Defining ecologically relevant effects sizes for changes in fish measurements after exposure to pulp mill effluents

K.R. Munkittrick*, C. Portt[1], B. Kilgour[2], M. McMaster[3] and R. Lowell[4]

* Canadian Rivers Institute, University of New Brunswick, Saint John, NB, Canada; krm@unb.ca
[1] C. Portt and Associates, Guelph, ON, Canada; cportt@sentex.net
[2]Stantec Consultants, Ottawa, ON, Canada; bkilgour@stantec.com
[3]National Water Research Institute, Environment Canada, Burlington, ON, Canada; mark.mcmaster@ec.gc.ca
[4]National Water Research Institute, Environment Canada, Saskatoon, SK, Canada rick.lowell@ec.gc.ca

There has been considerable emphasis since the first pulp mill conference in 1991 on trying to understand the ecological relevance of changes in endpoints used in fish monitoring programs. The Canadian EEM program has collected data from more than 300 field studies using fish, and contains data from more than 400 comparisons of different species and sexes with which to evaluate fish responses. The average fish response in Canadian environments receiving pulp mill effluent is one of reduced gonad sizes, and increased liver size and condition factor, identical to the first Canadian studies conducted at Jackfish Bay in the late 1980s. The average benthic community response reflects eutrophication in freshwater receiving environments. With the experience from these collections, we evaluate different approaches to setting critical effect sizes (differences we wish to detect) including (1) statistical significance, (2) natural variability, (3) best professional judgment/experience, and/or (4) an arbitrary number that is either (a) a threshold that would trigger a regulatory response or (b) a published number that others have defined as an impact at the population or community level. We have reviewed these approaches using data from the Canadian EEM program, and suggest that changes in critical performance parameters in a such as growth rate, organ size and critical endpoints that exceed 25% in a single monitoring cycle represent a change that reflects a that is a major concern. Changes that are smaller than this (i.e. 10%) but are consistent between cycles may represent a concern that warrants further study. Critical effects sizes that would warrant management action may be larger, or may be represented by changes that are getting worse over time. Decisions regarding management action are site-specific and will be dependent on a number of factors, including the magnitude and extent of changes, the species sensitivity, the number of endpoints responding, their trend over time, etc.

Keywords: fish populations, critical effect sizes, gonad size, condition, liver size, growth

Introduction

The Canadian Environmental Effects Monitoring (EEM) program was implemented within the Pulp and Paper Effluent Regulations in 1992 (Walker et al., 2002), and function to evaluate the effects of pulp and paper effluent on fish, fish habitat and the use of fisheries resources. Pulp and paper EEM is structured into approximately three-year sequences of monitoring and interpretation phases, and reports were submitted by more than 100 pulp mills in April of 1996 (cycle 1), 2000 (cycle 2) and 2004 (cycle 3). The average fish response in Canadian environments receiving pulp mill effluent over this time period has been one of reduced gonad sizes, and increased liver size and condition factor, which is the pattern observed in the first Canadian studies conducted at Jackfish Bay in the late 1980s (Munkittrick et al., 1991). The average benthic community response reflects eutrophication in freshwater receiving environments (Lowell et al. 2003, 2004, 2005). Similar programs have been developed for metal mining industries in Canada (2002), and are under consideration for sewage discharges (Kilgour et al., 2005) and other sectors. The fish component of the EEM program monitors impacts on individual-level endpoints reflecting age, energy investments (growth and reproduction) and energy storage. The program was designed to evaluate effects on key endpoints, measured on a minimum number of fish, on males and females of at last two species. An “effect” is defined as a statistical difference in a measured endpoint.

The first cycle of monitoring was reviewed by “expert working groups” comprised of both industry and government scientists (FSEWG, 1997), and was focused on developing a sufficient database to design future cycles. The study design component requires an assessment of variability in measured endpoints and power analysis to estimate the number of fish required to detect effects of a certain magnitude. In the absence of relevant data for cycle 1, target sample sizes for fish were set at 20 males and 20 females of two species. A site that finds no differences in required endpoints, and confirms the absence in a confirmatory cycle of monitoring, reflects a site defined as having no significant effects.

The data from the first cycle were reviewed by “expert working groups” comprised of both industry and government scientists (FSEWG, 1997). During this review (SFEWG, 1997), and in preparation for cycle 2, decisions had to be made regarding the magnitude of differences that the studies should be able to detect, and the acceptable levels of error, to allow study designs to be developed. For fish it was decided that α and β be set at traditional values (0.05 and 0.20, respectively), that the standard deviation be used from cycle 1, and that an effect size be set at 25% for gonad size (FSEWG, 1997), to enable calculations to be made regarding the number of fish to be collected. This was based on a decision that reproduction was the primary endpoint of concern, and that a 25% difference was approximately the magnitude of change that had been seen in earlier studies at Jackfish Bay (Canada) and Norrsundet (Sweden). The rationale was that it was desirable to know how often reproductive impacts occurred that were as large as those that had initiated the reproductive concerns, and precipitated the development of the EEM program. Different endpoints have different levels of variability, but the power for the other endpoints in fish was dictated by the sample sizes required for achieving sufficient power for gonad size. With the exception of growth (size at age) power levels are usually higher for the other endpoints).

There are challenges in getting consensus in interpretation of data from the monitoring studies (Munkittrick and Sandström, 2003; Munkittrick, 2004). During the review of cycle 2 data, the study design guidelines were adjusted to improve interpretability (Environment Canada, 2005a). Several changes were made, including an adjustment to α and β levels. The α level protects industry by protecting against declaring there is a difference when there is not one (Type I error), while β protects the environment by protecting against declaring there is no effect when there is one (Type II error). The revisions to the Pulp and paper Effluent Regulations stipulated that α and β should be set equal at a value of 0.1 or less. *A priori* decisions are needed on statistical significance (α), and power (1-β) levels to be achieved during the study.

Now that there are data from three cycles of monitoring (a total of >300 comparisons), there is an opportunity to review these earlier decisions about effects sizes. This paper examines different approaches to evaluating biological/ecological/statistical significance to determine what size of changes in individual-level endpoints should create a concern.

Ecological significance of changes in key endpoints

The fish population component of the EEM program defined what monitoring data would represent the absence of a concern, and that was defined as the presence of two species that show growth, reproduction, energy storage and age distribution that is not distinguishable from reference populations (Hodson et al., 1996; Environment Canada, 1997). There is recognition that there are species differences in sensitivity, but the assumption allows a starting point to identify the worst situations (Munkittrick et al., 2002).

Growth, reproduction, survival and energy storage are ecologically relevant processes, but the question remains how big a change in these key processes reflects an impact that warrants action. There are a number of types of relevance that can be assigned to differences: statistical, biological, ecological and regulatory significance may have different magnitudes. There is general agreement that the loss of a fish species or the presence of contamination that affects the consumption potential of fisheries resources are changes that are important to the public. A retrospective assessment process like EEM is meant to provide some information on how close effects might be to thresholds where species or communities may be at risk.

Under the EEM program, fish responses need to be beyond a defined size, and need to be confirmed, before monitoring is triggered into defining the geographic extent and magnitude and/or investigation of cause (Environment Canada 2002, 2005a). Exceedance of critical fish population effect sizes do not define whether changes are acceptable or not, but define whether the situation warrants collecting more information on the extent and magnitude of effects that will ultimately establish whether the situation is acceptable.

Within the EEM monitoring framework, effect sizes are used for the purposes of study design, and the effect size is defined as the magnitude of an effect that we want to be able to detect. The term "critical effect size" is currently used several ways, and has traditionally been used as a means of documenting the minimum effect size that will trigger management actions (Keough and Mapstone, 1995). The question is raised as to why the EEM program wants to be able to detect changes that do not warrant management action? The framework is designed with endpoints that are more sensitive than the regulatory endpoints (species presence and absence) to be able to allow response time to mitigate serious impacts. Changes in these whole-organism or population-level endpoints provide a warning that the performance of fish has been compromised, and that further study is needed.

The question then becomes what size of a warning do we want to be able to detect to allow some response time to evaluate the causes and initiate mitigation, if necessary? There is a lot of debate on the number that should be chosen to trigger further studies. At the current time within the EEM program, benthic community changes in richness, abundance or key diversity indices need to be greater than two standard deviations from the mean of the reference area to trigger further, more detailed monitoring (Lowell et al. 2003, 2005; Kilgour et al., 2005). For fish population parameters, there are a number of ways that the effect size could be developed, and they can be based on the presence of (1) statistically significant differences, (2) natural variability, (3) best professional judgment/experience, and/or (4) an arbitrary number that is either (a) a threshold that would trigger a regulatory response or (b) a published number that others have defined as an impact at the population or community level. Each of these is discussed below.

Best professional judgment (BPJ)

The Technical Guidance documents for pulp and paper (Environment Canada 1997, 2005) and metal mining (Environment Canada 2002) currently state that the objective is to be able to detect an approximate difference of 20-30% in gonad size between sites, for the purpose of designing the fish population studies. This magnitude was selected because it was the approximate size of differences in gonad size seen in early studies on pulp mill impacts at Jackfish Bay (Munkittrick et al., 1991) and Norrsundet, Sweden (Andersson et al., 1988; Sandstrom and Thoresson, 1988), and it was important to define how many studies detected differences as large as that. Since power, variability, and sample size requirements will vary with endpoint selected, the power required to detect differences in gonad size was selected as the endpoint of importance for study design.

Based on rare responses seen historically in studies

The pulp and paper EEM cycle 2 review showed that gonad size, liver size, mean age, and size-at-age differences were normally distributed (see Figure 1), with most responses being within ± 20%, and condition factor ± 10% (Munkittrick et al., 2002; Lowell et al., 2004). Changes beyond these size ranges were relatively rare, and represent situations where more information collection is warranted (Kilgour et al., 2005).

A more stringent interpretation has been recently implemented to determine which mills could go to Investigation of Cause. It was decided that mills that demonstrated more than a 10% decrease in gonad size, consistently in successive cycles, could participate in a national effort at identifying the causes and sources of chemicals in pulp mill effluent that are responsible for impacting reproduction. The magnitude of difference of effects that are important for defining subsequent cycles is still under evaluation, and a recent "Smart Reg" task force has recommended that effects sizes be further evaluated (Environment Canada, 2005b).

Outside a range seen in comparisons between comparable reference sites

A study in the Moose River basin in northern Ontario sampled 12 reference sites over a period of 7 years (Munkittrick et al., 2000). A comparison of GSI? values between reference sites, within years showed that in 33/36 comparisons differences were <30% (Figure 2). The 3 comparisons that exceeded 30% all involved a comparison with one site (the Missinaibi River, Thunderhouse Falls, MISS_NTH). Fish at that site showed differences with other reference sites in terms of liver size, condition, and growth compared to all other reference sites. The site also had dramatic differences in fish community structure. The average difference (n=36) was 3.4 ± 3.6 %.

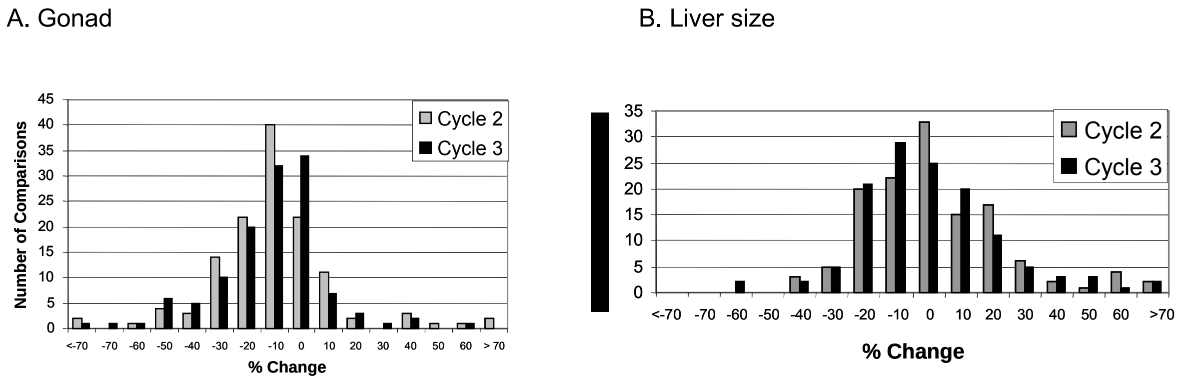

Figure 1. Distribution of effect sizes (% change from reference site) for cycle 2 fish comparisons (based on corrected electronic data, modified from Lowell et al., 2004)

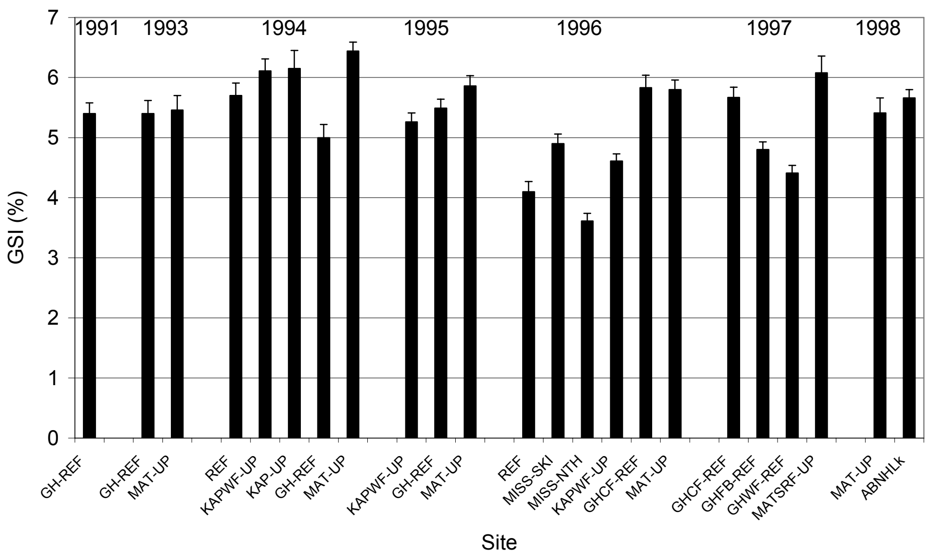

Figure 2. A comparison of female reference site gonad sizes (GSI=100*gonad weight/body weight) for Moose River reference sites (data adapted from Munkittrick et al., 2000).

Statistically different

Statistical differences depend on the magnitude of the underlying true difference, within-treatment (site) variability and sample size. Even when there are very small differences, those differences can be made statistically significant by collecting enough samples. The use of statistical significance as a criterion is therefore difficult to justify, unless sample sizes are limited to an upper maximum. In the Canadian EEM programs for pulp and paper and mining, sample sizes for sentinel fish populations could be limited to the number of fish required to detect effects of specified magnitudes (e.g., 25% for gonad size). However, because the gonad size versus body size relationship tends to be the most variable of the endpoints examined, small differences in the other endpoints are often statistically significant. With particularly large sample sizes, it is possible to get small changes that are statistically significant, and large changes that are not (Figure 3). A statistical difference "*alone*" can not be used to identify relevance, since it will " *partly* " be a consequence of study design, sample size and variability.

More than two reference standard deviations from the mean

The use of the "2 standard deviation (SD) in reference area" philosophy for setting critical effect sizes for fish population endpoints has not been fully explored to date (Kilgour et al., 1998), although it has proved satisfactory for benthic invertebrate community EEM data (Lowell et al., 2003). Data from pulp mill studies in Canada show a general relationship between effect sizes expressed as percentage differences and in reference area standard deviation units: the larger the percentage difference between reference and exposure populations, the larger the difference is in reference area standard deviation units (as one might expect) (Figure 4). For gonad size, differences of about 50% between reference and exposure populations coincides with an effect of roughly 2 SDs. Using the 2 SD concept would, therefore, result in targeting the detection of effects that are considered, a priori, to be very large, thereby missing large effects that are less than 2 SD.

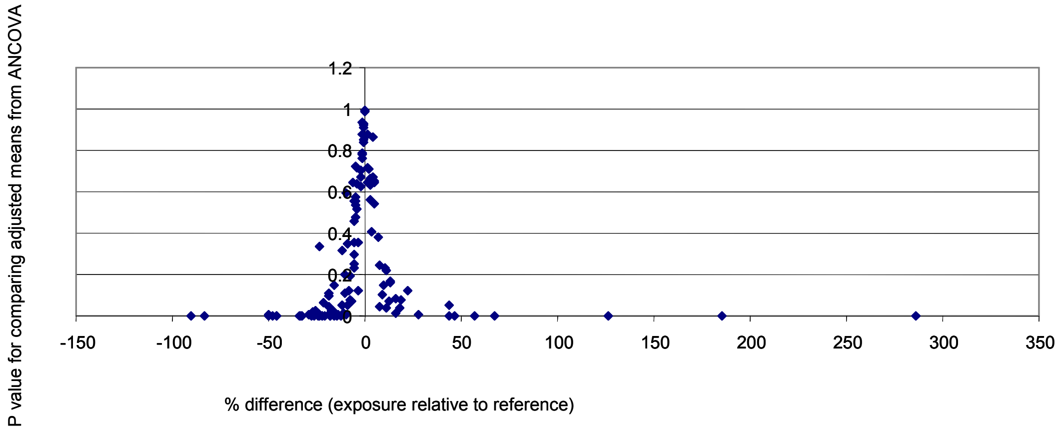

Figure 3. Effect size for gonad differences (cycle 2 corrected electronic data) versus the probability of detecting a difference (p value from ANCOVA model).

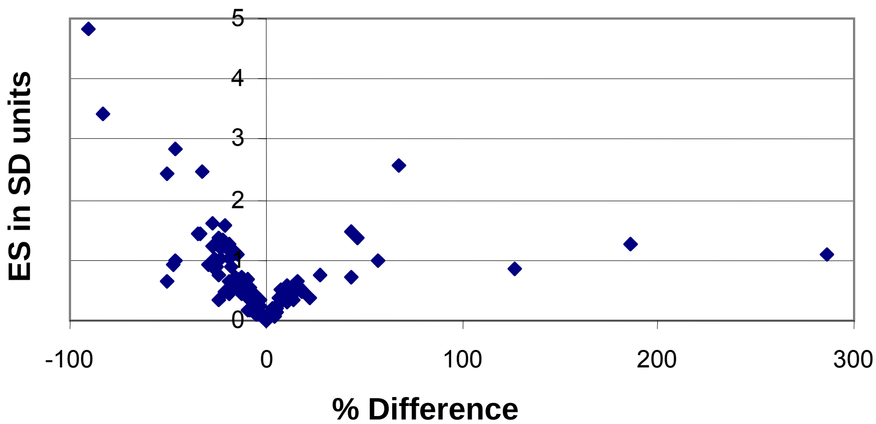

Figure 4. Scatterplot of effect sizes expressed as percent difference and in reference area standard deviation (SD) units for gonad size. Data are from corrected, electronic Cycle 2 EEM for pulp and paper.

Outside the range of natural variability

Whereas the previous section addresses natural variability *within studies*, there have been several investigations that have cited the importance of using natural variability *among studies* to identify the range of expected response levels. But this latter approach has definite disadvantages, including that endpoints like gonad size change seasonally, and that it is difficult to define what range of reference sites should be considered to represent “natural” variability. Furthermore, large differences between reference sites can occur when those sites do not have comparable habitat. A review of cycle 1 data for white sucker studies shows that there is enormous variability between sites in gonad sizes (Figure 5). This is not surprising since there is not any consistency between studies in the timing of collections. In most cases, however, you cannot be certain that you are capturing them at the same point in gonad maturation, because gonad maturation is affected by things like temperature, or in some species flow, and probably in some species changes in flow, and you have no way of knowing how much those things are advancing or delaying gonad maturation in a given year. Similar variability can be seen between years at the same sites (Figure 6). And endpoints like condition, which shows less variability within a season than gonad size for white sucker (*Catostomus commersoni*) still show inter-site differences in excess of 60% (Figure 5b). This has previously been reported even within a single province like New Brunswick (Galloway et al., 2003) and Ontario (Munkittrick et al., 1994). Exceeding the normal range of variability would require that about half the observed fish in an exposed site actually have a condition that is outside the normal range of variability observed in reference sites. For tangible endpoints like gonad size, liver size or condition, this is less acceptable than it is for less tangible endpoints like indices of benthic community composition where the use of the normal range is generally accepted (e.g., Kilgour et al., 1998).

A. GSI

B. Condition

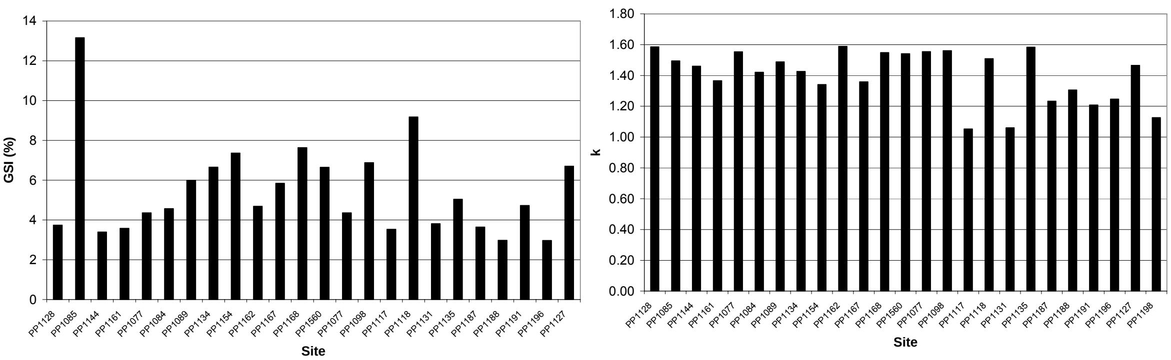

Figure 5. Between site variability in reference size gonads (GSI+100*gonad weight/body weight) and condition (k=100*wt/length^3) for corrected cycle 1 electronic data for white sucker females collected during different studies.

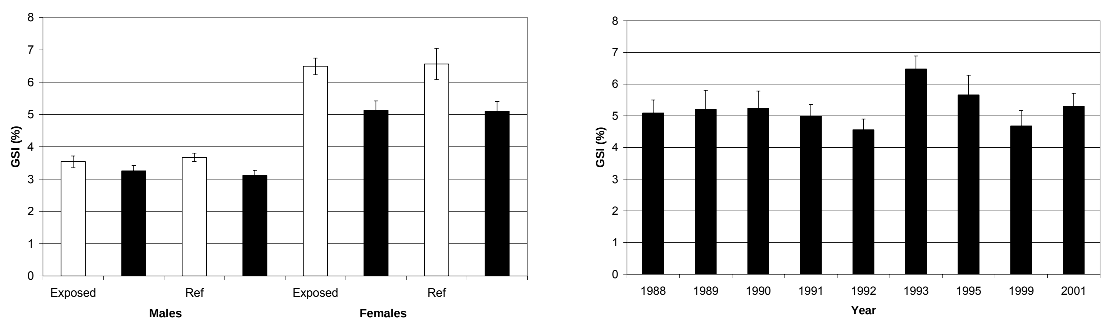

Figure 6. Between year variability in gonad size for (A) GSI in fall male and female white sucker from reference sites near Kenora, ON in 1999 (open bars) and 2003 (closed bars) (C. Portt, unpubl. Data) and (B) GSI in prespawning male white sucker from Mountain Bay, Lake Superior (K. Munkittrick and M. McMaster, unpubl. Data).

Arbitrary number

Beyond an arbitrary number that could include either a) a threshold that would trigger a regulatory response, b) a published number that others have defined as an impact at the population or community level, c) a level relative to other responses seen, or d) a number negotiated through multi-stakeholder consultation. There is not good information available at the current time to base a criterion on approaches reflecting regulatory responses. Analysis would require reviewing the regulatory decisions that have been made as a consequence of fisheries information, and relevant effects sizes, and making a decision regarding how large a difference in EEM endpoints has been associated with decisions. There are also insufficient numbers of studies that have associated changes in EEM endpoints that have been associated with changes at the population and community level.

A level relative to other responses seen

It may be possible to select an arbitrary number based on data collected during a cycle of monitoring. Figure 7 shows cycle 2 fish data for the pulp and paper EEM plotted in a multi-dimensional scaling ordination (Lowell et al., 2003). Points drawn to the bottom right represent sites with larger gonads, larger livers and larger condition; sites drawn to the upper left represent sites where fish had smaller livers, smaller gonads and smaller condition. The circle encompasses the 90% of sites closest to the origin of the plot. Under this approach, sites outside the circle would represent the 10% of the mills with the worst impacts in the country. Disadvantages to this approach include that the circle would depend on which data were included, and that the magnitude of impact determined to be "important" would vary from cycle to cycle.

A number negotiated through multi-stakeholder consultation

One of the major challenges involved in determining the existence, importance and causes of changes is achieving consensus between stakeholders on how to interpret the results. Although there have been descriptions of the impacts of pulp mill effluent on sexual maturation and reproductive development in fish for over 15 years (Munkittrick et al., 1991), there has not been consensus on whether these changes are real (Munkittrick et al., 2003), consistent (Kovacs et al., 1997), or important (Munkittrick, 2004). The inability to reach consensus on the existence of impacts, or their causes, is based on a number of factors, including a lack of agreement on the importance of measurement endpoints (Munkittrick and Sandström, 2003) and confusion about the relevance of changes.

These challenges may be avoided by making fundamental changes in the questions that are asked of the data. It is not possible to easily get consensus on the existence of impacts, and the effects-based approach subdivides the question of "Is there an impact?" into a series of questions that simplify interpretation, and allow an increased possibility of consensus. Rather than simply determining whether there are adverse impacts, or damage, or a problem, the important questions driving the approach include: are there significant changes present, are they consistent between years, are they getting better or worse, is the current situation sustainable, is it acceptable, and what are the consequences in terms of future developments or climate events?

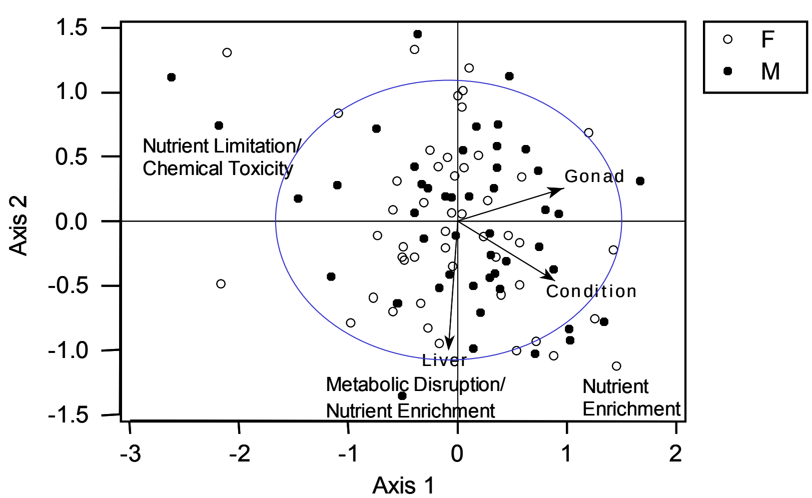

Figure 7. A MDS plot of cycle 2 fish data describing responses of gonad size, liver size and condition for male (M) and female (F) fish (adapted from Lowell et al., 2003). The circle encompasses approximately 90% of the study sites; sites further from the origin represent the "worst" sites for fish responses.

By separating the questions of sustainability and acceptability, it is possible to define a situation that we can expect to persist for some time without passing judgment on whether mitigation is necessary. The decision as to the acceptability of changes that are seen can be based on science or can be developed through consensus agreement on the nature of changes that will be socially or economically unacceptable. These decision points can be developed *a priori,* prior to data collection, or *a posteriori* through stakeholder consultation. In Sweden, a multi-stakeholder working group defined the changes associated with exposure to pulp mill effluent which would be considered unacceptable, including biochemical, individual, population and community-level changes (Swedish EPA, 1997). The level of change considered unacceptable after this process is not always defendable based on science, but participants have agreed to interpretation thresholds prior to data collection.

The EEM program has not currently defined the magnitude of change that will require mitigation. The site-specific relevance of changes detected will be dependent on a number of factors, including the magnitude and extent of changes, the species sensitivity, the number of endpoints responding, their trend over time, etc (Environment Canada, 2000). The relevance of the changes needs to be evaluated after sufficient data on extent and magnitude has been collected. There are differences in a 75% decrease on gonad size that extends for 50 m downstream, and a 20% difference that extends for 100 km. Under the EEM program, fish responses need to be beyond a defined size, and need to be confirmed, before monitoring is triggered into defining the geographic extent and magnitude (Environment Canada 2002, 2005a). That trigger is not to define whether changes are acceptable or not, but to define whether additional information is required.

The definition of acceptability needs to be strictly outlined in government policy, but has not been to date. The effect sizes described in this paper are used to trigger the need for more detailed studies, but do not address acceptability. They do not mean that effects beyond this size are unacceptable. Unacceptable changes could be defined based on a) levels of change which have previously resulted in regulatory action, b) statistically-based decisions as to the relative level of change requried to be ecologically relevant, or c) a government policy based on scientific research which defines the amount of damage that will be tolerated. It is the goal of the EEM process to provide the necessary amount of interpretable information on the types of changes, the distribution of changes, and the relevance of changes to other levels of organization so that an evaluation of sustainability can be conducted and relevance can be assessed.

There are some human impacts that most members of society would agree are clearly unacceptable, including the contamination of fishery resources with anthropogenic contaminants that limit their consumability, or impacts that have resulted in the loss of species that should be present in substantial numbers. Beyond those ‘severe' impacts, it becomes more difficult to define what is unaaceptable. The cumulatve effects literature deals with thresholds within which cumulative effects are either considered insignificant or acceptable, and beyond which significant or unacceptable effects are expected to occur (EIP, 1998). Even when the scientific process has defined that changes are acceptable, social concerns may define the impacts as exceeding a threshold that the public is comfortable with. Under CEAA, the definition of acceptable levels of change requires substantial public consultation that is consistent with a planning approach, and should consider economic and sociological considerations (Greig et al., 1998). A process needs to be described to reach a decision on the acceptability of changes, and to define what constitutes an unacceptable change.

The EEM process is currently completing a review under a Canadian “Smart Reg” process (Environment Canada, 2005b). In that process, stakeholders have suggested that 25% change in gonad size may be too conservative, and that a tighter guideline be adopted. For instance, in the current review for mills participating in Investigation of Cause for cycle 4 or cycle 5 pulp and paper EEM, a guideline of 10% reduction in gonad size over consecutive cycles has been adopted for potential inclusion. It may be that further multi-stakeholder negotiation derives a number different than those proposed here.

In the absence of this negotiated value, the existence of impacts at the whole organism level can not be interpreted as unacceptable without supporting data. There are economic and social impacts of overprotecting the environment through response triggers that are too restrictive. Significant changes in growth, reproductive development, energy storage and age distribution represent a “warning” that populations may be at risk for significant damage that could threaten the sustainability of the species. The magnitude of the change that represents a significant effect is still a matter of debate, and may be adjusted based on future science development. However, at the present time, the magnitude of change that represents a warning is 25 % for key fish endpoints reflecting gonad size, liver size, size-at-age and age distributions. Since the variability of values for condition factor are much lower than the other endpoints, a difference between sites of 10% represents a level of change that warrants further study. Data that exceed a warning level do not constitute an adverse effect on their own. However, based on a definition of sustainability that includes preserving the environment for future generations, a fundamental assumption is that the initial goal must be to reduce and eliminate environmental degradation Therefore, if monitoring over time has demonstrated that a key endpoint response is above a warning level, and getting worse, it must represent an unacceptable trend.

Responses that are increasing over time reflect evidence of environmental degradation, and should be unacceptable. If changes are unacceptable, then studies should proceed to include both a surveillance program and an investigation of cause.

It is possible to define unacceptable changes through three different mechanisms: (a) responses are deemed unacceptable to stakeholders based on consultative processes, b) responses in key endpoints are above the defined warning levels and getting worse over time; and c) community level impacts have occurred and one or more species that are expected to be present are missing.

Conclusions

There are a variety of potential approaches to defining an effect size that would be considered important for the purposes of EEM. We have reviewed these approaches using data from the Canadian EEM program, and suggest that changes in critical performance parameters in a such as growth rate, organ size and critical endpoints beyond 25% in a single monitoring cycle represent a change that reflects a that is a major concern (Table 1). Changes that are smaller than this but are consistent between cycles represent a concern that warrants further study. Critical effects sizes that would warrant management action may be larger, or may be represented by changes that are getting worse over time. Decisions regarding management action are site-specific and will be dependent on a number of factors, including the magnitude and extent of changes, the species sensitivity, the number of endpoints responding, their trend over time, etc.

Table 1. A comparison of methods for determining unacceptable effect sizes for EEM using different methods.

Method	Recommended	Magnitude
Best professional judgment regarding impacts of concern	??	25%
Rare responses previously seen in similar monitoring	??	25%
Outside of differences seen between comparable reference sites	??	25-30%
Based on tests of statistical significance	No	
More than two standard deviations	No	
Outside the range of natural variability	No	
Based on specific experience	Yes	20-25%
A threshold that would trigger a regulatory response	Possible	?? challenges
A published number that others have defined as an impact at the population or community level	Potential	?? challenges
A number based on responses relative to other sites	Yes	Variable
A number negotiated through multi-stakeholder consultation	Yes	Has not been done yet

Bibliography

Andersson, T., L. Förlin, J. Härdig and Å. Larsson. 1988. Physiological disturbances in fish living in coastal water polluted with bleached kraft mill effluents. Can. J. Fish. Aquat. Sci. 45: 1525-1536

Environmental Information Partnership (EIP). 1998. Cumulative effects assessment in the Moose Rive Basin - Background literature review. Prepared for the Environmental Information Partnership, South Porcupine, Ontario. Prepared by D.M. Abraham, ESSA Technologies Ltd., Richmond Hill, Ontario. 62 pp.

Environment Canada. 1997. Aquatic Environmental Effects Monitoring Requirements (revised EPS 1/RM/18). EEM/1997/1.

---- 2000. Guidance for determining follow-up actions when effects have been identified in EEM. Ottawa, ON. Draft document available at www.ec.gc.ca/eem/pdf_publications

------ 2002. Metal Mining Guidance document for Aquatic Environmental Effects Monitoring. Environment Canada, Ottawa ON.

----- 2005a. Pulp and paper EEM guidance document. Environment Canada, Ottawa, ON.

----- 2005b. Improving the effectiveness and efficiency of pulp and paper Environmental Effects Monitoring: A Smart Regulation Opportunity. Ottawa, ON

FSEWG (Fish Survey Expert Working Group). 1997. Recommendations from Cycle 1 review. Environment Canada, Ottawa, ON EEM/1997/6. 262 p.

Galloway, B.J., K.R. Munkittrick, S. Currie, M.A. Gray, R.A. Curry, and C.S. Wood. 2003. Examination of the responses of slimy sculpin (*Cottus cognatus*) and white sucker (*Catostomus commersoni*)

collected on the Saint John River (Canada) downstream of pulp mill, paper mill, and sewage discharges. Environ. Toxicol. Chem. 22: 2898-2907.
Greig, L.I., G.A. Duckworth, R. McCrea and C. Daniel. 1998. Conceptual framework & considerations for cumulative effects assessment in the Moose River Basin: Workshop Report. Moose River Basin Environmental Partnership, Ontario Ministry of Natural Resources, South Porcupine, Ontario.
Hodson, P.V., K.R. Munkittrick, R. Stevens and A. Colodey. 1996. A tier-testing strategy for managing programs of environmental effects monitoring. Water Pollut. Res. J. Can. 31: 215-224.
Keough, M. J., and B. Mapstone. 1995. *Protocols for designing marine ecological monitoring programs associated with BEKM operations.* National Pulp Mills Research Program, Technical Report 11. CSIRO, Canberra, 177 pp. ISBN 0 643 05847 8
Kilgour, B.W., K.M. Somers, and D.E. Matthews. 1998. Using the normal range as a criterion for biological significance in environmental monitoring and assessment. Écoscience, 5:542-550.
Kilgour, B.W., K.R. Munkittrick, C.B. Portt, K. Hedley, J. Culp, S. Dixit, G. Pastershank. 2005. Biological criteria for municipal wastewater effluent monitoring programs. Water Quality Research Journal of Canada 40:374-387.
Kovacs, T.G., J.S. Gibbons, P.H. Martel and R.H. Voss. 1997. Perspective on the potential of pulp and paper effluents to affect the reproductive capacity of fish: a review of Canadian field studies. Journal of Toxicology and Environmental Health 51: 305-352.
Lowell, R.B., S.C. Ribey, I.K. Ellis, E.L. Porter, J.M. Culp, L.C. Grapentine, M.E. McMaster, K.R. Munkittrick and R.P. Scroggins. 2003. National assessment of the pulp and paper environmental effects monitoring data. NWRI Publ. 03-521. National Water Research Institute, Burlington, ON 124 p.
Lowell, R.B., K.R. Munkittrick, J.M. Culp, M.E. McMaster, and L.C. Grapentine. 2004. National response patterns of fish and invertebrates exposed to pulp and paper mill effluents: metabolic disruption in combination with eutrophication and other effects. Pages 147-155 in *Pulp and Paper Mill Effluent Environmental Fate and Effects* , D.L. Borton, T.J. Hall, R.P. Fisher, J.F. Thomas, eds., DEStech Publications, Lancaster, PA.
Lowell, R.B., B. Ring, G. Pastershank, S. Walker, L. Trudel and K. Hedley. 2005. National Assessment of Pulp and Paper Environmental Effects Monitoring Data: Findings from Cycles 1 through 3. National Water Research Institute, Environment Canada. NWRI Scientific Assessment Report Series No. 5. 40 p.
Munkittrick, K.R. 2004. The evolution of study approaches with pulp mill effluents, 1991-2003. p xv-xxvii In: (D.L. Borton, T.J. Hall, R.P. Fisher and J.F. Thomas, Eds.) Pulp & Paper Mill Effluent Environmental Fate & Effects. DesTech Publications, Lancaster, PA.
Munkittrick, K.R. and O. Sandström. 2003. Ecological assessments of pulp mill impacts: issues, concerns, myths and research needs. Chapter 4.1, p 352-362 In Environmental Impacts of Pulp and Paper Waste Streams. T. Stuthridge, M. van den Heuvel, N. Marvin, A. Slade and J. Clifford (Eds.), Proceedings from the 3rd International Conference on Environmental Fate and Effects of Pulp and Paper Mill Effluents. Rotorua, New Zealand.
Munkittrick, K.R., G.J. Van Der Kraak, M.E. McMaster, C.B. Portt, M.R. van den Heuvel and M.R. Servos. 1994. Survey of receiving water environmental impacts associated with discharges from pulp mills. 2. Gonad size, liver size, hepatic EROD activity and plasma sex steroid levels in white sucker. Environ. Toxicol. Chem. 13: 1089-1101.
Munkittrick, K.R., M. McMaster, G. Van Der Kraak, C. Portt, W. Gibbons, A. Farwell and M. Gray. 2000. *Development of Methods for Effects-Based Cumulative Effects Assessment Using Fish Populations: Moose River Project.* SETAC Press, Pensacola, FL. 236 pp. + xviii
Munkittrick, K.R., S.A. McGeachy, M.E. McMaster and S.C. Courtenay. 2002. Overview of freshwater fish studies from the pulp and paper Environmental Effects Monitoring program. Water Quality Res J Can. 37: 49-77
Munkittrick, K.R., O. Sandström, Å. Larsson, G.J. Van Der Kraak, L. Förlin, E. Lindesjöö, M.E. McMaster and M.R. Servos. 2003. A reassessment of the original reviews of Norrsundet and Jackfish Bay field studies. Ch. 4.12, p 459-477 In Environmental Impacts of Pulp and Paper Waste Streams. T. Stuthridge, M. van den Heuvel, N. Marvin, A. Slade and J. Clifford (Eds.), Proceedings from the 3rd International Conference on Environmental Fate and Effects of Pulp and Paper Mill Effluents. Rotorua, New Zealand.
Sandström, O. and G. Thoresson. 1988. Mortality in perch populations in a Baltic pulp mill effluent area. Mar. Pollut. Bull 19:564-567.
Swedish EPA (Environmental Protection Agency). 1997. Environmental impacts of pulp and paper mill effluents: a strategy for future environmental risk assessments. Report 4785. ISBN 91-620-4785-X.
Walker, S.L., K. Hedley and E. Porter. 2002. Pulp and paper environmental effects monitoring in Canada: An overview. Water Quality Research Journal of Canada 37: 7-19