat{\alpha})}{\hat{\alpha}} \frac{\lambda_r}{\lambda_h} > 0. \qquad (13)$$

The importance of this constraint and the associated Equations (11) and (12) is that they highlight the following points.

- At a concentration $C^{\#}$ of the contamination, a fish dies only if the constraint (13) is satisfied as only then is $T^{\#}$ finite.

- On the basis of Equations (12) and (13), it follows that the threshold concentration between survival and death of Equation (9) becomes

$$\hat{C} = \hat{C}(\hat{\alpha}) = \frac{(1 - \hat{\alpha})}{\hat{\alpha}} \frac{\lambda_r}{\lambda_h} \, ,$$

  with, for given values of $\hat{\alpha}$, λ_r and λ_h, survival corresponding to situations where $C^\# < \hat{C}$. In reality, this threshold is artificial and a more realistic value would be one where $\hat{C}$ corresponds to an appropriate large finite time.

- If, for a given $C^\#$, the corresponding $\hat{\alpha}$ is measured experimentally, then estimates for the ratio λ_r/λ_h can be obtained using the following upper bound

$$\frac{\lambda_r}{\lambda_h} < \frac{C^\# \hat{\alpha}}{1 - \hat{\alpha}} \, .$$

- If estimates for λ_r and λ_h are available, then $\hat{\alpha}$ can be predicted for a given $C^\#$. Biologically, this would allow discussions about the effect of contamination levels $C^\#$ to focus on how the gills of the fish are likely to be compromised.

- The plot in Figure 11 of

$$\hat{C}(\hat{\alpha}) = K\frac{(1 - \hat{\alpha})}{\hat{\alpha}} \, , \quad K = \text{constant},$$

  with $\hat{C}(0) = \infty$ and $\hat{C}(1) = 0$, has the basic hyperbolic monotone structure that agrees with the essence of the biology being modelled; namely,

  - at low levels of the concentration of the contaminant, the lifetime starts from an extremely high value which corresponds to fish survival,

  - at intermediate levels of the concentration, the changeover from survival to death occurs, with lifetime decreasing with increasing contamination, and
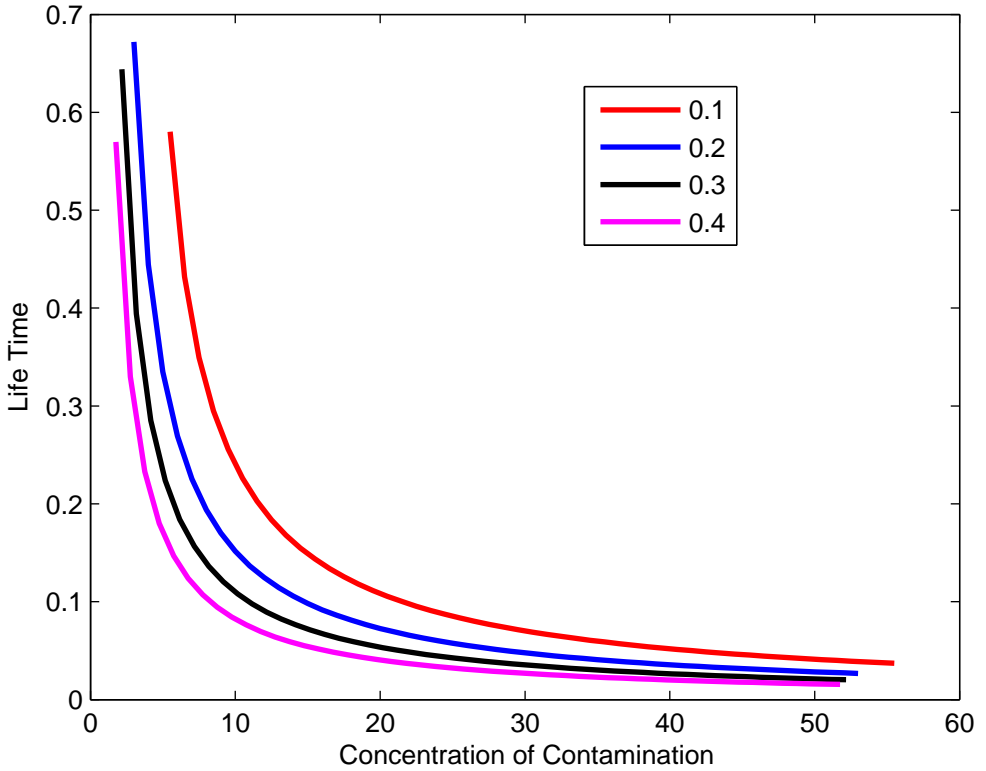
Figure 11: A plot of $\mathsf{T}^{\#}$ as a function of $\mathsf{C}^{\#}$ for some representative values of $\hat{\alpha}$ given in the legend.

– at high concentrations, survival, even for a short time, becomes problematic.

Equation (11), in conjunction with the constraint (13), defines $\hat{\alpha}$ as a function of $\mathsf{T}^{\#}$ and $\mathsf{C}^{\#}$. Since the reciprocal of $\hat{\alpha}$ represents a measure of *'severity'*, while lifetime corresponds to Newcombe and Jensen's concept of *'duration'*, a comparison can be made between the results generated by the deterministic model and the severity plots given in Newcombe and Jensen [10]. For scaling purposes, *severity* is defined to be $1/(0.1 + \hat{\alpha})$. For a representative choice
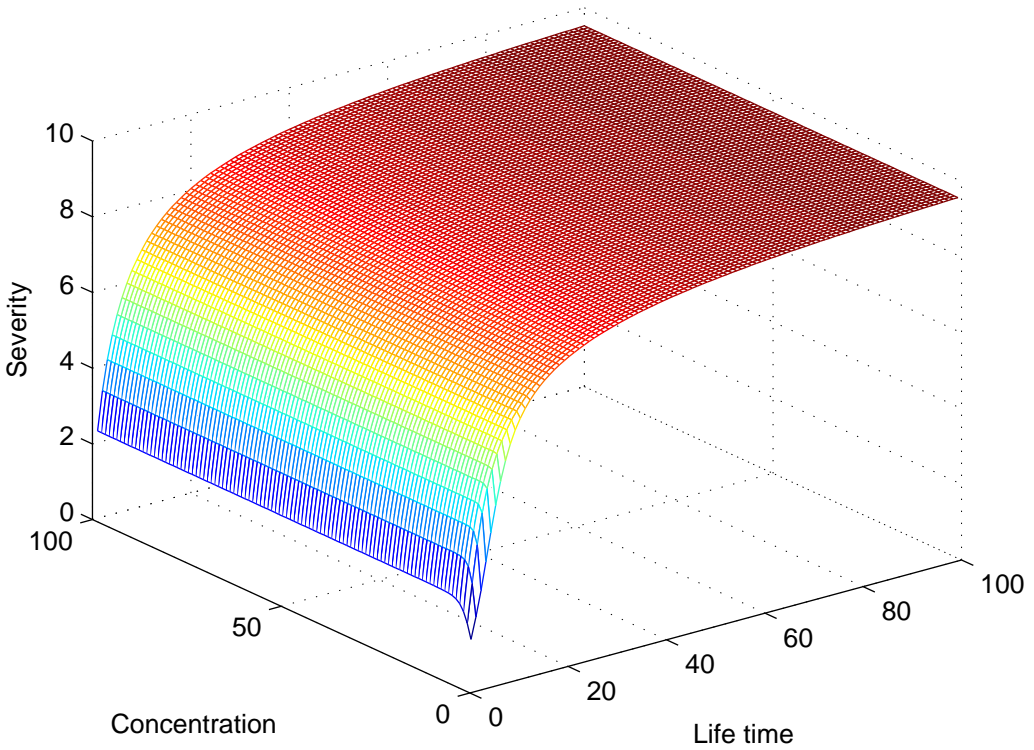
Figure 12: A plot of *severity* as a function of $T^{\#}$ and $C^{\#}$.

for the values of λ_r, λ_h and $\hat{\alpha}$, a plot of severity as a function of $T^{\#}$ and $C^{\#}$ is given in Figure 12. It clearly has a structure which is consistent with the underlying biology. A plot of severity as a function of $\ln T^{\#}$ and $\ln C^{\#}$ is given in Figure 13. It has a structure which is consistent with the log-log model proposed and utilized by Newcombe and Jensen [10] in their review and analysis of experimental fish survival data and with their plots in Figures 7 to 12.

Remark 2. As is clear from Figure 13, the log-log plot for severity has a planar-like structure not too dissimilar from the plots of Figures 7 to 12 by Newcombe and Jensen [10]. This therefore represents validation for the
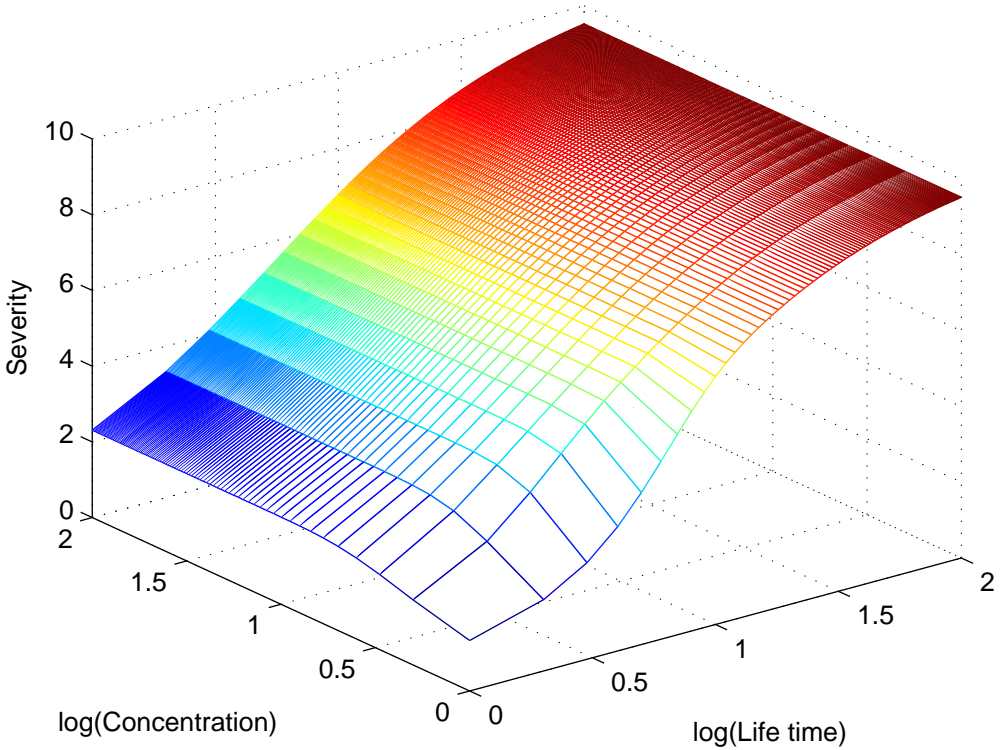
Figure 13: A plot of *severity* as a function of $\ln \mathsf{T}^{\#}$ and $\ln \mathsf{C}^{\#}$.

deterministic model (1). Its advantage is that it is a model, although basic, of the biology whereas the log-log model proposed by Newcombe and Jensen [10] is to a certain extent an exploitation of the 'linearizing effect' of a log-log plot.

6.1   Discrete time series analysis

Contamination, such as suspended solids in a river, are monitored at discrete times (for example, daily sampling). It is therefore appropriate to input

monitoring data and plot trajectories using a difference equation version of Equation (3). Some basic algebra, exploiting the properties of the exponential function, yields

$$\bar{A}(\tau + \theta) = \bar{A}(\tau)\exp(-k\theta) + \frac{\lambda_r A_G}{k}\left[1 - \exp(-k\theta)\right], \quad k = \lambda_r + \lambda_h \bar{C}, \quad (14)$$

where τ and $\tau + \theta$ denote two successive time steps. For an appropriate choice of parameter values, a representative example of its application is given in Figure 14 where the concentration data are the discrete black dots and the curve plots the changing relative effective gill area. It illustrates how a large concentration reduces the effective gill area of the fish, which recovers as the concentration drops providing the large concentration was not lethal.

7   Future possibilities

There are various ways in which the above modelling framework can be utilized.

Survival analysis In the future utilization of the above modelling framework, the possibility exists, once biologically representative values for the parameters are available, of performing survival analysis studies for different fish species with respect to various exposure to suspended solids scenarios.

Given the probability (in the frequentist sense) of survival of a fish as a function of gill area, the modelling developed here could be utilized to inform ecological decision-making about the likelihood of species survival with respect to different scenarios.

An immune response model Even though the modelling of the effect of contamination on effective gill area has been insightful, it ignores the more fundamental biology involved. A more appropriate model would allow for the recovery to be driven by an immune process, with the harm

Figure 14: Some representative daily time series concentration data (black dots) and the corresponding relative effective gill area.

calculated as before. Then, Equation (1) generalizes to some coupled system of ordinary differential equations such as

$$\frac{dG(t)}{dt} = \lambda_r G(t) E(t) - \lambda_h G(t) C(t),$$
$$\frac{dE(t)}{dt} = k[1 - G(t)],$$

where $G(t)$ and $E(t)$ model the proportion of damaged gill area and the nature of the immune response respectively, λ_r and λ_h are the recovery rate and harm rate, and k is the rate at which the immune response can act as a function of the remaining effective gill area.

In this more comprehensive model, that $G(t)$ depends on the concentration $C(t)$, in come complex manner related to the biology of the fish, can be taken into account, since it is known that oxygen tension in the fish varies with gill area and, hence, depends on $C(t)$.

8   Conclusions

A clustering analysis and a statistical analysis was performed on the Newcombe and Jensen [10] data. The clustering analysis yielded a new severity threshold for the sublethal to lethal categorization. Using this new threshold, the statistical analysis highlighted the dominant role of duration in determining the risk to fish health. Independent ordinary differential equation modelling of gill area yielded a connection back to the importance of duration of contamination.

In this way, it has been established how statistical modelling in conjunction with basic ordinary differential equation modelling, using representative data and knowledge about the known biological processes involved with fish health in contaminated water, has the potential to yield a framework to support ecological decision making of the type required by EPA Victoria.

References

[1] D. W. T. Au, C. A. Pollino, R. S. S Wu, P. K. S. Shin, S. T. F. Lau, and J. Y. M. Tang. Chronic effects of suspended solids on gill structure, osmoregulation, growth, and triiodothyronine in juvenile green grouper *epinephelus coioides*. *Marine Ecology Press Series*, 266:255–264, 2004. M70, M85

[2] J.C. Bezdek, S.K. Chuah, and D. Leep. Generalized k-nearest neighbor rules. *Fuzzy Sets and Systems*, 18:237–26, 1986. M74

[3] E. T. Champagne, K. L. Bett-Garber, A. M. McClung, and C. Bergman. Sensory characteristics of diverse rice cultivars as influenced by genetic and environmental factors. *Cereal Chem.*, 81:237–243, 2004. M70

[4] S. G. Cheung and P. K. S. Shin. Size effects of suspended particles on gill damage in green-lipped mussel perna viridis. *Marine Pollution Bulletin*, 51(8–12):801–810, 2005. M70, M85

[5] D. H. Evans. The fish gill: site of action and model for toxic effects of environmental pollutants. *Environmental Health Perspectives*, 71:44–58, 1987. M85

[6] G. C. Grigg. The failure of oxygen transport in a fish at low levels of ambient oxygen. *Comp. Biochem. Physiol.*, 29:1253–1257, 1969. M91

[7] G. Holmes, A. Donkin, and I.H. Witten. Weka: A machine learning workbench. In *Proceedings of the Second Australia and New Zealand*

Conference on Intelligent Information Systems, volume 24, pages 357–361, Brisbane, Australia, 1994. IEEE Computer Society. M74

[8] D. D. Macdonald and C. P. Newcombe. Utility of the stress index for predicting suspended sediment effects: response to comments. *North American Journal of Fisheries Management*, 13:873–876, 1993. M66

[9] C. P. Newcombe. Suspended sediment in aquatic ecosystems: ill effects as a function of concentration and duration of exposure. Technical report, British Columbia Ministry of Environment, Lands and Parks, Habitat Protection branch, Victoria, 1994. M66

[10] C. P. Newcombe and J. O. T. Jensen. Channel suspended sediment and fisheries: A synthesis for quantitative assessment of risk and impact. *North American Journal of Fisheries Management*, 16(4):693–727, 1996. M66, M67, M68, M70, M72, M74, M75, M77, M78, M85, M94, M95, M96, M99

[11] C. P. Newcombe and D. D. Macdonald. Effects of suspended sediments on aquatic ecosystems. *North American Journal of Fisheries Management*, 11(1):72–82, 1991. M66

[12] K. Schmidt-Nielsen. *Scaling. Why is animal size so important?* Cambridge University Press, NY, 1984. M91

[13] J. S. Schwartz, A. Simon, and L. Klimetz. Use of fish functional traits to associate in-stream suspended sediment transport metrics with biological impairment. *Environmental Monitoring and Assessment*, 179(1–4):347–369, 2011. M67

[14] E. Al Shaw and J. S. Richardson. Direct and indirect effects of sediment pulse duration on stream invertebrate assemblages and rainbow trout (*Oncorhynchus mykiss*) growth and survival. *Canadian Journal of Fish and Aquatic Science*, 58:2213–2221, 2001. M66

[15] P. Tiwari and H. Hasegawa. Demand for housing in Tokyo: A discrete choice analysis. *Regional Studies*, 38:27–42, 2004. M70

[16] Y. Tramblay, A. Saint-Hilaire, T. B. M. J. Ouarda, F. Moatar, and B Hecht. Estimation of local extreme suspended sediment concentrations in california rivers. *Science of the Total Environment*, 408:4221–4229, 2010. M67

[17] J. H. Ward. Hierarchical grouping to optimize an objective function. *JASA*, 58:236, 1963. M71, M72

[18] D. B. Weaver. Hard-core ecotourists in Lamington National Park, Australia. *J. Ecotourism*, 1:19–35, 2002. M71

[19] Q. Yang, G. Benoy, Z. Zhao, T. L. Chow, P. A. Charles, and F. R. Meng. Watershed-level analysis of exceedance frequencies for different management strategies. *Water Quality Research Journal of Canada*, 46(1):64–73, 2011. M67

Author addresses

1. **J. M. Dunn**, School of Mathematical and Geospatial Science, RMIT University, GPO Box 2476, Melbourne, Victoria 3001, Australia.
   mailto:jessica.dunn@rmit.edu.au

2. **J. Rong**, School of Information Technology, Deakin University, 221 Burwood Hwy, Burwood, Victoria 3125, Australia
   mailto:jiarong@tulip.org.au

3. **M. McKubre-Jordens**, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.
   mailto:maarten.jordens@canterbury.ac.nz

4. **B. Henderson**, CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra, ACT 2601, Australia.
   mailto:Brent.Henderson@csiro.au

5. **R. Goudey**, Center for Environmental Sciences, Environment

Protection Authority, Ernest Jones Drive, Macleod, Victoria 3085, Australia.
mailto:Rob.goudey@epa.vic.gov.au

6. **R. S. Anderssen**, CSIRO Mathematics, Informatics and Statistics, CSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra, ACT 2601, Australia.
mailto:bob.anderssen@csiro.au