



Environmental Hazard Classification of Metals and Metal Compounds

- A Probabilistic Assessment of Classification Power
for Data Generated by the T/D Protocol

Reinhard Meister & Jonas Falck

Environmental Hazard Classification of Metals and Metal Compounds

- A Probabilistic Assessment of Classification Power for Data Generated by the T/D Protocol

TemaNord 2008:518

© Nordic Council of Ministers, Copenhagen 2008

ISBN 978-92-893-1663-7

Print: Ekspresen Tryk & Kopicenter

Printed on environmentally friendly paper

This publication can be ordered on www.norden.org/order. Other Nordic publications are available at www.norden.org/publications

Printed in Denmark

Nordic Council of Ministers

Store Strandstræde 18
DK-1255 Copenhagen K
Phone (+45) 3396 0200
Fax (+45) 3396 0202

Nordic Council

Store Strandstræde 18
DK-1255 Copenhagen K
Phone (+45) 3396 0400
Fax (+45) 3311 1870

www.norden.org

Nordic co-operation

Nordic cooperation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and three autonomous areas: the Faroe Islands, Greenland, and Åland.

Nordic cooperation has firm traditions in politics, the economy, and culture. It plays an important role in European and international collaboration, and aims at creating a strong Nordic community in a strong Europe.

Nordic cooperation seeks to safeguard Nordic and regional interests and principles in the global community. Common Nordic values help the region solidify its position as one of the world's most innovative and competitive.

Table of Contents

Preface	7
Reports on "Environmental Hazard Classification and Labelling" issued by the Nordic Council of Ministers.....	8
Executive Summary.....	9
1. Background	11
1.1 Classification of metals and sparingly soluble metal compounds.....	11
1.2 Intrinsic T/D properties and classification rule.....	12
1.3 Validation of the Transformation/Dissolution protocol.....	14
2. Classification and measurement error.....	17
2.1 A note on logarithms and notation	17
2.2 Classification based on one measurement resulting from one loading	18
3. Probabilities of Classification into Category I - III.....	21
3.1 Classification: hypothetical examples	22
3.2 Classification: data from validation ring test	25
4. Conclusions	29
References.....	30
Svensk sammanfattning (Summary in Swedish)	31

Preface

For the purpose of determining the rate and extent to which metals and sparingly soluble metal compounds can produce soluble available ionic and other metal-bearing species in aqueous media under a set of standard laboratory conditions representative of those generally occurring in the environment, a Test Guidance was designed (see A10 of /1/). The generation of Transformation/Dissolution (T/D) data according to this Test Guidance, which is referred to as the T/D protocol, is an integrated part of the recommended strategy to assess metals and metal compounds and transformation data can generally only be considered as reliable for the purposes of classification if conducted according to this Test Guidance (see A9.7.1.3 of /1/).

The present publication consist of *a probabilistic assessment of classification power for data generated by the T/D Protocol*, following the guidance on aquatic hazard classification of metals and metal compounds given in Chapter 7 of Annex 9 to /1/.

The assessment was conducted by Dr. Reinhard Meister, TFH Berlin – University of Applied Sciences, meister@tfh-berlin.

The publication is written by Reinhard Meister and Jonas Falck, Swedish Chemicals Agency, jonas.falck@kemi.se and produced by The Nordic Council of Ministers.

The Nordic project group on Classification and Labelling is responsible for this report and the probabilistic assessment presented, apart from being issued by the Nordic Council of Ministers (NCM), is by permission from NCM also intended for publication by the OECD concerning the validation of the transformation/dissolution (T/D) protocol.

The Nordic project on "Environmental Hazard Classification and Labelling" was launched in the spring of 1990. Until now, the results of the work have been presented in 15 reports issued by the Nordic Council of Ministers (see next page).

Reports on "Environmental Hazard Classification and Labelling" issued by the Nordic Council of Ministers

- "Miljöfarliga kemiska ämnen och produkter- System för klassificering, märkning och varuinformationsblad", (Nord 1990:087), pp. 62. (Publication 2 but in Swedish)
- "Substances and Preparations Dangerous for the Environment - A System for Classification, Labelling and Safety Data Sheet", (Nord 1990:087E), pp. 65.
- "Environmental Hazard Classification - data collection and interpretation guide", (TemaNord 1994:589, 2nd edition 1995:581), pp. 167.
- "Environmental Hazard Classification - classification of selected substances as dangerous for the environment (I)", (TemaNord 1994:643), pp 101.
- "Environmental Hazard Classification criteria for chemical substances: Terrestrial environment – Fate in the soil and soil compartment effects", (unpublished, 1997), pp 192.
- "Hazard classification of metals and metal compounds – Effect of pH on dissolution of priority substances", COWI Report: 29855-02, (unpublished, June 1997), pp 22.
- "Modelling of metal speciation in Hazard Classification", COWI Report: 03, (unpublished, September 1997).
- "Environmental Hazard Classification - classification of selected substances as dangerous for the environment (II)", (TemaNord 1997:549), pp 137.
- "Environmental Hazard Classification – The N-CLASS Database", CD-ROM version, (TemaNord 1999:538), pp 44 and CD-ROM.
- "Transformation/Dissolution of Specified Inorganic Metal Substances", VKI Project: 12434, (unpublished, December 1999), pp 22.
- "Transformation/Dissolution Screenings Test of Inorganic Metal Compounds", DHI project: 50602.01, (unpublished, February 2000).
- "The N-CLASS Database", Internet version release (2001).
- "The N-CLASS Database", Folder on content and reliee of the New Internet 5.0 version (2004).
- "Identifying relevant parameters that can be used for monitoring trends in the use of environmentally hazardous substances", (In Danish) (unpublished, 2004), pp 87.
- "Environmental Hazard Classification of Metals and Metal Compounds - A Probabilistic Assessment of Classification Power for Data Generated by the T/D Protocol", (TemaNord 2008:518, present report).

Executive Summary

This report serves as a guide for understanding the performance and reliability of a classification rule, when applied to data generated by Annex 10 to /1/ (the transformation/dissolution (T/D) protocol).

The classification of sparingly soluble metal compounds is based on ecotoxicity reference values (ERV) and on the measured concentrations of metal ions of experiments performed according to the T/D protocol and the operating procedure (OP, Annex 1 to /2/).

Experimental values have inevitably a component of random noise and classification of sparingly soluble metal compounds into category I – III cannot be made error free. Results of a statistical analysis in /2/ of the validation ring test of the T/D protocol allow, however, to derive a framework for assessing error rates and an approach tackling this rate is presented.

The summary of /2/ establishes constant variability on log scale (with respect to expected level of measurement) within and between laboratories can be assumed. In addition, a log-normal distribution of measured concentrations appears as a reasonable model for the data. Based on these findings, false negative and false positive decision rates can be calculated for hypothetical settings, if the ecotoxicity reference value as well as the T/D-characteristics (e.g. median concentrations for different loadings) are specified. Hence, the probability of classifying a substance into a given category can be predicted.

A 95% uncertainty-factor for classification can be calculated and the classification scenario can be studied, using a schematic plot of classification probabilities, assumed median concentrations and ERV. Thereby, based on the results from three laboratories, it can for concentrations measured at pH 6, be concluded that if the ratio (or its inverse) of the median concentration of a substance to the ERV is greater than 2, the false decision rate is below 2%.

As long as the T/D measurements show no extreme sublinear dependence on loading, misclassification is limited to a one-category difference. On the basis of the experimental findings, Cu_2O , Co_3O_4 , and Ni metal powder preparations from the validation ring test can be classified consistently for the three laboratories, providing examples of classification into categories I – III.

1. Background

1.1 Classification of metals and sparingly soluble metal compounds

The recommended strategy to derive an environmental hazard classification can be found in Annex 9, section A9.7, GHS, p. 483 of /1/. Most of the text under 1.1 and 1.2 in this publication is taken from Annex 9, section A9.7 of /1/.

The harmonized system for classifying chemical substances is a hazard based system, and the basis of the identification of hazard is the aquatic toxicity of the substances, and information on the degradation and bioaccumulation behaviour. The recommended strategy to classify aquatic hazardous metals and metal compounds deals only with the hazards associated with a given substance when the substance is dissolved in the water column. Exposure from this source is limited by the solubility of the substance in water and bioavailability of the substance in species in the aquatic environment. Thus, the hazard classification schemes for metals and metal compounds are limited to the hazards posed by metals and metal compounds when they are available (i.e. exist as dissolved metal ions, for example, as M^+ when present as $M-NO_3$), and do not take into account exposures to metals and metal compounds that are not dissolved in the water column but may still be bioavailable, such as metals in foods (see A9.7.1.1 of /1/).

The industry associations responsible for the supply of these materials have taken a positive attitude in seeking to assess the toxicity of their products. It has been agreed that, rather than conduct further aquatic toxicity testing, the key issue is to determine the extent to which bioavailable forms can be produced during standard conditions, the rate at which these forms are generated, and whether this rate and extent of formation should lead to classification.

The level of the metal ion which may be present in solution following the addition of the metal and/or its compounds, will largely be determined by two processes: the extent to which it can be dissolved, i.e. its water solubility, and the extent to which it can react with the media to transform to water soluble forms. The rate and extent at which this latter process, known as “transformation” for the purposes of the classification guidance, takes place can vary extensively between different compounds and the metal itself, and is an important factor in determining the appropriate hazard category (see A9.7.1.2 of /1/).

1.2 Intrinsic T/D properties and classification rule

Generally speaking, the rate at which a substance dissolves is not considered relevant to the determination of its intrinsic toxicity. For metals and many poorly soluble inorganic metal compounds, however, the difficulties in achieving dissolution through normal solubilization techniques is so severe that the two processes of solubilization and transformation become indistinguishable. Thus, where the short-term aquatic toxicity for the metal ions of concern (expressed as $L(E)C_{50}$) is less than or equal to 100 mg/l and the compound is sufficiently poorly soluble that the levels dissolved following normal attempts at solubilization do not exceed the available $L(E)C_{50}$, consideration must be given to the data available on the rate and extent to which these ions can be generated from the metal or metal compound (see A9.7.1.3 of /1/).

For the purpose of determining the rate and extent to which metals and sparingly soluble metal compounds can produce soluble available ionic and other metal-bearing species in aqueous media under a set of standard laboratory conditions representative of those generally occurring in the environment, a Test Guidance was designed (see A10 of /1/). The generation of Transformation/Dissolution (T/D) data according to this Test Guidance, which is referred to as the T/D protocol, is an integrated part of the recommended strategy to assess metals and metal compounds and transformation data can generally only be considered as reliable for the purposes of classification if conducted according to this Test Guidance (see A9.7.1.3 of /1/).

Where such data are unavailable, i.e. there is no clear data of sufficient validity to show that the transformation to metal ions will not occur, the safety net classification (Chronic Category 4) should be applied since the known classifiable toxicity of these soluble forms is considered to produce sufficient concern (see A9.7.5.2.3 of /1/).

However, where data have been generated using the T/D protocol, the evaluation of the aquatic hazard of metals and sparingly soluble metal compounds is to be accomplished by comparison of (a) the concentration of the metal ion in solution with (b) appropriate standard ecotoxicity data as determined with the soluble metal salt, known as the ecotoxicity reference values (ERV). If the ERV is exceeded, irrespective of whether the toxicity and dissolution data are at the same pH and if this is the only data available then the substance should be classified. If other solubility data are available to show that the dissolution concentration would not exceed the $L(E)C_{50}$ across the entire pH range then the substance should not be classified on its soluble form (see A9.7.2.3 of /1/).

The aquatic hazard classification of metals and sparingly soluble metal compounds are based on the dissolved metal ion concentrations obtained after a seven day T/D period. Normally massive forms and/or powders are introduced into the aqueous medium at three different load-

ings: 1, 10 and 100 mg/l. The loading that delivers a dissolved metal concentration that equals or exceeds the ERV, defines the classification level (see A10.2.3.1 of /1/).

If the 100 mg/L loading does not deliver a dissolved metal concentration exceeding the ERV, then the metal does not receive any of the core set of aquatic hazard classification categories (Category 1-3). For the purpose of this paper they will be called unclassified. They may, however, still be subject for a safety net classification (Chronic Category 4).

Basically, the classification considers the loading necessary to reach a limit where aquatic toxicity is expected. Thus the classification depends on both: the intrinsic toxicity of the metal-ions and the metal and/or its compounds ability to deliver certain concentrations of metal-ions in the aquatic medium. Figure 1 provides a condensed overview of this concept.

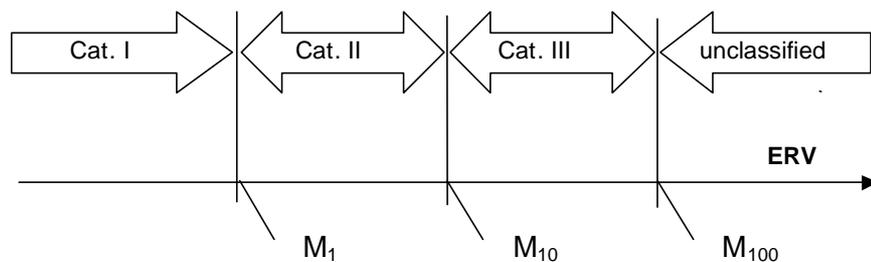


Figure 1: Intrinsic T/D properties and classification rule. The classification depends on the actual ecotoxicity reference value (ERV) and the T/D properties of a substance determined for different loadings (1, 10, 100 mg/L). The true median metal ion concentrations resulting from loadings of 1, 10, 100 mg/L are denoted by M_1 , M_{10} , M_{100} .

Example:

Assume that a compound has true median concentrations (M_1 , M_{10} , M_{100}) of 7, 50, and 400 $\mu\text{g/L}$ corresponding to loadings of 1, 10, and 100 mg/L in a 7 day testing period. An ERV of 25 $\mu\text{g/L}$ should result in a classification into Category II while an ERV > 400 $\mu\text{g/L}$ should leave the sparingly soluble compound unclassified.

However, classification cannot be based on true median concentrations; experimentally derived values have to be used instead. All measurements come with an error causing unavoidable classification error, which is analysed in the following section.

1.3 Validation of the Transformation/Dissolution protocol

The T/D protocol was first published in 2001 as an OECD Test Guidance document (3) before it was incorporated into GHS as Annex 10 to (1).

As explained in the foreword of (3) the T/D protocol was considered provisional and subject to changes depending on the outcome of a validation work.

In January 2002, the OECD Task Force on Harmonisation of Classification and Labelling (HCL) discussed the approaches for the validation of the T/D protocol, and agreed to establish the Validation Management Group on the Transformation/Dissolution protocol (VMG) (see (4)). In December 2002, December 2004 and December 2006, the UN sub-Committee of Experts on GHS requested the OECD to complete work on the validation (see (5), (6) and (7)).

It had originally been agreed that the validation should not cover the metal strategy as described in Chapter 7 of Annex 9 to (1), but that the validation could cover both reliability (Phase 1) and relevance for classification purposes (Phase 2) of the T/D protocol, Annex 10 of (1) (see (4) and (6)).

It was later clarified that the Phase 2 validation on relevance could possibly also cover discussions on the utility and applicability of the classification strategy (see Forward to (2)).

A central part of the Phase 1 validation was a ring test on transformation/dissolution, performed in 2005 with three metal substances (Cu_2O , Co_3O_4 , and Ni powders) and one INVAR alloy to determine its potential to provide consistent results between laboratories.

Four laboratories participated in the ring test. Two of these laboratories had been involved in the development of the Test Guidance and had much experience prior to the ring test with performing tests using the T/D protocol. The same two laboratories had co-operated in writing the Operating Procedure (OP) for the ring test (Annex 1 to (2)). Two other laboratories participated with less or no experience of the protocol.

For one of the latter laboratories, the measured metal ion concentrations, as a result of the test performed, were in line with the first two laboratories. The fourth laboratory showed, however, obvious discrepancies.

These discrepancies were not limited to a single or few measurement. Nor were they limited to the measurement of a single metal or metal loading. For example, while the fourth laboratory measured near-zero copper concentrations, the other three laboratories identified significant levels of dissolution with an average dissolved concentration of 131 and 391 $\mu\text{g/l}$ for the 7 day and 28 day test, respectively (target pH 6 and loading of 1 mg $\text{Cu}_2\text{O/l}$). Similar pattern was shown for target pH 8 (see Paragraph 50 and Figure 1.2a of (2)).

Note:

Already back in 2001, the European Copper Institute (ECI) reported T/D test data on Cu_2O following the T/D protocol. The report revealed for the target pH of 6 and a loading of 1 mg/l a measured copper concentration of 236 $\mu\text{g/l}$ after 7 days (see /8/). Hence, the ECI results are very much in line with the results reported from three of the four laboratories in the validation ring test.

For the loadings of 10 and 100 mg $\text{Cu}_2\text{O/L}$ and target pH of 6 and 8, the laboratory that had shown obvious discrepancies reported significantly greater copper concentrations than the other three laboratories. And while the other three reported a significant difference in copper concentrations between the tests performed under the two target pHs the fourth laboratory could not show this clear difference (see Figure 1.2 a and 1.2.b of /2/).

Note:

A linear regression of dissolved copper concentration, between pH 6 and 8 was studied and reported by the European Copper Institute in 2001, with thermodynamic calculation as supportive evidence (see /8/).

It is important to realise that just because the results from one laboratory in the ring test of a new protocol reported discrepancies in their measurements most assuredly does not mean that this laboratory will represent 1 in any 4 laboratories worldwide that may in the future conduct T/D testing. All this discrepancy means that procedures at one of the four participating laboratories alone were at some point faulty, somewhere.

Even if there are no obvious reasons at this time as to why there should be such considerable discrepancies between one laboratory and the other three laboratories, and a thorough investigation was not done to investigate this further, avoiding results like those of laboratory 4 is considered not to be a statistical issue. Therefore, laboratory 4 was excluded from the statistical analysis in /2/, as well as in this report.

From a statistical point of view, variability for the measured concentrations of the components of the alloy tested could not be assessed with sufficient precision. The alloy is not included in the probabilistic assessment in this report.

2. Classification and measurement error

In reality, classification is based on the experimentally determined dissolved concentrations C . In this section, the probability of a classification given a specific ERV and the T/D properties of a substance will be derived. As a first step, false-positive and false negative decision rates will be defined and a rule for their computation will be derived.

For the introduction of the concept, only one measurement and one loading is considered. More complicated cases, considering measurements for several loadings simultaneously are discussed later.

For understanding the formulae used in computation, the following note should be useful.

2.1 A note on logarithms and notation

It had been expected, and the data analysis of the ring test study results showed, that a logarithmic transformation of the concentration measurements is very convenient (*cf. /2/*). In particular, the Normal-distribution appears as acceptable model for characterizing the distribution of the log-transformed concentrations. Therefore, this report uses log₁₀-transformed data for computations, and back-transformed values for presentations and graphs. When used for graphs, concentrations are displayed on logarithmically scaled axes.

Figure 2 displays some of the facts, helpful for understanding the transformation.

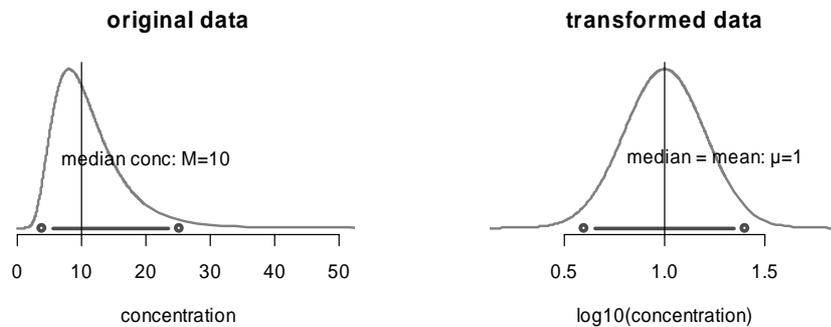


Figure 2: Distribution of concentration measurements, displayed on original and on logarithmic scale including prediction intervals (o--o), where a randomly sampled value will be observed with 95% probability.

The log10-transformation is monotonic, therefore percentiles are preserved. In particular, the median of the log-transformed concentrations equals the log-transform of the median of the original distribution. An example is given in Figure 2.

For understanding the behaviour of the classification procedure, it is important to know what to expect from a new concentration measurement of a given compound by a randomly chosen laboratory. Standard normal distribution theory gives the 95% prediction intervals for new measurements.

Assuming a normal distribution for the log-concentrations with expectation μ and variance σ^2 , a $1-\alpha$ prediction interval is given by $\mu \pm z_{1-\alpha/2} \times \sigma$. For the example data in Figure 2 values approximately equal to $1 \pm 2 \times 0.2 = (0.6, 1.4)$ are obtained when $\alpha = 5\%$. The anti-log transformation gives the interval $10^{(\mu \pm z_{1-\alpha/2} \times \sigma)}$ approximately equal to (4, 25). Using $M = 10^1 = 10$ and $f = 10^{2 \times 0.2} \approx 2.5$ it is obvious, that the boundaries of the interval are given by $(M / f, M \times f)$. One-sided limits (that is, $(-\infty, M \times f)$ or $(M / f, \infty)$) could be computed similarly, just replacing $z_{1-\alpha/2}$ by $z_{1-\alpha}$.

Notations:

variable	original scale	log10 scale
concentration	C	c or $\log_{10}(C)$
true median	M	μ
stand.deviation		σ
ecotox.ref.val.	ERV	erv

2.2 Classification based on one measurement resulting from one loading

The classification rule is very simple given a specific ERV:

Rule: classify the substance if $ERV < C$

There are four possible situations that have to be considered. These are explained in the table 1.

Table 1: There are four possible situations when comparing the measured concentration (C) based on one measurement from one loading with the ecotoxicity reference value (ERV). M denotes the true median metal ion concentration.

		Not classified $ERV > C$	Classified $ERV < C$
True unclassified	$ERV > M$	correct	false positive
True Cat. 1	$ERV < M$	false negative	correct

The probability for a classification decision can easily be calculated, after making some assumptions about the true intrinsic properties of a substance.

Assumption: $\log_{10}(C) \sim N(\mu, \sigma^2)$ normally distributed \log_{10} transformed concentration measurements.

This assumption covers both aspects of the experimental determination of a concentration: the median value μ and the standard deviation σ quantified on \log_{10} scale.

Given these intrinsic characteristics of the metal ion concentration observed after the T/D process at a certain loading, the probability of being classified is a function depending solely on the actual value of the ecotoxicity reference limit.

Probability of classification:	$P(ERV < C)$	$= 1 - P(C \leq ERV)$ $= 1 - \Phi(\{erv - \mu\} / \sigma)$ $= \Phi(-\{erv - \mu\} / \sigma)$
---------------------------------------	--------------	--

Here, Φ denotes the cumulative distribution function of the standard normal distribution.

The calculation of expected classification rates is now straightforward, just by inserting numerical values into the equation given above. Figures 3 and 4 illustrate the cases of false negative and false positive classification.

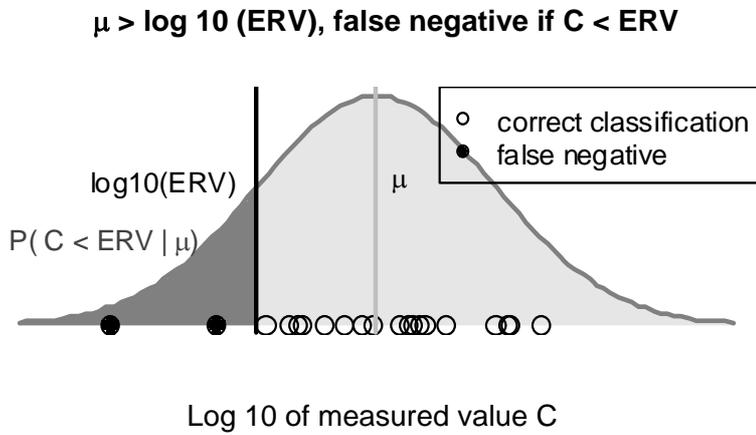


Figure 3: Application of classification rule to measured concentrations. Probability of **false negative classification**: not classified, true mean above ERV. For illustration purpose simulated measurements, displayed as filled and empty circles, are added.

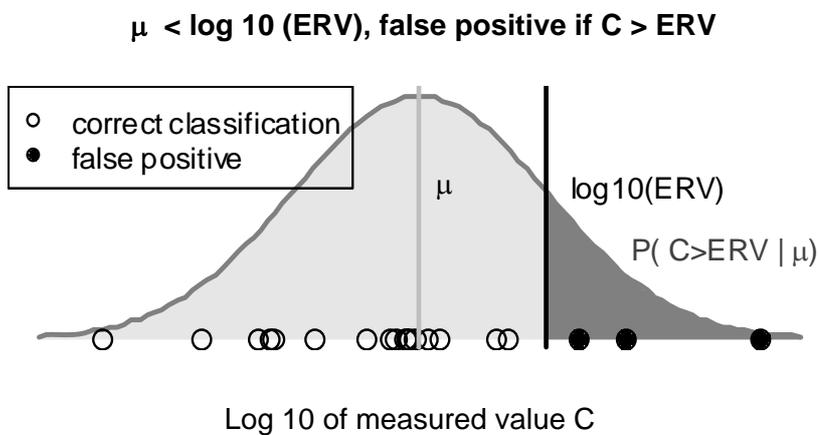


Figure 4: Application of classification rule to measured concentrations. Probability of **false positive classification**: classified, true mean below ERV. For illustration purpose simulated measurements, displayed as filled and empty circles, are added.

The concept described above will be used in the next part, considering the real procedure with three plus one categories.

3. Probabilities of Classification into Category I - III

In this part, probabilities for correct and for false classifications are derived. Obviously, these probabilities depend on the following variables:

- the ecotoxicity reference value
- the true median concentrations at the different loadings
- the variability of the measured concentrations

From the report on the “Statistical Reanalysis of the T/D Validation Study” (Meister 2006) two conclusions can be drawn:

- the T/D measurements are appropriately described by a lognormal probability distribution
- the variability within and between laboratories can be regarded as constant on log scale

For an evaluation of the performance of the classification scheme the probabilities needed can be calculated, under the assumptions mentioned above. Given the intrinsic log-median concentrations of a substance (μ_{10} , μ_{100}) and the common standard deviation σ , the desired probabilities are derived under the assumption of independent measurements at different loadings. Again Φ denotes the cumulative distribution function of the standard normal distribution.

Table 2: Formulae for computing probabilities of classification into different categories

Classification	Event	Probability
Category I	$C_{1mg} > ERV$	$1 - \Phi(\{erv - \mu_{1mg}\} / \sigma)$
Category II	$C_{1mg} < ERV \leq C_{10mg}$	$\Phi(\{erv - \mu_{1mg}\} / \sigma) \times [1 - \Phi(\{erv - \mu_{10mg}\} / \sigma)]$
Category III	$C_{1mg} < ERV \cap C_{10mg} < ERV \cap C_{100mg} > ERV$	$\Phi(\{erv - \mu_{1mg}\} / \sigma) \times \Phi(\{erv - \mu_{10mg}\} / \sigma) \times [1 - \Phi(\{erv - \mu_{100mg}\} / \sigma)]$
unclassified	$\neg(Cat.I \cup Cat.II \cup Cat.III)$	$1 - [P(Cat.I) + P(Cat.II) + P(Cat.III)]$

The derivation of the formulae given above is straightforward and on a basic level of probability calculus. Within the range of settings considered in this paper, the assumption of independent measurements is not critical and could be replaced by appropriate assumptions for intra-laboratory correlations, without making any substantial change to the classification probabilities derived. This has been demonstrated by some numerical integration.

3.1 Classification: hypothetical examples

A schematic plot of classification probabilities will be presented, illustrating the properties of the classification procedure. To this end information about the essential parameters is needed:

- assumed true median concentrations corresponding to different loadings
- assumed standard deviation of log10 transformed concentration measurements

True median concentrations

The assumed median concentrations differ from substance to substance. In addition, different behaviour of the T/D process results in different ratios of the medians corresponding to the loadings. For example 1:10:100 has linear behaviour, but 1:5:15 has extremely sublinear behaviour (observed if the T/D process shows an apparent plateau).

Standard deviation

The standard-deviation is taken as constant on the log scale. This assumption is fully supported for the substances tested in the validation study.

For further calculations a standard deviation of 0.14 on log10 scale is assumed. This value corresponds to the sum of variances within and between laboratories ($\sigma = \sqrt{\sigma_{within}^2 + \sigma_{between}^2}$) derived in the validation ring test. A measurement taken in a randomly selected laboratory will have exactly this standard deviation under the assumptions made in the evaluation of the ring test.

Schematic plot

The schematic plot (figure 5) displays the probabilities of classification into different categories.

Once the intrinsic properties of a metal are specified, this probability depends only on the magnitude of the ecotoxicity reference value (ERV).

Therefore, the horizontal axis displays all potential values of ERV. Ticks and divided grey boxes indicate the assumed true median concentrations as well as regions of uncertainty in classification. The lower half of the boxes controls the false negative classification rate; the upper half controls the false positive rate. The horizontal axis displays original values using a logarithmic scaling.

The construction of the boxes follows the ideas presented in 2.1. One-sided limits ($\alpha = 5\%$) are used with $\sigma = 0.14$ as specified above. The width of the boxes $w = z_{0.95} \times \sigma = 1.65 \times 0.14 = 0.23$ guarantees that the risk of a false classification is below 5% outside the boxes. Consider, for example, a situation where ERV is below the true median M_1 for the 1 mg/L loading corresponding to a true Category I substance. The corresponding values on log10-scale will be denoted by erv and μ_1 . It could happen though, that a measured concentration C_1 below the ERV is observed. This observation would result in a false negative classification into Category II. The dashed curve in figure 4 displays the probability of a Category II classification. The false negative rate $P\{\log_{10}(C_1) < erv\}$ is smaller than 5% if $erv < \mu_1 - w$ holds true. This condition can also be expressed for the original data. The false negative rate is below 5% if $ERV < M_1 / (10^w)$. With the values assumed the factor is $10^w = 1.7$.

Note:

The calculation of classification probabilities is possible only if the true median concentrations are known. There is no way to estimate misclassification rates from single observed concentrations. However, if the ratio of a measured metal ion concentration to the ERV is larger than the factor given above (or smaller than its inverse), a misclassification appears unlikely.

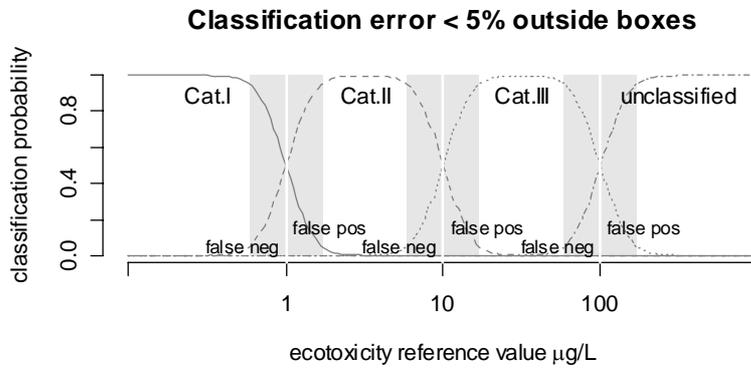


Figure 5: Schematic plot: Probabilities of classification for an artificial substance. The true median concentrations corresponding to loadings of 1, 10, and 100 mg/ are assumed as 1, 10 and 100 µg/L.

From figure 5 the behaviour of the classification scheme can be read off. As an example, the uninterrupted line gives the probability for classification into Category I. If the ERV is within the shaded box around the median concentration for 1 mg/L loading, a substantial risk of false classification occurs. ERV values outside the boxes would result in correct classification with a probability greater than 95%.

Numerical results are listed in table 3, which is just a read-out from figure 5.

Table 3: Probabilities of classification for an artificial substance. The true median concentrations corresponding to loadings of 1, 10, and 100 mg/L are assumed as 1, 10 and 100 µg/L.

ERV	P(Category I)	P(Category II)	P(Category III)	P(unclass)
0.50	0.98	0.02	0.00	0.00
0.67	0.90	0.10	0.00	0.00
1.00	0.50	0.50	0.00	0.00
1.50	0.10	0.90	0.00	0.00
2.00	0.02	0.98	0.00	0.00
5.00	0.00	0.98	0.02	0.00
6.67	0.00	0.90	0.10	0.00
10.00	0.00	0.50	0.50	0.00
15.00	0.00	0.10	0.90	0.00
20.00	0.00	0.02	0.98	0.00
50.00	0.00	0.00	0.98	0.02
66.67	0.00	0.00	0.90	0.10
100.00	0.00	0.00	0.50	0.50
150.00	0.00	0.00	0.10	0.90
200.00	0.00	0.00	0.02	0.98

The behaviour of the classification probability is similar for all categories. This could be different, if the T/D process shows a sub-linear characteristic.

Such a case is provided by the situation shown in figure 6.

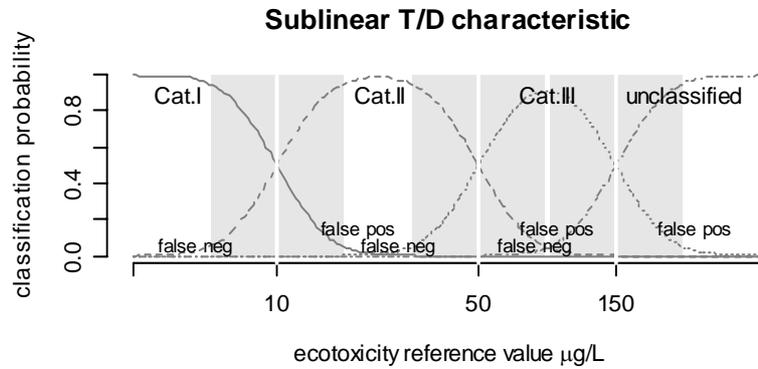


Figure 6: Probabilities of classification for an artificial substance. The true median concentrations corresponding to loadings of 1, 10, and 100 mg/L are assumed as 10, 50 and 150 µg/L.

Figure 6 demonstrates, that substantial sub-linear T/D characteristics of a substance can result in difficulties concerning a correct classification, if the ERV is in the range of the concentrations reached by the 10 and 100 mg/L loadings. This example, however, shows a very extreme case, not seen in the validation study for the powder preparations.

Numerical results are listed in table 4, which is just a read-out from figure 6.

Table 4: Probabilities of classification for an artificial substance. The true median concentrations corresponding to loadings of 1, 10, and 100 mg/L are assumed as 10, 50 and 150 µg/L.

ERV	P(Category I)	P(Category II)	P(Category III)	P(unclass)
5.00	0.98	0.02	0.00	0.00
6.67	0.90	0.10	0.00	0.00
10.00	0.50	0.50	0.00	0.00
15.00	0.10	0.90	0.00	0.00
20.00	0.02	0.98	0.00	0.00
25.00	0.00	0.98	0.02	0.00
33.33	0.00	0.90	0.10	0.00
50.00	0.00	0.50	0.50	0.00
75.00	0.00	0.10	0.88	0.01
100.00	0.00	0.02	0.88	0.10
150.00	0.00	0.00	0.50	0.50
225.00	0.00	0.00	0.10	0.90
300.00	0.00	0.00	0.02	0.98

3.2 Classification: data from validation ring test

The concept given in the preceding paragraph is illustrated using data from the ring test study. The results from three laboratories and three substances are used. Here only concentrations measured at pH 6 are considered. The geometric mean of the concentrations provided by the three

laboratories is used as substitute of the true median concentration. The substances are classified identically when using the individual data from the labs.

Cu₂O: For Cu₂O the ERV equals 29µg/L, all measured concentrations are greater than this value, therefore, the substance is labelled as Category I (see figure 7). If the substitute median concentrations are taken as true, the chance of a false positive classification can be taken as zero.

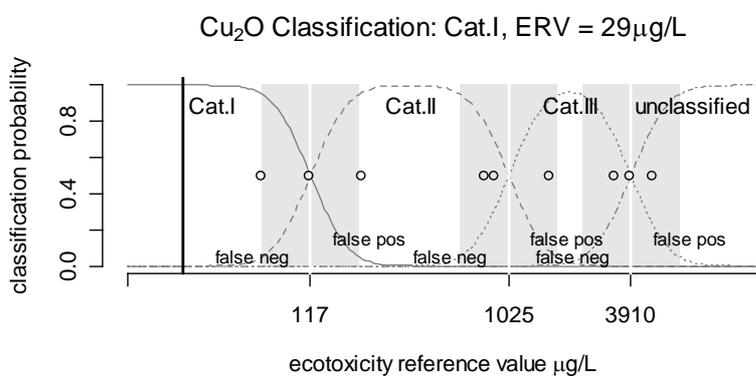


Figure 7: Classification of Cu₂O, pH=6. The ERV is far below all measured concentrations, so Cu₂O would be classified as Category I. Assumed median concentrations as indicated, individual measurements displayed as open circles.

Co₃O₄: The measured concentrations of Co₃O₄ give a Category II result (see figure 8). The ERV of 6.7µg/L is far from the assumed true median concentrations (3.1 µg/L for 1mg/L loading and 25.8µg/L for 10mg/L loading). The probability of a correct classification equals virtually 100%.

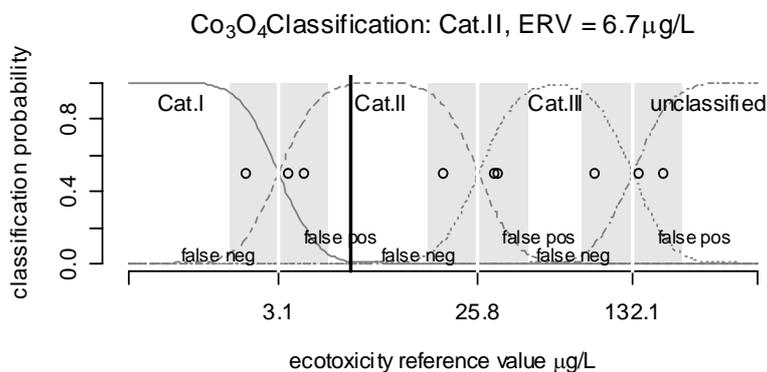


Figure 8: Classification of Co₃O₄, pH=6. The ERV is well in between the measured concentrations at 1mg/L and 10mg/L loading, so Co₃O₄ would be classified as Category II. Assumed median concentrations as indicated, individual measurements displayed as open circles.

Nickel is classified into Category III (see figure 9). The ERV of 67µg/L is just above the concentrations measured by the three labs for 10mg/L loading. Taking the geometric mean of 52.8µg/L as true median concentration, there would have been a chance of about 23% for a false positive classification into Category II. The observed per lab measurements, however, all result in a Category III decision.

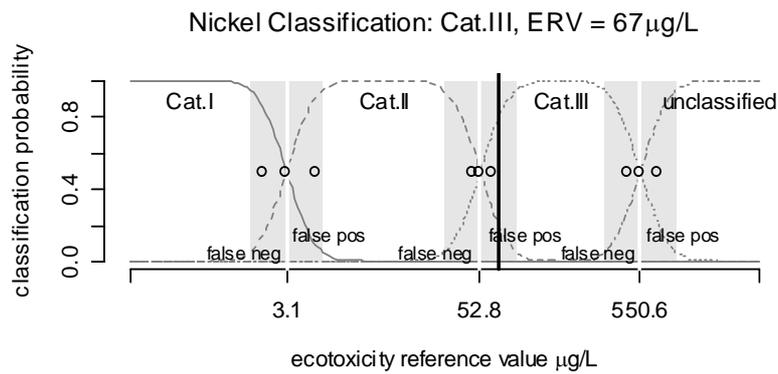


Figure 9: Classification of Nickel, pH=6. The ERV is just above all measured concentrations at 10mg/L loading, so Nickel would be classified as Category III. Assumed median concentrations as indicated, individual measurements displayed as open circles.

For further clarification the information used in classifying the three compounds is included in tabular form. The assumed T/D properties, their relation to the ERV values and the probabilities of classification are provided in tables 5-7.

Table 5: Numerical summary of T/D data. Ecotoxicity reference values and assumed median concentrations for loadings of 1, 10, and 100 mg/L at pH6 (Concentrations in µg/L).

compound	ERV	M1	M10	M100
Cu ₂ O	29	117	1025	3910
Co ₃ O ₄	6.7	3.1	25.8	132.1
Nickel	67	3.1	52.8	550.6

The ratios in table 6 show the relation between the ERV and the T/D response. Only one ratio lies within the interval (0.59, 1.7), where a substantial chance for misclassification exists: see Nickel and the ratio of M10 and ERV.

Table 6: Ratios of assumed median concentrations to ERV.

compound	M1/ERV	M10/ERV	M100/ERV
Cu ₂ O	4.0	35	135
Co ₃ O ₄	0.5	3.9	19.7
Nickel	0.0	0.8	8.2

The fact, that M10 and ERV are rather similar for Nickel is expressed in a lower rate of correct classification (see table 7). Such behaviour is unavoidable in principle.

Table 7: Probability of classification into different categories, given assumed median concentrations and ERV values of compounds. Correct classifications are indicated as bold.

compound	Cat.I	Cat.II	Cat.III	unclass.
Cu ₂ O	1.00	0.00	0.00	0.00
Co ₃ O ₄	0.01	0.99	0.00	0.00
Nickel	0.00	0.23	0.77	0.00

Probabilities are calculated according to formulae given in table 2.

4. Conclusions

- Classification of sparingly soluble metal compounds into category I – III cannot be made error free.
- Results of the statistical analysis of the validation ring test allow a framework for assessing error rates to be derived.
- False negative and false positive decision rates can be calculated for hypothetical settings, if the ecotoxicity reference value and the T/D-characteristics (e.g. median concentrations for different loadings) are specified.
- If the ratio (or its inverse) of the median concentration of a substance to the ERV is greater than 2, the false decision rate is virtually equal to zero.
- As long as the TD measurements show no extreme sublinear dependence on loading, misclassification is limited to a one-category difference.
- Classification scenarios can be studied, using a schematic plot of classification probabilities, assumed median concentrations and ERV.
- Metal powder preparations from the validation ring test were classified consistently for the three laboratories, thereby providing examples for categories I – III.

References

- /1/ United Nations (2007) *The Globally Harmonized System of Classification and Labelling of Chemicals*. ST/SG/AC.10/30/Rev. 2. United Nations, New York and Geneva.
- /2/ OECD (2008), Report of the ring test and statistical analysis of performance of the guidance on transformation/dissolution of metals and metal compounds in aqueous media (Transformation/Dissolution Protocol) (DRAFT, 18 January 2008), OECD Series of Testing and Assessment, No. XX. Organisation for Economic Co-operation and Development, Paris.
- /3/ OECD (April 2001), *Environment, Health and Safety Publications, Series on Testing and Assessment, No. 29*, Environment Directorate, Organisation for Economic Co-operation and Development, Paris.
- /4/ OECD (13 September 2002) ENV/JM/HCL/M(2002)1/REV1, Summary record of the 11th meeting of the Task Force on Harmonisation of Classification and Labelling 24–25 January 2002.
- /5/ United Nations (7 February 2003) ST/SG/AC.10/C.4/8, Report of the Sub-Committee of Experts on GHS on its fourth session, 9–11 December 2002.
- /6/ United Nations (21 December 2004) ST/SG/AC.10/C.4/16, Report of the Sub-Committee of Experts on GHS on its eighth session, 7–9 December 2004.
- /7/ United Nations (21 December 2006) ST/SG/AC.10/C.4/24, Report of the Sub-Committee of Experts on GHS on its twelfth session, 12–14 December 2006.
- /8/ European Copper Institute, *Synthesis of the information on hazard classification of di-copper oxide*; ECBI/61/95 Add. 135.

Svensk sammanfattning (Summary in Swedish)

Syftet med denna rapport är att ge en vägledning i förståelsen av en klassificeringsregels utformning och tillförlitlighet, när den tillämpas på data som tagits fram med hjälp av bilaga 10 till /1/, (Transformation/Löslighets (*Transformation/Dissolution* – T/D) protokollet.

Klassificeringen av svårlösliga metallföreningar baseras på ekotoxikologiska referensvärden (ERV) och på uppmätta koncentrationer av metalljoner vid tester utförda enligt T/D-protokollet samt på tillvägagångssättet (OP, Bilaga 1 till /2/).

Experimentella värden har oundvikligen en komponent av slumpmässiga fel och klassificering av svårlösliga metallföreningar i kategori I – III kan inte göras utan att fel uppstår. Resultat av en statistisk analys i /2/ av ett ringtest vid validering av T/D-protokollet gör det dock möjligt att ta fram ett ramverk för beräkning av felfrekvenser och en metod för att hantera dessa beskrivs.

I sammanfattningen till /2/ fastställs att en konstant variabilitet i logaritmisk skala (med avseende på förväntade mätnivåer) kan förutsättas inom och mellan laboratorerna. Dessutom framstår en lognormalfördelning av uppmätta koncentrationer som en rimlig modell för framtagna data. Baserat på dessa slutsatser, kan osäkerhetsmarginaler med negativ och positiv avvikelse beräknas för hypotetiska förutsättningar, om både ekotoxikologiskt referensvärde och T/D-karakteristika (t.ex. mediankoncentrationer motsvarande olika tillsatta mängder) är specificerade. Man kan därmed förutsäga sannolikheten för att ett ämne ska klassificeras i en given kategori.

En 95% osäkerhetsfaktor vid klassificering kan beräknas och klassificeringsscenarioet kan studeras med hjälp av en schematiskt gjord plottning av sannolikheterna för klassificering i de olika kategorierna, antagna mediankoncentrationer och ERV. Baserat på resultaten från tre laboratorier, kan man för koncentrationer uppmätta vid pH 6 dra slutsatsen att om förhållandet (eller dess inverterade värde) mellan mediankoncentrationen för ett ämne och dess ERV är större än 2, är osäkerhetsmarginalen vid klassificeringen mindre än 2 %.

Så länge T/D-mätningarna inte visar något extremt sublinjärt beroende av tillsatt mängd, är felklassificeringen begränsad till en skillnad mellan en kategori och den närmast liggande. Baserat på de experimentella resultaten från ringtestet vid valideringen kan metallpulver av Cu_2O , Co_3O_4 och nickel klassificeras konsistent för de tre laboratorerna, och ge exempel på klassificering i kategorierna I – III.