Repeatability and Reproducibility of the Luminescent Bacteria Bioassay

Juan M. Ribo,[1] Miguel Canela,[2] Eulalia Griful[3]

[1]Institut Català de Tecnologia, Pl. Ramon Berenguer 1, 08002 Barcelona, Spain

[2]Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain

[3]Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Colom 11, 08222 Terrassa, Spain

Received 23 *September* 1999; *revised* 31 *January* 2000; *accepted* 24 *March* 2000

ABSTRACT: This paper presents the statistical analysis of the results of an inter-laboratory study for the luminescent bacteria toxicity bioassay. It also contains a discussion on the statistical methods for the presentation and refinement of the evaluation of the precision of an assay method (including intra- and inter-laboratory variability), with special emphasis on the rejection of outliers, and the use of standardized parameters, like the repeatability and reproducibility values. @ 2001 by John Wiley & Sons, Inc. Environ Toxicol 16: 127-135, 2001

Keywords: luminescent bacteria bioassay; inter-laboratory studies; repeatability; reproducibility; outliers; Mandel's *h* and *k* statistics

INTRODUCTION

Background

The general objective of inter-laboratory comparison exercises for the bioluminescent bacteria toxicity assay is the study of the variability of the assay when used to determine the toxicity of waste water samples. The EC_{50} value of an industrial water discharge is used to calculate a tax in terms of volume and quality of the waste water eliminated through the public waterworks system. The different sensitivities and variabilities of this assay towards different types of contaminants have already been discussed in the literature (Bulich, 1979). A better knowledge of the assay sensitivity towards specific contaminants would allow a better use of this test for regulatory purposes, i.e., stating that the environmental hazard posed by certain types of waste water must be assessed by means of other, more sensitive, toxicity assays using a different test organism.

In the literature examined, reproducibility and repeatability of the luminescent bacteria bioassay has been mainly expressed as a coefficient of variation (CV) without mention of statistical methods for the detection of outliers. Usually CV values for this assay are high compared with CV values for analytical chemistry determinations, although they are normal for bio- logical assays.

Objective

The aim of this study is to explore and discuss the use of new statistical parameters in this type of inter-laboratory comparison exercise, which would enable a better description of the inter- and intra-laboratory

Correspondence to: J.M. Ribo; *e-mail:* jribo@ictnet.es

© 2001 by John Wiley & Sons, Inc.

variability of the luminescent bacteria toxicity assay, thus giving to participant laboratories more information and a more useful tool to control the quality of their results and to improve their performance. To the best of our knowledge the statistical methods used here have not been applied to inter-Iaboratory comparison exercises for the luminescent bacteria toxicity assay.

The Luminescent Bacteria Toxicity Assay

The toxicity of a chemical substance can be evaluated by means of biological assays, that is, by checking its effects on living organisms or isolated organs. In bioassays, test organisms are exposed to toxic samples, and a biological response is measured, allowing the assessment of the toxicity of the sample. The luminescent bacteria assay was developed to provide industries with a fast and reliable method to measure the toxicity of wastewater effluents, so that the toxicity values could be used in making decisions regarding effluent discharge (Bulich, 1979). The Microtox assay (Microtox® is a registered trademark of Azur Environmental) is a commercial version used as a screening technique for aqueous samples.

This assay uses non-pathogenic luminescent marine bacteria *(Vibrio fisheri NRRL B-11177)* as test organisms. These bacteria emit light as a result of their normal metabolism, and the intensity of the light emitted is an indication of their metabolic activity .When they are exposed to a toxic substance, the light emission is reduced, so that the reduction in light intensity provides a measure of the toxicity of the sample. Its advantages, with respect to the alternatives, are the simplicity of operations, the speed, and the low cost. The commercial Microtox Toxicity Analyzer uses marine bacteria in lyophilized form, so that a standard source of the test organisms is guaranteed, and results from different laboratories are comparable.

Numerous comparative studies between Microtox and other aquatic toxicity bioassays have been carried out, both on single compounds and complex effluents, and, in general, high correlation has been obtained. For example ;correlation between Microtox toxicity values and those obtained on Fathead minnows, for complex environmental samples, range from r=0.41 (Lebsack *et al.,* 1981) to r=1.0 (Chang *et al.,* 1981). For the same assay, Curtis *et al.*(1982) obtained r=0.85 on 68 single organic chemicals, and Indorato *et al.* (1983), also on a series of single toxicants, found r=0.91. On several chlorinated aromatic compounds, EC_{50} values related to the reduction of bacterial luminescence in *Photobacterium phosphoreum* have been compared to those related to: (a) the inhibition of bacterial dehydrogenase in *Bacillus sp,* (b)the inhibition of spore germination in *Bacillus subtilis.* (c) the acute toxicity on the brown trout *(Salmo truta)* and bluegill *(Lepomis macrochirus),* and (d) the semi-chronic toxicity on the guppy *(Poecilia reticulata).* These comparisons produced r^2 statistics ranging from 0.73 to 0.93 (Ribó and Kaiser, 1983). In another study, Vasseur and Ferard (1984) report r^2 values higher than 0.86 between Microtox and *Daphnia magna* EC_{50} values of 39 industrial effluents.

Inter-laboratory Studies

Inter-laboratory studies are very valuable for assessing the precision of an assay. The precision can be evaluated through the intra- and inter-laboratory components of the variance. These components are the two terms obtained by decomposing an estimate of the variance of the assay:

$$s^2_{TOTAL} = s^2_{INTRA} + s^2_{INTER}$$

Estimates of the total variance and the two components are obtained by means of a nested analysis of variance (Mandel, 1991), and they are often translated into coefficients of variation (CV).

An equivalent approach, supported by organizations such as ISO (ISO, 1994), ASTM (ANSI/ ASTM, 1979 and 1989), and IUPAC (IUPAC, 1990), is based on repeatability and reproducibility values. The repeatability is the degree of coincidence between results from the same laboratory, while the reproducibility is the degree of coincidence between results not necessarily from the same laboratory. Because of random errors, they are expressed in terms of confidence limits. Thus, the repeatability value *(r)* is the 95% confidence limit for the absolute value (without sign) of the difference between two measurements of the same specimen, in the same laboratory. Assuming a normal distribution of the measure errors, r is given by (Mandel, 1991; ISO, 1994):

$$r = 2.8 s_{INTRA}$$

In a similar way, the reproducibility value *(R)* is the 95% confidence limit for the absolute value of the difference between two measurements of the same specimen, not necessarily ill the same laboratory, and an estimate is given by

$$R = 2.8 s_{TOTAL}$$

Such definitions allow for a straightforward interpretation: the difference between two results obtained in the same laboratory for the same specimen is significant if it is greater than r. R can be interpreted in the same way, but it applies to differences between

results from different laboratories. In most cases, both *r* and *R* depend on the true value of the magnitude measured, and therefore they are obtained for different levels of this magnitude.

Inter-laboratory Studies for the Luminescent Bacteria Assay

Inter-Iaboratory studies have shown a satisfactory reproducibility for the Microtox assay (Qureshi *et al.,* 1987). Earlier, a very complete study of reproducibility and interchangeability, between Microtox and other aquatic toxicity bioassays, was carried out by the Canadian Petroleum Association. The average inter-Iaboratory CV for the Microtox assay was 13%, with a maxi- mum of 31 %. On the other hand, the average coefficient of variation for the fish assay was 30%, with a maximum of 98% (Strosher, 1984).

The coefficient of variation in these studies depends on several factors. Variation in repeatability and reproducibility results from variation in experimental methods, sample preparation, and, more importantly, com- position of the samples. In the literature, the variations reported when samples are constituted by inorganic contaminants, or complex effluents, are higher than those reported for solutions of pure organic com- pounds. Indeed, in two independent studies, the CY obtained for the Microtox toxicity of single chemicals was 10% (Curtis *et al.,* 1982; DeZwart and Slooff, 1983). For the toxicity of wastewater effluents, CVs range from 3% to 12% (Vasseur and Ferard, 1984; Mat thews and Short, 1983).

Over the past seven years, we have performed a yearly inter-Iaboratory study for the luminescent bacteria assay, aimed to improve the quality of toxicity data for aqueous samples (Ribó and Diez, 1994; Ribó, 1997). These studies were round robins in which four series of five samples were sent to 20 laboratories. Results and no. of participants are presented in Table I.

TABLE I. Variation coefficients and participants in interlaboratory comparison studies with bioluminescent bacteria toxicity assay

Study No.	Year	No. Participants	Mean CV
1	1993	11	34.40%
2	1994	19	27.22%
3	1995	24	19.52%
4	1996	21	25.95%
5	1997	19	30.96%
6	1998	22	24.31%
7	1999	15	31.48%

Statistical Analysis in Inter-laboratory Studies

As they are obtained from sums of squares, estimates of the components of variance can be affected by outliers. In a series of observations, an outlier is an observation isolated from the others. To avoid subjectivity in the rejection of outliers, a number of statistical tests have been proposed, most of them assuming a particular pattern of appearance for the outliers. The ANSI/ASTM guide (ANSI/ASTM, 1987) presents some of these tests, covering various situations. In the context of quality assurance of laboratories, Taylor (1989) presents two classical tests, the Dixon and Cochran tests, also recommended in a former version of ISO 5725 standard. The last version of the ISO standard (ISO, 1994) replaces the Dixon test with a pair of tests usually referred to as Grubbs tests (Grubbs and Beck, 1972). An elementary introduction to Grubbs and Cochran tests can be found in Burke (1998).

The Mandel *h* and *k* statistics (Mandel, 1991; ISO, 1994), also recommended in the ISO standard, can be used to examine the global pattern of the laboratories- by-sample data, by assessing differences between groups of results. The *h* and *k* statistics are useful in graphical presentations of inter-laboratory studies, as they help to get a quick understanding of the differences between laboratories. The *k* statistic is related to the discrepancies in precision, and the *h* statistic is related to the discrepancies in mean value.

MATERIALS AND METHODS

Assay Methods

The test organisms for the Microtox assay are the luminescent bacteria supplied in freeze-dried form by the manufacturer. The bacteria are reconstituted with ultra-pure water to prepare the stock solution, which is kept at 5°C. Five aliquots of the stock suspension are then taken to prepare five bacterial suspensions in a saline diluent. According to the standard operational procedure, these solutions must be incubated at 15°C. The light emission of the bacterial suspensions is measured before and after exposure to five different dilutions of the sample, including a control sample (diluent alone). Since bacteria come from a marine environment, sample dilutions are always prepared by adding saline solution to provide osmotic protection.

From the measurements of the light intensity, a value of the gamma function (Γ) is obtained for each test solution. Gamma is the ratio of the light lost to the light remaining, and, in log-log scale, there is a linear relationship between gamma and the concentration of the sample. As an indicator of the

inhibition effect, gamma is preferred to the percentage of light inhibition (Johnson *et a[.,* 1974).

The toxicity of a sample is expressed in terms of the concentration of the sample which reduces the light output by 50% (i.e., $\Gamma=1$), which can obtained graphically by fitting a straight line to a log-log plot of gamma versus concentration or by means of regression analysis. This concentration is referred to as EC50 (effective concentration of the sample causing a 50% reduction in light emission, or median effective concentration), and depends on the test temperature, the time of exposure, and, mainly, on the chemical nature of the toxicants present in the sample.

Bioluminescent Bacteria Toxicity Assay

Participant laboratories were asked to perform the bioluminescent bacteria toxicity assay, following the protocol used in their own laboratory and described in the literature. They were asked to determine the sample toxicity without any manipulation of the samples sent to them; that is, no filtration, dilution, centrifugation, nor pH neutralization was expected to be necessary for the test. These provisions were made to discard, as much as possible, the variability due to sample handling. In consequence, all sample solutions were prepared in such a way that they could be tested directly.

Original samples sent to participants were prepared using deionized water where needed. As recommended by the Microtox basic test protocol, participants were required to add 22% aqueous NaCI (Microtox Adjusting Solution) to provide osmotic protection to the microorganisms. A 2% NaCl aqueous solution (Microtox Diluent) was recommended to prepare test sample solutions, according to the test protocol.

The use of commercial Microtox solutions and reagents was recommended to ensure standardization of the experimental procedure and to eliminate the variability due to the use of differently prepared re- agents. The quality of solutions and bacterial suspensions can be a source of variability, as well as the cleanliness of the glassware used, pipetting skills of the lab personnel, etc. The influence of these variability sources was not considered in the study.

Toxic Agents Tested

Five samples containing toxic agents were prepared, two from pure substances and three from real waste- water effluents. The pure substances chosen, namely, Zinc sulfate (40 mg/L in deionized water) and Phenol (10 mg/L in deionized water), are widely used as standards for the Microtox assay. The effluents came from industrial discharges: (a) effluent 1 from a plastics manufacturing plant, (b) effluent 2 from a chemical industry manufacturing surface agents, and (c) effluent 3 from a production plant producing pharmaceutical raw materials.

All samples were prepared by independent person- nel and stored at low temperature. They were tested initially by the coordinator of the study, to ensure proper concentration. Then they were coded and sent to the participants. As mentioned above, participants were required to test samples without any further treatment.

Organization of the Inter-Laboratory Study

Twenty-two laboratories were voluntarily enrolled for this study. A code name, known only by the coordinator of the study, was given to each participant, so that his identity remained secret during the study, in the final report, and in any publication that could be derived from the data collected. The participants determined the toxicity of five samples once a month, during four consecutive months, and thus four replicates were ob- tained for each sample. The EC_{50} values were deter- mined after 5 and 15 minutes, so that we collected 10 data sets, 2 for each sample. Each data set contained 88 results, 4 for each laboratory (unpublished data). Nevertheless, one of the data sets (zinc sulfate after 5 minutes) was discarded because of the high number of failures, and therefore the statistical analysis was restricted to 9 data sets.

In each of the four runs, five indistinguishable samples were used, conveniently scrambled. The sample identity, known only by the coordinator, was not disclosed to the participants until the final report. The EC_{50} results were sent to the study coordinator. Participants were also asked to send in the computer print- out, including the raw data (the actual gamma values). Missing EC_{50} values from gamma values rejected by the Microtox software were recalculated manually. This includes EC_{50} values higher than 100%, and also those not calculated because of rejected gamma values.

Statistical Methods

The statistical analysis was performed separately for the nine data sets (i.e., separately for the 5 and the 15 m assays and for each toxic agent), using the methods recommended in ISO (1994), previously discussed in this paper. A combination of statistical tests was used for the rejection of outliers in each data set.

First, a Cochran test for homogeneity of intra- laboratory variance was applied to detect extra high variances. For every outlying variance, the first test of Grubbs was applied to detect outliers within the group of four replicates. The second test of Grubbs

was applied to detect outlying laboratory means. Finally, when evidence of a non-random pattern of variation was found by comparing the results of the 5 and the 15 m assays for a single agent, the whole laboratory was skipped for this agent.

The estimates of the intra- and inter-laboratory variance were obtained as the components of variance in a nested analysis of variance. Then the estimates of the repeatability *(r)* and reproducibility *(R)* values were derived from these variances. *h* and *k* statistics were also calculated for each laboratory.

Parameters *r* and *R* are required in laboratory methods validation according to future ISO/EIC 17025 Standard, currently in draft version (ISO, 1999).

RESULTS AND DISCUSSION

The mean values and the intra- and inter-laboratory coefficients of variation obtained in this study are presented in Table II for the nine data sets. The statistics of both assays are quite similar for the four toxic agents for which the comparison is possible (phenol and effluents 1, 2, and 3), except for the mean value for effluent 1. The results of Table II have been obtained after the rejection of outliers. From the nine data sets, a total of 24 results were rejected, about 3%. Even after the rejection of the outliers, the intra-laboratory coefficient of variation is higher than 15%, except for phenol.

The results obtained for the *k* and *h* statistics, for the 15 m assay are presented in graphical form in Figs. 1 and 2. Each figure consists on five bar diagrams, with 22 bars, one for each laboratory. The horizontal line corresponds to the 95% critical value. Laboratory 1 was rejected in the statistical analysis of the phenol data because of a suspicious non-random pattern of variation present in both 5 and 15 m assays, suggesting that something happened when manipulating the samples, and therefore a bar is missing in the phenol diagrams.

Figure 1 shows the *k* values. The 95% critical value is 2.01 (Mandel, 1991). It can be seen that laboratory 2 in the phenol assay, laboratory 3 in the effluent 1 assay and laboratory 10 in the effluent 2 assay have statistically significant *k* values. The diagrams do not show any systematic pattern for the differences between intra-laboratory standard deviations. Figure 2 shows the *h* values. The 95% critical values are *±2.58* (Mandel, 1991). It can be seen that laboratories 2 and 17 in the effluent 1 assay and laboratory 4 in the effluent 2 and effluent 3 assays have significant *h* values. We observe in Fig. 2 that the zinc sulfate and phenol assays do not produce statistically significant *h* values. We also did not find any systematic pattern for the mean.

In Fig. 3, the *r* and *R* values are plotted against mean values obtained in the 5 m assay. Figure 4 is the same for the 15 m assay. These estimates have also been obtained after the rejection of the outliers. The inclusion of the outliers in the calculations would pro- duce inflated *r* and *R* values, which in some cases can be 10% higher.

As could be expected from the results of Table II, *r* and *R* values increase when the EC_{50} increases, so that the toxicity of the more toxic samples can be more precisely assessed. Such a pattern supports the use of coefficients of variation for the assessment of the precision of the assay. Nevertheless, although the coefficients of variation can be useful, it can be observed in Table II that they are highest for the lowest EC50 values, i.e., for the most toxic samples. Therefore, we cannot assume that there is a linear relationship between EC50 values and the corresponding *r* and *R* values.

CONCLUSIONS

In view of the results presented, we can conclude:

- There is no linear nor even monotonic relationship between EC_{50} values and *r* or *R.* This confirms that the chemical nature of the sample is influencial on the assay_precision. The relationship between EC_{50} and *r* or *R* cannot be reduced to a simple mathematical formula for any type of substance.

TABLE II. Mean EC_{50} values and coefficients of variation

	5 minutes			15 minutes		
	Mean	Intra-lab CV	Inter-lab CV	Mean	Intra-lab CV	Inter-lab CV
SZn				11.73	34.31%	23.17%
Phenol	19.70	13.23%	14.61%	21.20	14.45%	15.31%
Effluent 1	15.35	32.82%	13.92%	9.15	36.02%	15.30%
Effluent 2	33.00	19.57%	21.10%	35.09	23.93%	21.32%
Effluent 3	8.36	23.79%	25.62%	8.29	25.30%	29.94%

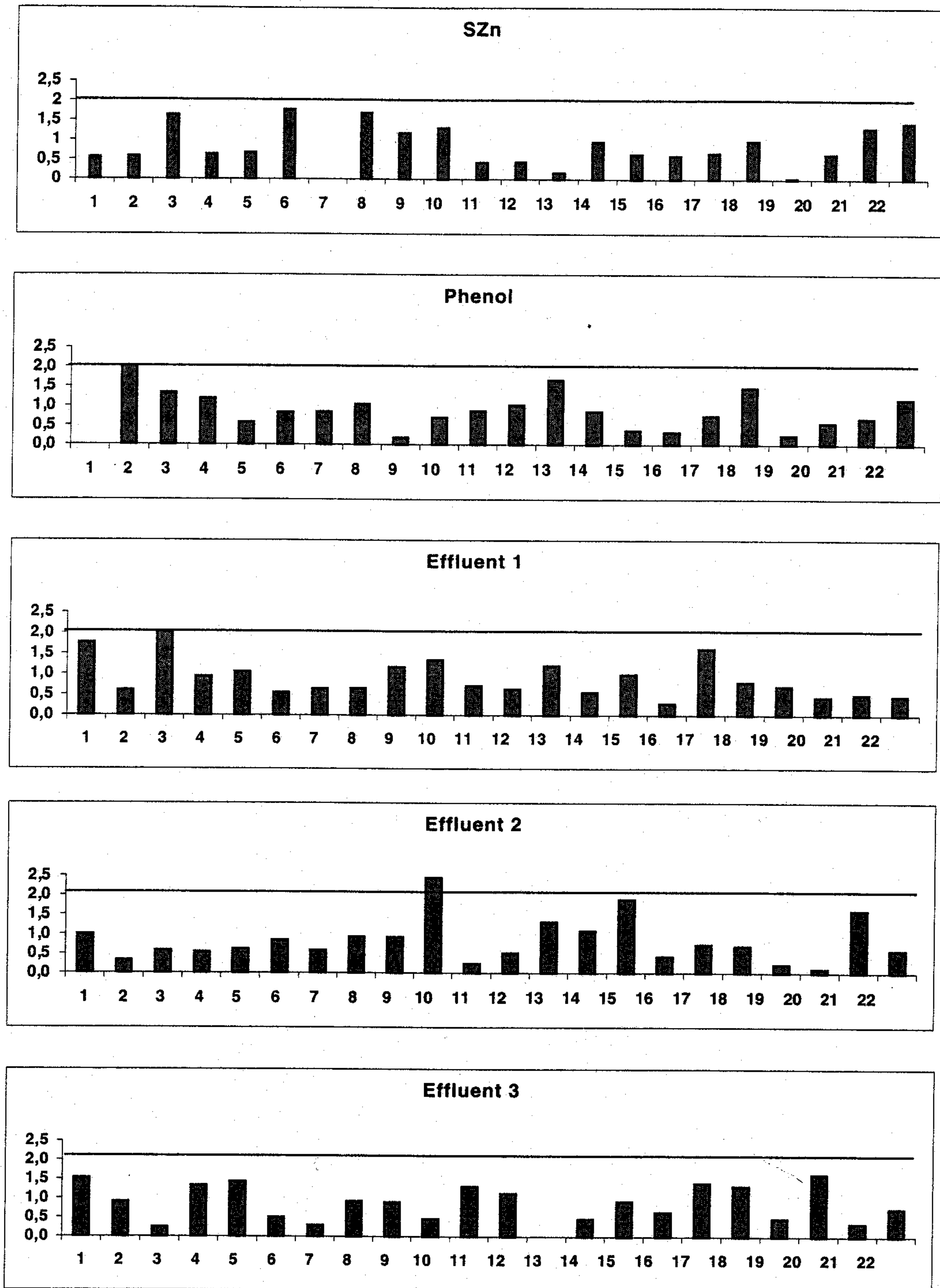

Fig. 1. *k* values of EC_{50} after 15 minutes for each laboratory.

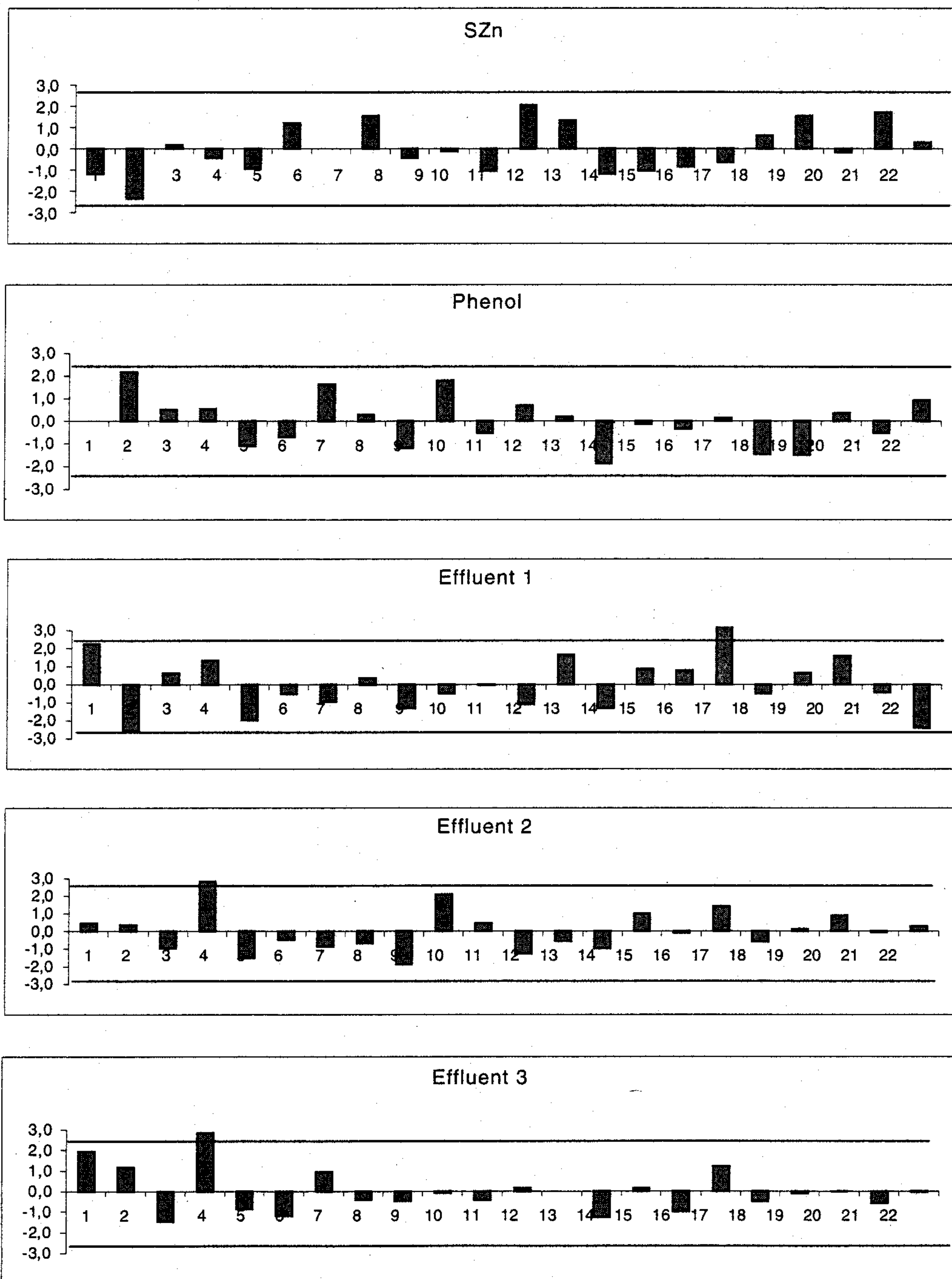

Fig. 2. *h* values of EC_{50} after 15 minutes for each laboratory.

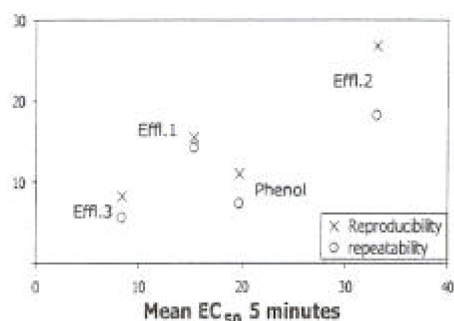

Fig. 3. *r* and *R* plotted against mean EC_{50} values for the 5 m assay

- Consequently, it makes no sense to characterize this assay by means of the CY or the *r* or *R* statistics in a universal way. These parameters would be of use only when referring to a family of substances with an analogous chemical nature. In any case they would not be applicable to complex samples such as industrial effluents.
- The rejection of outliers is a crucial issue in the evaluation of the precision, at least for this assay. Outliers are a common problem in experiments, as they distort the statistical summaries of experimental results, and, in particular, special tests for rejection of outliers are recommended for the inter-laboratory studies. In this study, these tests allowed us to refine our estimates of the repeatability and reproducibility of the Microtox assay. In the presentation of the conclusions of an inter-laboratory study, it is also important to provide clear information about the procedures used in the rejection of outliers and the percentage of results rejected.
- Mandel's *h* and *k* statistics are useful for graphical presentations of results of inter-laboratory studies, providing additional information on the differences between laboratories. These graphical presentations are also useful in verifying and highlighting systematic patterns in data from participant laboratories. They allow the immediate identification of laboratories differing from the rest, either in mean or in variance, and they present information more readily usable to improve the internal quality control of participant laboratories.

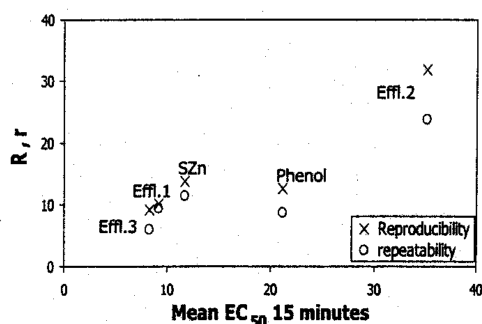

Fig. 4. *r* and *A* plotted against mean ECso values for the 15 m assay.

REFERENCES

ANSI/ ASTM. Standard Practices for Conducting an Inter-laboratory Test Program to Determine the Precision of Test Methods; 1979.

ANSI/ ASTM. Standard Practices for Dealing with Outlying Observations; 1989.

Bulich, A. A. Use of luminescent bacteria for determining toxicity in aquatic environments. In Aquatic Toxicology, ASTM 667; Markings, L. L.; Kimerle, R. A., Eds.; American Society for Testing and Materials: Philadelphia, 1979; pp 98-106.

Burke, S. Sci Data Manag 1998,2,28-35.

Chang, J. C.; Taylor, P. B.; Leach, F. R. Bull Environ Contam Toxico11981, 26, 150-156~

Curtis, C.; Lima, A.; Lozano, A.; Veith, G. D. Evaluation of a bacterial bioluminescence bioassay as a method for predicting acute toxicity of organic chemicals to fish. In Aquatic Toxicology and Hazard Assessment, 5th. Conf. ASTM STP 766; Pearson, J. G. *et al.,* Eds.; American Society for Testing and Materials: Philadelphia,1982; pp170-178.

DeZwart, D.; Slooff, W. Aquat Toxico1 1983, 4, 129-138.

Grubbs, F. E.; Beck, G. Technometrics 1972, 14,847-854.

Indorato, A. M.; Snyder, K B.; Usinowicz, P. J. Toxicity screening using Microtox analyzer. In Toxicity Screening Using Bacterial Systems; Dutka, B. J.; Liu, D., Eds.; Marcel Dekker: New York, 1983; pp 37-53.

ISO 5725-2. Accuracy (Trueness and Precision) of Measurement Methods and Results-A Basic Method for the De- termination of Repeatability and Reproducibility of a Standard Measurement Method; 1994.

ISO/IEC FDIS 17025. General Requirements for the Competence of Testing and Calibration Laboratories; 1999 (Final Draft).

IUPAC. Pure Appl Chem 1990,62,149-162. Johnson, F. H.; Eyring, H.; Stover, B. J., Eds. The Theory of Rate Processes in Biology and Medicine; John Wiley & Sons: New York, 1974.

Lebsack, M. E.; Yearson, A. D.; DeGraeve, C. M.; Bergman, H. L. In Aquatic Toxicology and Hazard Assessment, 4th Conference, ASTM STP 737; Branson, D. R.; Dickson, K L., Eds.; American Soc. for Testing and Materials: Philadelphia, 1981; pp 348-356.

Mandel, J. Evaluation and Control of Measurements; Marcel Dekker: New York, 1991.

Matthews, J. E.; Short, T. E. Screening Program for Demonstrating Hazardous Waste Land Treatability; Environmental Research Laboratory Report, Ada, 1983.

Qureshi, A. A.; Sharrna, A. K.; Paran, J. H. The Third International Symposium on Toxicity Testing Using Microbial Systems, Valencia, May 11-15,1987.

Ribó, J. M. Environ Toxicol Water Qua1 1997, 12,283-294.

Rib6, J. M.; Diez, M. Tecnologia del Agua 1994, 124, 49-58.

Rib6, J. M.; Kaiser, K. L. E. Chemosphere 1983, 12, 1421-1442.

Strosher, M. T. A Comparison of Biological testing Methods in Association with Chemical Analysis to Evaluate Toxicity of Waste Drilling Fluids in Alberta; Canadian Petroleum Association: Calgary, 1984; Vol. 1.

Taylor, J. K. Quality Assurance of Chemical Measurements ; Lewis Publ.: Michigan, 1989.

Vasseur, P.; Ferard, J. F. Environ Pollut 1984, 34, 1-11.