Preliminary Analysis of Pulp and Paper Environmental Effects Monitoring Data to Assess Possible Relationships between the Sublethal Toxicity of Effluent and Effects on Biota in the Field

Sherry L. Walker,1* Richard B. Lowell2 and James P. Sherry3

1National Environmental Effects Monitoring Office, National Water Research Institute, Environment Canada, 351 St. Joseph Boulevard, Gatineau, Quebec K1A 0H3
2National Environmental Effects Monitoring Office, National Water Research Institute, Environment Canada, 11 Innovation Boulevard, Saskatoon, Saskatchewan S7N 3H5
3National Water Research Institute, Environment Canada, 867 Lakeshore Road, Burlington, Ontario L7R 4A6

Data from the sublethal toxicity testing of effluents may or may not be predictive of field effects. Although qualitative studies have attempted to support a predictive relationship at select sites, few quantitative studies have been undertaken to establish whether general predictive relationships exist for diverse recipient environments. Since Canada’s Environmental Effects Monitoring (EEM) Program encompasses a strong field component as well as a suite of sublethal toxicity tests, the Cycle 2 data set of the Pulp and Paper EEM Program presented an opportunity to elucidate whether relationships exist between various sublethal toxicity endpoints used in EEM and field effects that were determined in surveys of benthic invertebrate communities and fish populations. Sublethal toxicity data and key endpoints from the fish (gonad weight, liver weight and condition) and invertebrate surveys (taxon richness and abundance) were quantitatively analyzed using simple bivariate correlation analysis. Our preliminary analysis of the data did not reveal any meaningful general relationships between the field biomonitoring and sublethal toxicity data collected under the Pulp and Paper EEM Program. Although the sublethal toxicity tests are useful to assess changes in effluent quality, their ability to predict the field effects for the key endpoints that are currently measured for fish and benthos in the Pulp and Paper EEM Program remains unsubstantiated.

Key words: toxicity tests, biological effects, laboratory to field extrapolation, field validation, pulp and paper mills

Introduction

Sublethal toxicity tests are widely used to measure effluent toxicity in the regulatory programs of Canada, the U.S. and other countries. Sublethal toxicity tests have the following suggested uses for the Environmental Effects Monitoring (EEM) Program (Environment Canada 1998b; Scroggins et al. 2002): (i) to measure changes in effluent quality as a result of changes in mill process or effluent treatment; (ii) to estimate, in multiple discharge situations, the relative contributions from various sources to any observed effects in the receiving environment; and (iii) to estimate the potential for effects in receiving waters. Effluent toxicity tests provide an opportunity to integrate interactions among complex mixtures of contaminants under controlled laboratory conditions, are highly standardized, and are cost effective (Chapman 2000). Also, since effluent toxicity tests are conducted in the laboratory, confounding factors caused by multiple discharges are avoided. The use of single species toxicity tests to predict effects on biota in the field has been criticized, however, because of uncertainty in making the following extrapolations: from single species to complex interactions at higher levels of biological organization, from surrogate test species to indigenous species, and from laboratory to field exposures (Cairns 1983, 1986, 1988a,b; Clements and Kiffney 1994). For those reasons, single species toxicity tests of effluent may not be an accurate surrogate measure(s) of the incidence of biological effects for all discharges to all receiving environments.

Under the 1992 Pulp and Paper Effluent Regulations (PPER), amended in the 2004 Regulations Amending the Pulp and Paper Effluent Regulations (RAPPER), of the Canadian Fisheries Act, pulp and paper mills are required to undertake an Environmental Effects Monitoring (EEM) Program in addition to meeting discharge limits for total suspended solids (TSS), biochemical oxygen demand (BOD) and acute lethality to rainbow trout. EEM is a cyclical monitoring and evaluation of impacts in waters into which effluents are discharged. This monitoring program encompasses a strong biological field component including both a fish population survey and benthic invertebrate community survey. In addition, sub-
lethal toxicity testing of effluent is conducted on a fish, invertebrate and algal species. EEM was included as a component of the PPER because of the large variability in mill processes and diversity of recipient environments across Canada, and the uncertainty that uniform discharge standards would adequately protect all receiving environments. The data collected through EEM provide a national database of quantitative endpoints from the biological surveys and a suite of sublethal toxicity tests for pulp and paper mills across Canada. This EEM database presents a unique opportunity to quantitatively examine whether general relationships exist between data from the laboratory sublethal toxicity tests and endpoints from the fish population and benthic invertebrate community surveys.

Previous studies that have used EEM pulp and paper mill data to examine potential relationships between sublethal toxicity endpoints and biological endpoints in the field have undertaken qualitative rather than quantitative comparisons, with the exception of a study by Sarakinos and Rasmussen (1998). To date, no studies have quantitatively examined EEM data to see whether such relationships are upheld for effluent discharges into diverse receiving environments. Moody (2002), in a report to Environment Canada, used two types of qualitative assessment techniques, the zone of potential effects (ZPE) and a novel lab-to-field (LTF) rating scheme, to compare the relationship between laboratory sublethal toxicity and field measurements for 16 Ontario pulp and paper mills. The relationship between the ZPE (based on the lowest IC25) and the extent of biological effects observed in field studies was rated strong, moderate or weak based on the proximity of the ZPE to the location of measured effects on fish or benthic invertebrates in the recipient environment. As well, the LTF rating scheme was created to assess the strength of the relationship between sublethal toxicity results and field measurements. A scale of 1 to 5 was used to rate the field results at each site based on the number of statistically significant field effects (exposure compared with reference) relative to the number of endpoints measured in the field. Similarly, the level of toxicity was rated on a 5-point scale. The strength of the relationships between the toxicity results and field measurements was then rated by observing the degree of similarity in the ratings. The terms strong, moderate or weak were assigned to the strength of the relationship if the ratings were equal, differed by 1 point, or differed by 2 or more points, respectively. Using both of these approaches, Moody (2002) found better relationships between Ceriodaphnia and Selenastrum sublethal toxicity test results and the measured field effects than for the fathead minnow toxicity test. Similarly, Borgmann et al. (2004) used the LTF to rate toxicity and field survey results for 16 mills in Ontario. Using a regression analysis of LTF scores, they reported a significant relationship between the Cerio-

daphnia reproduction tests and benthic invertebrate field survey measurements. They concluded, however, that there was insufficient data to determine if this approach could be used as a predictive tool.

In a quantitative analysis, Sarakinos and Rasmussen (1998) used a canonical correlation analysis to examine the ability of laboratory-derived toxicity thresholds to predict the response thresholds of an invertebrate community using EEM data (adjusted for effects of physical/chemical variables) from a single Quebec pulp and paper mill. They concluded that the laboratory-derived toxicity thresholds exceeded the response thresholds of the invertebrate community in the field and impacts were observed in areas where no impact was expected (i.e., sublethal toxicity tests were under-protective).

The need to assess the relationship between data from sublethal toxicity tests and field effects for the Pulp and Paper EEM Program has been identified on several occasions (Environment Canada 1997, 2002; EEM Science Committee 2003). Our purpose in the present study is to undertake a quantitative analysis of national Cycle 2 EEM data to examine whether there are general relationships between the data from the sublethal toxicity testing of pulp mill effluents and the biological endpoints that were measured in the recipient environments into which the effluents were discharged.

Methods

Sublethal Toxicity Testing

The EEM Program requires pulp and paper mills to conduct a battery of three sublethal toxicity tests on effluent twice per year (summer and winter) for each year of the EEM Cycle. The toxicity tests assess effects on: (1) early life stage development of fish, (2) invertebrate reproduction, and (3) algal toxicity (Table 1). Toxicity testing must follow prescribed test methods that specify conditions for quality assurance and quality control (QA/QC) to ensure high-quality data. Environment Canada (1998b, 2005) recommends that laboratories contracted by pulp and paper mills to conduct sublethal toxicity tests should be accredited by the Standards Council of Canada through the Canadian Association for Environmental Analytical Laboratories (CAEAL). Regional Environment Canada staff also undertake quality assurance checks of the toxicity data. Full descriptions of the sublethal toxicity testing requirements are provided in Environment Canada (1998b, 2005) and the test methods are referenced in Table 1.

The inhibiting concentration for 25% effect (IC25) is the effluent concentration that causes a 25% decrease in performance compared to the control organisms. The IC25 from the sublethal toxicity test for each species was selected as the value for comparison with the biological endpoints from the receiving environment. To provide a
single numerical value for use in the correlation analysis for each site, the geometric means (n ≤ 6) of the IC25s over Cycle 2 for each mill were calculated. For the rainbow trout early life stage embryo test, the endpoint was the similarly defined effective concentration for 25% reduction in embryo viability of the embryos (EC25).

**Fish Survey**

Under EEM, mills are required to conduct a fish survey to assess effects on fish growth, reproduction, condition and survival (Environment Canada 1998b, 2005). Minimum sample sizes of 20 males and 20 females are recommended from two sentinel species resident in the reference and receiving (i.e., exposure) environments. Required measurements include age, length, weight, liver size, gender, gonad size, fecundity, egg size and external condition. Alternatives to the fish survey, such as mesocosms or caged bivalves, are recommended for sites where use of the standard fish survey is not possible or appropriate. A guidance document was developed by Environment Canada (1998b, 2005) for use in the fish surveys to ensure consistency and quality of data.

We selected condition (weight relative to length; a measure of energy storage), gonad weight (relative to body weight; a measure of energy use) and liver weight (relative to body weight; a measure of energy storage) as the endpoints from the fish survey to be compared with the data from the sublethal toxicity tests (Environment Canada 1998b, 2005; Lowell et al. 2002). A preliminary evaluation of EEM fish data indicated that information for those endpoints was the most consistently reported among pulp and paper mills (Munkittrick et al. 2002). Those endpoints were identified in the National Assessment of Pulp and Paper Environmental Effects Monitoring Data as being particularly useful for interpreting fish responses to effluent exposure (Lowell et al. 2003). Briefly, data submitted electronically by mills and stored in the national EEM database were screened for errors and exported into spreadsheets for analyses. ANCOVA analyses, calculated for this national assessment, provided the mean values for each of these endpoints, adjusted for a size covariate. The covariates were fish length for the condition endpoint and fish weight for the gonad and liver endpoints. A full description of the methods and criteria can be found in Lowell et al. (2002, 2003). From this national data set, the effect size, as the percent difference between the adjusted mean values of the exposure versus reference areas for each of the three endpoints (condition, gonad weight and liver weight), was calculated. The effect size was then used as the quantitative value for comparison with the sublethal toxicity endpoint (IC25 or EC25).

Acceptable data were obtained from 47 or 48 freshwater mills depending on whether the respective endpoints of condition or liver weight/gonad weight were measured. For the mills discharging to estuarine or marine receiving environments, acceptable data were available for three or four mills depending on the endpoint (i.e., condition and gonad weight/liver weight, respectively). The data for each sentinel species and gender sampled were considered separate values. As such, there were potentially four data sets per mill (i.e., two sentinel species and two genders). Data from alternative approaches such as mesocosms and caged bivalves studies were excluded from our analysis.
Benthic Invertebrate Community Survey

The benthic invertebrate community survey is used in EEM to provide an indication of the health of the fish habitat. Structural changes in benthic invertebrate communities (i.e., abundance, number of taxa, shifts in the kinds, and dominance of taxa) can indicate impairment that results from contaminant or nutrient stressors. Briefly, five categories of sampling designs are recommended for community assessments in the EEM Program: (1) control/impact (CI) design, (2) multiple CI design, (3) simple gradient design, (4) radial gradient design, and (5) multiple gradient design. Guidance is also provided by Environment Canada (1998b, 2005) on selection of dominant habitat type and reference areas, sampling methods (i.e., sample size, effort, sorting and subsampling), QA/QC, taxonomy, reference collections and data analysis. Four endpoints are used to assess effects of pulp mill effluent on benthic invertebrate community structure: abundance of individuals, taxon richness (number of taxa), Simpson's evenness (how evenly individuals are distributed among taxa) and the Bray-Curtis index of dissimilarity (change in overall community structure, particularly composition). A detailed description of the benthic invertebrate community survey for the Pulp and Paper EEM Program is provided in Environment Canada (1998b, 2005).

To examine the potential relationship between the laboratory toxicity data and field results for the benthic invertebrate community survey, two endpoints were considered (taxon richness and abundance). Those endpoints were identified in the National Assessment of Cycle 2 EEM data as indicative of the predominant response patterns observed at mills across Canada (Lowell et al. 2003). Similar to the analysis of the fish survey, the data from the invertebrate community survey that were submitted by pulp and paper mills were first screened for errors. Area means were then calculated as part of the ANOVA undertaken for the national assessment (see Lowell et al. [2002, 2003] for a full description of methods). From this analysis, the effect sizes for both taxon richness and abundance were quantified as the percent difference between exposure and reference areas for each mill. The effect sizes were then used as the quantitative point for comparison with the sublethal toxicity endpoint (IC25 or EC25). Acceptable invertebrate community survey data were obtained for 53 freshwater and six estuarine/marine mills.

Comparison of Biological Field Effects with Sublethal Toxicity Data

Sublethal toxicity data and field data from Cycle 2 (1997–2000) of the Pulp and Paper EEM Program were analyzed using simple bivariate correlation analysis and scatter plots to assess the degree of association between the laboratory sublethal toxicity tests and biological measurements in the field studies. Correlation analyses were undertaken on the untransformed data, on the absolute values of the field endpoints, and on the log/log transformations of both the field biomonitoring and sublethal toxicity test endpoints. Estuarine/marine and freshwater data were initially pooled and analyzed together to maximize the number of comparisons. Thus, the toxicity data for fathead minnow (Pimephales promelas), rainbow trout (Oncorhynchus mykiss), inland silverside (Menidia beryllina) and topsmelt (Atherinops affinis) were combined in the initial analysis. Similarly for invertebrates, toxicity data for Ceriodaphnia dubia and echinoids were initially combined. For the algae, the toxicity data for freshwater alga, Selenastrum capricornutum, and the marine red macro alga, Chlamydomonas parvula, were combined. The freshwater data were also analyzed separately from the marine data. There were insufficient acceptable marine data to warrant a separate analysis.

To examine the influence of dilution effects in the recipient environment on the relationship between the data from the laboratory toxicity tests and the biological data from the fish and benthic invertebrate surveys, we attempted to obtain dilution factors for each mill. Submission of dilution information, however, was not required for Cycle 2 and although some dilution data were available for Cycle 3, such data were reported inconsistently among mills. In addition, it was determined in preliminary analysis of Cycle 3 data that field sampling locations for many mills had changed from Cycle 2 and thus use of the Cycle 3 dilution data could introduce errors in our analyses. Only near-field data from both control-impact (CI) and multiple CI studies (the most commonly used designs) were used to minimize variability in effluent exposure concentration among mills (i.e., to minimize variability in dilution), and to maximize the chances of finding a laboratory-to-field relationship, if one existed. Mills will be required to report effluent dilution factors for Cycle 4 and future cycles. Gradient designs, which were used if there was a high degree of variability in dilution within the near-field area, and alternative studies (mesocosms), were excluded from the analysis.

Results

Endpoints from the feral fish surveys were compared against both fish and algal sublethal toxicity results (IC25). Similarly, invertebrate field results were compared against both invertebrate and algal sublethal toxicity tests. Correlation coefficients (r) for ten laboratory-to-field comparisons using the combined freshwater and marine data are presented in Table 2. Significant (p < 0.05), although small, r values were found for two of the 10 comparisons: condition factor versus fish toxi-
city and condition factor versus algal toxicity. Inspection of the scatter plots revealed no apparent relationships between the field endpoints and IC25 for either untransformed (Fig. 1–5) or transformed data (not shown).

For the combined freshwater and marine data, significant but small inverse r values were found for the comparisons of absolute value of condition factor versus fish sublethal toxicity ($r = -0.202$, $n = 115$, $p < 0.05$; Table 2) (pooled fathead minnow, rainbow trout, inland silverside and topsmelt) as well as for the comparison of log/log transformed condition factor (absolute value) and fish sublethal toxicity ($r = -0.259$, $n = 115$, $p < 0.01$; Table 2) (Fig. 1). The r values for liver weight and gonad weight versus fish sublethal toxicity were not significant (Fig. 2 and 3).

The r values were not significant for the comparisons between the combined marine and freshwater benthic invertebrate survey and the invertebrate sublethal toxicity data (Ceriodaphnia dubia and echinoderms) for the two endpoints considered (taxon richness and abundance; Fig. 4 and 5).

For the comparisons of the fish and benthic invertebrate endpoints (combined freshwater and marine) against the algal toxicity data (Selenastrum capricornutum and Champa parvula), an inverse r value, which though small was significant, was found for untransformed fish condition and algal sublethal toxicity ($r = -0.198$, $n = 117$, $p < 0.05$; Table 2). Correlations were not significant for the other feral fish endpoints (liver weight or gonad weight) or for the invertebrate endpoints (taxon richness and abundance) against the algal sublethal toxicity data. Scatter plots of the benthic invertebrate and fish endpoints versus the algal sublethal toxicity data did not reveal any apparent relationships (not shown).

Similar results were obtained when the data for freshwater mills were analyzed separately (Table 3). Significant, although small, inverse relationships were found for the comparison of the absolute value of fish condition versus fish sublethal toxicity ($r = -0.207$, $n = 107$, $p < 0.05$; Table 3) and the log/log transformation of this same comparison ($r = -0.249$, $n = 107$)

---

**TABLE 2.** Correlation coefficients (r) for comparisons of field endpoint versus sublethal toxicity (IC25) (combined freshwater and marine data)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Untransformed</th>
<th>Absolute value</th>
<th>Log field endpoint$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>field endpoint</td>
<td>of field endpoint</td>
<td>vs. IC25</td>
</tr>
<tr>
<td>Condition vs. fish toxicity ($n = 115$)$^b$</td>
<td>-0.136</td>
<td>-0.202$^c$</td>
<td>-0.259$^d$</td>
</tr>
<tr>
<td>Liver weight vs. fish toxicity ($n = 120$)</td>
<td>-0.104</td>
<td>0.016</td>
<td>0.035</td>
</tr>
<tr>
<td>Gonad weight vs. fish toxicity ($n = 118$)</td>
<td>0.174</td>
<td>0.026</td>
<td>0.056</td>
</tr>
<tr>
<td>Invertebrate abundance vs. invertebrate toxicity ($n = 59$)</td>
<td>-0.036</td>
<td>-0.056</td>
<td>0.043</td>
</tr>
<tr>
<td>Invertebrate taxon richness vs. invertebrate toxicity ($n = 59$)</td>
<td>0.161</td>
<td>-0.004</td>
<td>-0.146</td>
</tr>
<tr>
<td>Condition vs. algal toxicity ($n = 117$)</td>
<td>-0.198$^c$</td>
<td>-0.188</td>
<td>-0.191</td>
</tr>
<tr>
<td>Liver weight vs. algal toxicity ($n = 122$)</td>
<td>-0.060</td>
<td>-0.029</td>
<td>-0.085</td>
</tr>
<tr>
<td>Gonad weight vs. algal toxicity ($n = 120$)</td>
<td>-0.009</td>
<td>0.013</td>
<td>-0.003</td>
</tr>
<tr>
<td>Invertebrate abundance vs. algal toxicity ($n = 60$)</td>
<td>-0.215</td>
<td>-0.235</td>
<td>-0.141</td>
</tr>
<tr>
<td>Invertebrate taxon richness vs. algal toxicity ($n = 60$)</td>
<td>0.001</td>
<td>-0.143</td>
<td>-0.218</td>
</tr>
</tbody>
</table>

$^a$Log values were calculated from absolute values for field endpoints.

$^b$n; Number of comparisons.

$^c$Significant at $p < 0.05$.

$^d$Significant at $p < 0.01$.

---

**Fig. 1.** Fish condition (percent difference between exposure and reference sites) versus fish sublethal toxicity (combined marine and freshwater).

**Fig. 2.** Fish liver weight (percent difference between exposure and reference sites) versus fish sublethal toxicity (combined marine and freshwater).
p < 0.05). Significant inverse relationships were also found between fish condition and the algal (*Selenastrum*) sublethal toxicity test, although r values were low, for the untransformed comparison ($r = -0.232$, $n = 109$, $p < 0.05$), the absolute value of fish condition versus toxicity ($r = -0.214$, $n = 109$, $p < 0.05$), and the log/log transformation ($r = -0.194$, $n = 109$, $p < 0.05$). As well, significant inverse relationships were found between *Selenastrum* toxicity and invertebrate abundance for both the untransformed ($r = -0.287$, $n = 54$, $p < 0.05$; Fig. 6) and the comparison of absolute values of invertebrate abundance and the *Selenastrum* IC25 ($r = -0.283$, $n = 54$, $p < 0.05$). Correlation coefficients for the other comparisons analyzed for freshwater fish (liver weight and gonad weight) versus fish sublethal toxicity, invertebrate community endpoints (taxon richness and abundance) versus *Ceriodaphnia* sublethal toxicity, and invertebrate community taxon richness versus *Selenastrum* toxicity were not significant.

**Discussion**

There has been considerable controversy and discussion in the literature about the ability of effluent toxicity tests to predict biological impacts in recipient environments. Although some key articles are briefly discussed below, the reader is referred to recent literature reviews in the “grey” and published literature for a comprehensive discussion (Sprague 1997; de Vlaming and Norberg-King 1999; Chapman 2000; Diamond and Daley 2000; LaPoint and Waller 2000). In an evaluation that was prepared for the U.S. EPA of the ability of single species toxicity tests to predict community responses in aquatic ecosystems, de Vlaming and Norberg-King (1999) concluded that laboratory single species effluent toxicity tests are reliable qualitative predictors of ecosystem impacts. They noted that much of the criticism of single species toxicity tests has been that only qualitative (i.e., toxicity test results are associated with some degree of biological community impairment) rather than quantitative (i.e., a specific percent or degree of toxicity test response can be related to a specific percent or degree of impairment in the biological communities) relationships have been established. Similarly, Sprague (1997) reviewed 29 published studies yielding 73 comparisons (mostly qualitative) and found that there were 53 cases of general agreement between laboratory results and...
Comparisons of EEM Laboratory to Field Data

TABLE 3. Correlation coefficients (r) for comparisons of the field endpoints versus sublethal toxicity (IC25) (freshwater only)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Untransformed field endpoint vs. IC25</th>
<th>Absolute value of field endpoint vs. IC25</th>
<th>Log</th>
<th>field endpoint</th>
<th>vs. log IC25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition vs. fish toxicity (n = 107)b</td>
<td>-0.140</td>
<td>-0.207c</td>
<td>-0.249c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver weight vs. fish toxicity (n = 112)</td>
<td>-0.075</td>
<td>0.027</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gonad weight vs. fish toxicity (n = 111)</td>
<td>0.157</td>
<td>0.028</td>
<td>-0.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrate abundance vs. invertebrate toxicity (n = 53)</td>
<td>-0.082</td>
<td>-0.089</td>
<td>0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrate taxon richness vs. invertebrate toxicity (n = 53)</td>
<td>0.063</td>
<td>0.073</td>
<td>-0.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition vs. algal toxicity (n = 109)</td>
<td>-0.232c</td>
<td>-0.214c</td>
<td>-0.194c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver weight vs. algal toxicity (n = 114)</td>
<td>0.002</td>
<td>-0.012</td>
<td>-0.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gonad weight vs. algal toxicity (n = 113)</td>
<td>-0.062</td>
<td>0.021</td>
<td>0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrate abundance vs. algal toxicity (n = 54)</td>
<td>-0.287c</td>
<td>-0.283c</td>
<td>-0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invertebrate taxon richness vs. algal toxicity (n = 54)</td>
<td>-0.195</td>
<td>-0.092</td>
<td>-0.105</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aLog values were calculated from absolute values of field endpoints.
bNumber of comparisons.
cSignificant at p < 0.05.

field, and ten cases of disagreement, for 84% agreement. Sprague excluded ten other cases, however, because they were comparisons of values derived from the literature, because field effects were absent or inconsistent, or because they considered only lethal effects. La Point and Waller (2000) noted that most comparisons between effluent toxicity tests and biological responses in the field have focused on benthic macroinvertebrates and on effluent-dominated streams in which effluents demonstrate little or no significant acute toxicity, whereas few studies have examined the predictability of effluent toxicity tests for other ecosystems (e.g., large rivers, estuaries) or other biota (e.g., fish, primary producers). As well, La Point and Waller (2000) suggest that toxicity tests and chemical-specific measurements are inadequate for the prediction of potential environmental impacts, particularly for moderately toxic effluents under low dilution scenarios. Diamond and Daley (2000) indicated, in a comprehensive analysis of 250 discharges in the U.S., that fish acute and chronic endpoints were more related to in-stream biological condition than Ceriodaphnia effluent toxicity test endpoints; however, no single endpoint was capable of accurately portraying in-stream conditions for all discharges. Chapman (2000) assessed the general status of effluent toxicity tests and found that, although effluent toxicity tests are useful to identify a hazard, comparisons of effluent toxicity results to field responses indicate that effluent toxicity tests are not reliable predictors of the incidence (or lack) of effects in recipient environments.

Early attempts by eight U.S. EPA-sponsored studies under the Complex Effluent Toxicity Testing Program (CETTP) to validate the ability of effluent and ambient toxicity tests to predict effects in aquatic ecosystems (Mount et al. 1984, 1985, 1986a,b,c; Mount and Norberg-King 1985, 1986; Norberg-King and Mount 1986) were criticized because of study design limitations, lack of replication, limited geographical distribution, pseudoreplication and lack of statistical analysis (Dickson et al. 1992; Marcus and McDonald 1992). Both Marcus and McDonald (1992) and Dickson et al. (1992) re-evaluated the CETTP and associated studies (Birge et al. 1989; Dickson et al. 1996). Dickson et al. (1992), using robust canonical correlation analysis and classification methods, reported statistically significant relationships between ambient toxicity (Ceriodaphnia reproduction and fathead minnow growth and survival) and in-stream impacts on fish and benthic richness. Marcus and McDonald (1992), however, criticized the CETTP studies because of their sampling design and the selection of effluent-dominated streams where toxic impacts were expected to occur. Those shortcomings, they argued, prevent statistically based inferences to other sites or to other times at the selected sites. Marcus and McDonald's (1992) re-analysis of the data concluded that ambient toxicity tests can provide useful information about biological community structure in stream waters where ambient toxicity has a controlling influence. Those authors stated, however, that it would require more research to establish general relationships between whole effluent toxicity and field effects in aquatic communities.

Our preliminary analyses revealed no meaningful general relationships between the laboratory sublethal toxicity data and biological survey data for fish and benthic communities collected under the Pulp and Paper EEM Program. Differences in the dilution and modification of effluents by the receiving environment might explain the absence of apparent relationships. Both abiotic and biotic factors can modify an effluent's toxicity in the field (Marcus and McDonald 1992; Clements and Kiffney 1996; Chapman 2000). Abiotic factors include climate, temperature, environmental quality, nutrient loadings, dissolved oxygen, physical habitat limitations and the possible presence of other stressors. Biotic fac-
tors include species and life stage, sex and reproductive status, nutritional and disease status, competition and predation (Marcus and McDonald 1992; Chapman 2000). In addition, Clements and Kiffney (1996) discuss the challenge of separating natural variation from anthropogenic disturbance, as well as bias caused by endpoint selection, temporal variation, variation in sensitivity among locations, acclimation and adaptation, and other indirect effects.

Since pulp and paper mills require large volumes of water for processing, they tend to be located in Canada on large water bodies or rivers, thus dilution might be one reason that we did not find a relationship between field and laboratory data in the present study. To reduce the differences among mills in dilution, the present analyses were restricted to near-field areas. We recommend that the dilution hypothesis be further explored using Cycle 4 EEM data, which will require submission of effluent concentration data for each station. Examination of the results of our analyses, however, suggests that dilution likely only plays a partial role in explaining our results. If dilution was the major factor in our inability to identify an existing relationship, then one would expect to find scatter and outliers around an apparent trend of increasing impact in the field associated with increasing toxicity as measured by the sublethal toxicity tests. The scatter and outliers would correspond to mills that showed toxicity in the sublethal toxicity tests but no, or reduced, field effects because of dilution. We see no evidence of such a relationship. Instead we find basically flat relationships (slope = 0) or slopes in the opposite direction to that predicted by the laboratory tests. The latter is indicative of a weak negative relationship, which suggests that differences in dilution among mills is not the only factor limiting a potential relationship between laboratory and field data. These findings are likely related to sublethal toxicity tests (conducted in the laboratory) not accounting for the multitude of indirect and interacting effects that occur in the field (see previous paragraph), and to some additional factors that we discuss below. Similarly, Diamond and Daley (2000) found poor overall agreement between effluent toxicity test endpoints (Ceriodaphnia and fathead minnow) and results of in-stream benthic invertebrate assessments for 250 discharges in the U.S. They indicated, however, that better relationships were found between effluent toxicity tests and observed in-stream biological effects at sites where the recipient environment was comprised of >80% effluent. Effluents that comprised <20% of the stream were associated with a low probability that effluent toxicity tests would predict field effects.

Another possible reason for the absence of apparent relationships in the present study is the difference in endpoints between the sublethal toxicity tests and the types of responses observed in the field. In the national assessment of the EEM pulp and paper data, Lowell et al. (2003, 2004) found that moderate to pronounced eutrophication was the predominant effect for benthic invertebrates at a majority of sites. For the fish field survey, the predominant response pattern (decrease in gonad weight and increases in liver weight, condition factor and weight at age) was indicative of some form of metabolic disruption or impairment of endocrine functioning in combination with a nutrient enrichment effect (Lowell et al. 2003, 2004). The sublethal toxicity tests used in the EEM Program, however, are neither designed to detect a nutrient enrichment effect nor an endocrine disruption response (fish toxicity test endpoints are growth and survival for fathead minnow, topsmelt and inland silverside and embryo survival for rainbow trout; Table 1). La Point and Waller (2000) also noted that effluent toxicity tests do not address eutrophication effects in receiving systems. Since effects on gonad size are indicative of an endocrine disruption type of response, there is no mechanistic reason to suggest this response would be detected by the existing toxicity tests. The lack of mechanistic links between the laboratory tests and the field effects may be another factor limiting our ability to find relationships between the laboratory and field data.

Although significant correlation coefficients (r) were found for feral fish condition factor versus both fish sublethal toxicity and algal sublethal toxicity, as well as for freshwater invertebrate abundance versus Selenastrum toxicity, the low r values are indicative of weak negative relationships. The data patterns suggest that the significant correlations were an incidental consequence of the large sample sizes and large number of analyses that were run. It appears that the significance levels were driven by a subset of high positive values of condition factor at low IC25s, indicating higher condition of feral fish in association with higher levels of toxicity measured in the laboratory. Similarly, higher freshwater invertebrate abundance was associated with higher algal toxicity measured in the laboratory. These associations are the opposite of what would be expected if there was a general predictive relationship between toxicity in the laboratory tests and toxic effects in feral fish or benthic invertebrates. In addition, we would expect that a predictive relationship would yield much higher r values. As a result of the low r values, we did not undertake further analysis using multivariate statistical techniques.

For the fathead minnow test, there was a high reporting (e.g., ~50% of test results) of no toxic effects (i.e., IC25 >100) for mills where effects on liver weight, condition factor and/or gonad weight were reported for feral fish. Although the invertebrate sublethal toxicity tests for reproduction and survival of Ceriodaphnia dubia and the fertilization assay using echinoids (sea urchins and sand dollars) appeared to be more “sensitive” (i.e., a higher incidence of toxicity) to effluent toxicity, our analyses revealed no relationship with the ben-
thic invertebrate community survey endpoints of invertebrate abundance and taxon richness.

In conclusion, our preliminary analysis of the Cycle 2 data for pulp and paper mills did not find any meaningful general relationships between the sublethal toxicity tests and field effects for the Pulp and Paper EEM Program. Sublethal toxicity testing, however, has proved useful to verify improvements in effluent quality as a result of treatment and process changes (Lowell et al. 2003). Our preliminary analysis does not support the ability of the current sublethal toxicity tests used in EEM to predict the key field effects (i.e., benthic abundance and richness, fish condition, gonad weight and liver weight) measured for the Pulp and Paper EEM Program.

Acknowledgements

We would like to thank Bonna Ring for her assistance with summarizing the Cycle 2 field data and Jennifer Miller and Rick Scroggins for summarizing the sublethal toxicity data.

References


Mount DI, Norberg-King TJ. 1985. Validity of effluent and ambient toxicity tests for predicting biological impact, Scipio Creek, Circleville, Ohio. EPA/600/3-85/044, U.S. Environmental Protection Agency, Office of Research and Development, Duluth, Minn.


Norberg-King TJ, Mount DI. 1986. Validity of effluent and ambient toxicity tests for predicting biological impact, Skeleton Creek, Enid, Oklahoma. EPA/600/6-91/003, U.S. Environmental Protection Agency, Office of Research and Development, Duluth, Minn.


Received: February 9, 2005; accepted: July 25, 2005.