Estimating toxicity thresholds for aquatic ecological communities from sensitivity distributions 11-13 February 2014, Amsterdam

Workshop Report No. 28

EUROPEAN CENTRE FOR ECOTOXICOLOGY AND TOXICOLOGY OF CHEMICALS

Estimating toxicity thresholds for aquatic ecological communities from sensitivity distributions 11-13 February 2014, Amsterdam

Workshop Report No. 28

Brussels, November 2014

ISSN-2078-7200-28 (print)

ISSN-2078-7219-28 (online)

ECETOC Workshop Report No. 28

European Centre for Ecotoxicology and Toxicology of Chemicals
2 Avenue E. Van Nieuwenhuyse (Bte 8), B-1160 Brussels, Belgium.

Estimating toxicity thresholds for aquatic ecological communities from sensitivity distributions

CONTENTS

1. SUMMARY

This 3-day workshop organised by ECETOC and the Environment Agency for England and Wales took place in Amsterdam on the 11-13th February 2014.

The aims of the workshop were to review current thinking on when and how species sensitivity distributions (SSDs) should be used in environmental protection and management of chemicals and discuss how the methodology could be further developed to improve the quality of decisions. The workshop covered 3 specific areas:

a) What is the ecological relevance of an SSD?

b) What SSD statistical models are available for deriving toxic thresholds (HC5/PNEC) (hazardous concentration affecting p% of species / predicted no effect concentration) for aquatic communities?

c) What are the regulatory applications?

There were 41 attendees with experience in ecological risk assessment, ecology and statistics from academia, the chemical industry and the regulatory community. Appendix A provides a list of the attendees and Appendix B the meeting programme.

18 presentations were given during the workshop which covered the broader aspects of the use of SSDs in environmental protection and management, recent developments and specific case studies. In addition there were 3 syndicate sessions (with 4 discussion groups each) which focused on ecological, statistical and regulatory considerations.

There was general consensus that scientifically sound extrapolation approaches based on agreed but perhaps minimal and pertinent dataset SSDs for deriving toxicity threshold concentrations/PNECs should provide a more useful and transparent assessment of risks than a deterministic approach using generic factors applied to simple aquatic toxicity tests. The ability to understand the uncertainty was consider to be very important. For regulatory tools to be useful, they must not be overcautious (i.e. the tools become over protective and lead to unnecessary costs). When large datasets are available, the risk of being over protective is reduced and the use of SSDs becomes an option. When datasets are small, uncertainty is greater and consequently the more cautious deterministic approach may be more appropriate. The additional use of SSDs, (e.g. deriving expected impact magnitudes for focusing risk management of chemical contamination) was also acknowledged.

Continued validation against field and mesocosm data is required to ensure that thresholds expressed as an environmental quality standard (EQS) or PNEC have ecological relevance. The results of any extrapolation process (including SSDs) should always be critically assessed based on all available knowledge on the substance and related substances. It was agreed that SSD methodology is a valuable regulatory and management tool since it can give greater insight into the potential ecological effects than the assessment factor method, enabling better problem definitions and decisions. Use of SSD methodology should yield more generally applicable and acceptable results than those obtained from mesocosm-based methods.

The workshop considered the scientific and regulatory use of SSDs in chemical risk assessment and compared a range of accepted tools and their implementation. A novel approach, hSSD, to predict thresholds for defined species assemblages that uses knowledge of the general trends in how species sensitivity is related to their taxonomic distance was also discussed. A representative ETX model, the US Environmental Protection Agency (EPA) Web-ICE tool and the hSSD prototype were compared using case studies on the surfactant linear alkylbenzene sulfonate (LAS) and the pesticide, chlorpyrifos. The advantages and disadvantages of each approach were reported.

SSDs are used in both prospective risk- and retrospective impact assessment of chemicals. A prospective risk assessment needs to establish that there will be acceptable risk. In contrast, retrospective impact assessment uses diagnostic tools to identify the cause of existing adverse effects, using SSDs to quantify expected impacts. 3 distinct regulatory activities were identified:

1. The derivation of generic thresholds that need to be applied to many different locations, perhaps over very large geographical regions. These are assumed to offer sufficient protection everywhere, even in the most sensitive systems.
2. The derivation of scenario-specific thresholds that more closely reflect local conditions but which may not be transferable from one place to another.
3. Identifying the causes of biological impact (‘diagnosis’) or expected impact magnitudes of existing (mixture) contamination, in order to inform the need and focus for any remedial action.

Looking to the future, although it was noted that expert judgement will always be needed, a compilation of current best practices, a review of the state of the science and answers to frequently asked questions would facilitate acceptance of SSDs by regulators and risk managers. This compendium of best practice should be a technical document aimed at users with knowledge of SSDs and ecosystems. A less technical document suited to a more general audience is also recommended.

Knowledge gaps were identified and a list of research topics were developed. Although not prioritised during the workshop, the report authors will seek an indication of priority for the work from the workshop participants.

2. WORKSHOP OVERVIEW

2.1 Introduction

In 2008 ECETOC held a workshop on Probabilistic Approaches for Marine Hazard Assessment (ECETOC, 2008) which included discussion of the application of ETX-type SSDs for marine hazard assessment. The outcome identified key statistical and ecological challenges and further research to address them. Since then refined regulatory guidance has helped applicability of SSDs in risk assessment but there remain ecological and statistical challenges. Other approaches to SSDs have emerged over this period, e.g. US EPA Web-ICE and a prototype hierarchical model from Durham University, UK, both of which utilise prior knowledge of interspecies sensitivity relationships in deriving HC5 values.

These developments suggested that a workshop bringing together SSD practitioners from industry, regulatory agencies and academia would be valuable in discussing and report current thinking on when and how SSDs, should be used and how the methodology might be further developed. With this in mind, ECETOC organised this latest workshop with the aim of addressing 3 key aspects:

1. What is the ecological relevance of an SSD?
2. What SSD statistical models are available for deriving toxic thresholds (HC5/PNEC) for aquatic communities?
3. Regulatory application

2.2 Workshop structure

41 scientific experts from academia, governmental agencies and industry with experience in ecological risk assessment, ecology and statistics participated in a workshop held in Amsterdam from the 11th to 13th of February 2014. The workshop was organised around 18 presentations, 2 case studies and 3 syndicate discussion sessions where the ecological relevance of an SSD, SSD statistical models for deriving toxic thresholds (HC5/PNEC) for aquatic communities and regulatory applications were discussed. The deliberations from the 4 session subgroups for each of the 3 syndicate themes were addressed in subsequent plenary sessions and a final plenary session identified key points, consensus conclusions and research needs. A list of participants is given in Appendix A, and the programme is detailed in Appendix B.

2.3 Workshop aims and objectives

Workshop aims

The workshop reviewed current thinking on when and how species sensitivity distributions (SSDs), should be used and discussed how the methodology might be further developed. The workshop considered 3 key aspects:

1) What is the ecological relevance of an SSD?

- Are ecologically relevant assessments being made? Are regulatory protection goals explicit and clear? Are they set in relation to environmental quality? How do prospective and retrospective approaches differ?
- Are all species of equal importance, or are there keystone species that are more important than others? If so, how might these be accounted for?
- Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness? Should representative assemblages/communities (archetypes) to represent different typologies be developed? Should protection goals account for local community composition?
- How does aquatic community sensitivity vary with species composition? (summary of and developments since Pellston Classic workshop 2001- Ecological Variability: Separating Natural from Anthropogenic Causes of Ecosystem impairment)
- How can knowledge of chemical mode of action (MoA) help construct SSDs for HC5 estimation?

2) What SSD statistical models are available for deriving toxic thresholds (HC5/PNEC) for aquatic communities?

- Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.
- As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?
- Do models that utilise prior knowledge, e.g. aquatic toxicity data sets on many species, provide advantages over other methods?
- Are current modelling success criteria, such as those identified in the REACH Technical Guidance Document (TGD) (ECHA, 2011), sufficient, overly prescriptive or insufficient?

3) Regulatory application

- Would the methods reviewed in this workshop be accepted for use in regulatory assessments under current guidance? If not, what steps would be needed to facilitate their acceptance in the future?
- Should current guidance on the use of SSDs be revised in the light of the issues and approaches discussed in this workshop?
- What implications are there for the interpretation of SSDs and HC5s in risk assessment and risk management?

3. PRESENTATION SUMMARIES

3.1 Sense, simplicity and successes of SSDs in environmental protection, assessment and management

Leo Posthuma
RIVM, the Netherlands

This contribution presented the versatility of the use of species sensitivity distributions (SSDs) in the contemporary practices of environmental protection, assessment and management in the context of environmental stress. There is sense, simplicity and success - despite various shortcomings of the approach.

Some decades ago, the observation was made that - like many phenomena that exhibit variation - the sensitivities of different species towards a toxic compound were distributed in a way that could be described by a statistical model: this marked the birth of the concept known as the species sensitivity distribution. This was a timely concept, since it helped to solve the questions of that time. Soils and waters were affected by various compounds emitted to- or present in the environment, naturally or from past emissions, and the SSD concept helped to set Environmental Quality Criteria. SSDs were one of the methods used to derive the so-called PNEC for ecosystems. Comparison of a PEC (Predicted environmental concentration) with a PNEC became an important practical approach in environmental policies that aimed at reducing and limiting adverse impacts in ecosystems, since a PEC/PNEC-ratio higher than unity signals a potential for undesired effects. In this context, key issues of concern were and are amongst others quality, number and representativity of input data of SSDs, statistical model choice, and the definition of the PNEC itself, given an SSD. The current workshop focuses, with the science of today, on strengths and weaknesses of this use of the method - to support the best possible decision making.

Holistic goals have been set in environmental regulations, and they need be made operational. The Water Framework Directive (EC, 2000) states that water bodies should reach Good Ecological Status next to Good Chemical Status. When a water body is impacted, this can be the result of a suite of stressors, not necessarily being chemicals or their mixtures. Furthermore, monitoring has revealed that exceedances of the Quality Criteria are observed frequently. These kinds of triggers have prompted attention for the use of SSDs in another way, namely: to derive a local hazard level from an ambient concentration. In combination with mixture modelling, this SSD-use yields an estimated value for the toxic pressure (of single chemicals or mixtures) of an environmental sample. This use has been applied many times, in disciplines and approaches as variable as eco-epidemiological diagnosis of local impacts in aquatic ecosystems on a landscape scale, determination of sanitation urgency in soil management, derivation of impacts of chemical emissions in product Life Cycle Analysis, derivation of the Chemical Footprint of current chemical emissions in an area, and assessment and management of chemical disasters around the globe by UN-field teams.

In total, the sum of the ‘reverse use’ (deriving PNECs) and of the ‘forward use’ (deriving local hazard) of SSDs has grown into a wide field, which encompasses not only single chemicals, but also themes as variable as underwater noise, temperature and radionuclides, and techniques such as field-SSDs. As a response to the high practical importance of SSDs in environmental protection, assessment and management, various

studies focused on the validity of model outcomes. What does it mean when environmental concentrations increase? Does the predicted fraction of species affected relate to observed impacts? And if so, is this a straightforward one-to-one relationship, or at least a linear one? Various studies allow for the derivation of confirmation statements on SSD output, and suggest that a nuanced view is needed.

SSDs have sense, in that they apparently help to address contemporary questions for a variety of environmental problem definitions.

SSDs are simple, and they contain not a single bit of ecology - but despite that, confirmation studies highlight a basic relevance of SSD output for those problem definitions. The use of SSDs has resulted in successes, which might best be envisaged by imagining the absence of concepts like the PNEC on the one hand, and the potentially impacted fraction of species on the other.

Reflections on the sense, simplicity and successes of SSDs provide the context within which SSD-oriented improvements can be designed. These reflections, though not necessarily complete, can serve as a basis for the workshop and for further developments and use of SSDs.

3.2 Ecological limitations of SSDs

Lorraine Maltby
The University of Sheffield, UK

Species sensitivity distributions are generally derived using data from single-species toxicity tests. The species used in these tests are often from a limited geographic and/or habitat range and toxicity is measured in the absence of interspecific interactions. SSDs are used to assess the risk of chemicals to ecological assemblages containing many interacting species, often in a range of habitats (e.g. rivers, ditches, ponds) in different geographic regions. An extensive and detailed analysis of toxicity data for 67 pesticides (16 insecticides, 9 herbicides, 42 fungicides) has considered the implications of generating SSDs using data sets that contain toxicity data for species from different broad taxonomic groups (e.g. primary producers, fish, arthropods etc), from different habitats (e.g. fresh water, sea water, lotic, lentic) and from different parts of the world (e.g. temperate, tropical, UK). Moreover, this analysis considered how toxicity thresholds derived from SSDs compared to toxicity thresholds derived from multispecies, semi-field studies. The key findings of the analysis reported by Maltby *et al,* 2005, 2009; van den Brink *et al*, 2006, are:

- the species sensitivity profiles for 30 fungicides could be described by a single SSD, but separate SSDs for different taxonomic groups were required for herbicides, insecticides and the remaining 12 fungicides. Herbicides were most toxic to primary producers, insecticides were most toxic to arthropods but most fungicides were general biocides.
- Toxicity data for species from different geographical areas can be combined as long as the SSD is based on the sensitive taxonomic group(s). Similarly, toxicity data for species from different habitats can be combined as long as taxonomic differences are accounted for. The potential effects of test conditions on exposure, and hence sensitivity, should be considered whenever data are collated across different studies, irrespective of the geographical region in which the data were generated or the habitat from which the species were obtained.
- Threshold values derived from SSDs can be regarded as protective when compared with threshold values in multispecies studies. SSD-derived values (HC1, HC5, LLHC5 [Lower limit hazardous concentration for 5% of species]) were compared with the NOECeco values derived from the most sensitive structural or functional endpoint in each of 32 mesocosm studies. For the majority of pesticides, the HC1 or LLHC5 were lower than the NOECeco, as was the HC5/3.

The main conclusion of this analysis was that, if based on the most sensitive taxonomic group (determined by mode of action), SSDs derived from a collection of species from different habitats and geographies tested in the absence of interspecific interactions, can be used to derive threshold values that are protective of effects in more ecological complex systems. However, a limitation of this analysis is that the toxicity data sets used to derive SSDs generally do not contain information on all taxonomic groups and information on heterotrophic microorganisms, which are known to play key roles in many ecosystems, is generally absent. The very limited information available indicates that microbially-mediated functions (e.g. decomposition) may be protected by threshold values based on non-microbial toxicity data, but this is an area that requires more investigation.

3.3 How do species traits influence sensitivity and herewith species sensitivity distributions?

Paul J. van den Brink (Cancelled due to ill health)
Alterra and Wageningen University, the Netherlands

Species sensitivity distributions assume that sensitivity to toxicants within target species is random. While the SSD approach has shown to be promising, it is limited by the fact that data are sparse for most compounds, and that these data are largely based on the lethal responses of a small group of testing laboratory species. Here an alternative approach, based on the hypothesis that organisms' sensitivity to stress is a function of their biology, and can be predicted from species traits such as morphology, life history, physiology and feeding ecology is presented.

Examples of how species traits have been used to explain the differences in sensitivity between species will be shown in this talk.

1. Using data from the US EPA's AQUIRE database, we found that 4 species traits explained 71% of the variability in sensitivity to toxicants within a group of 12 species exposed to 15 chemicals. Our results indicate that this approach is promising, but effort is needed to compile species trait information to increase the power, precision and taxonomic representativeness of this approach.
2. Secondly, we mined existing data on organophosphate, carbamate and pyrethroid toxicity and mode of action and also species trait information. We linked taxon sensitivity to their traits at the family level in order to generate empirical and mechanistic hypotheses about sensitivity-trait relationships. In this way, we developed a Mode-specific sensitivity (MSS) ranking method, and tested this at the taxonomic level of family and genus. The MSS rankings were successfully linked to existing trait data in order to identify traits with predictive potential. Single traits as well as combinations of traits can be used to predict laboratory sensitivity to the substances tested, although associations were not as strong as in previous studies.
3. We also explored whether and in what ways traits can be linked purposefully to mechanistic effect models to predict intrinsic sensitivity using available data on the acute sensitivity and toxicokinetics of a range of freshwater arthropods exposed to chemicals, using the insecticide chlorpyrifos as an example. The results of a quantitative linking of 7 different endpoints and 12 traits demonstrate that while quantitative links between traits and/or trait combinations and process based (toxicokinetic) model parameters could be established, the use of simple traits to predict classical sensitivity endpoints yields less insight. Future research in this area should include a quantitative linking of toxicodynamic parameter estimations and physiological traits, and requires further consideration of how mechanistic trait-process/parameter links can be used for prediction of intrinsic sensitivity across species for different substances in environmental risk assessment (ERA).

3.4 Field validation of species sensitivity distributions

Adam Peters
WCA Environment, UK

There is a requirement in the technical guidance for quality standards derived under the European Water Framework Directive (WFD) to consider evidence from field and mesocosm studies, where such data exists. The same principle can also be applied to any chemical substance for which a robust ecological threshold (e.g. PNEC) has been derived, for example through the derivation of a species sensitivity distribution. Several different approaches towards performing these types of assessments were outlined, including examples of real assessments. The advantages and limitations of various assessment approaches were considered for both whole community assessments and assessments that are targeted at particularly sensitive organisms.

In order to evaluate relationships between metal exposures and benthic community metrics, the bioavailability of the metals must be calculated for each site. Several approaches can be taken towards the assessment of PNEC values, including simplistic assessments of ecological quality at different exposure levels and the derivation of limiting functions (comparable to a traditional dose response relationship). Assessments can be based on the whole community, subsets of the community, groups of taxa, or an individual taxon. Analyses based at the level of the whole community may lack the sensitivity to identify slight effects on particularly sensitive species or families. Reducing the diversity of organisms assessed increases the uncertainty in the assessment, particularly for reference based methods. This presentation reviewed approaches towards the identification of those taxa that should be considered as sensitive to a particular pollutant.

A novel approach for bridging the gap between quality standards based on laboratory ecotoxicity studies and site-specific local aquatic communities was also outlined. This approach aims to take account of variation in the composition of ecological communities, and the effect that this may have on the sensitivity of the community to a particular pollutant. This was illustrated with an example for deriving site-specific thresholds for zinc in an area affected by historic mining activities.

3.5 Derivation of toxicity thresholds for LAS – integration of QSARs, SSDs, mesocosms, and field data

Scott Belanger
Procter & Gamble, USA

Linear alkylbenzene sulfonate (LAS) has been one of the most heavily used anionic detergent surfactants globally since its introduction to the market in the 1960s. As such, it has a rich information base spanning physico-chemical properties, specific analytical methods applicable to all environmental matrices, acute and chronic toxicity, bioaccumulation, field monitoring data, and assessments using stream mesocosms. In this talk, this information was reviewed in support of an integrated approach that translates chronic toxicity data on pure LAS materials and technical mixtures for comparison to experimental stream mesocosm studies on LAS. Using the toxicity normalization method using local quantitative structure activity relationships (QSARs), chronic laboratory toxicity data for 19 species representing 9 phyla were summarized in various species sensitivity distributions that were also probed to understand the robustness of the SSD itself. The resulting HC5 was 0.19 mg/L (95% confidence interval of 0.06-0.38 mg/L). Leave-one-out and add-one-in Monte Carlo simulations were used to quantitatively and qualitatively evaluate 'what-if' scenarios regarding the generation of additional data and clearly demonstrated that the HC5 would not benefit from additional data generation. A high quality experimental stream mesocosm study yielded a long-term NOEC value of 0.27 mg/L suggesting the SSD remained appropriately conservative. In order to provide perspective on the relationships between SSDs, mesocosms, and field studies with regards to application factors, the ecological context of the stream mesocosm was also reviewed. The mesocosm was demonstrated to be an ecological equivalent of natural, low order, relatively unperturbed streams systems. Ecological investigations on trophic dynamics, nutrient processing regional community structure, combined with statistical and biological sensitivity of the test system support the use of low application factors for the mesocosm and SSD outputs as well as their predictive nature to derive safe concentrations for tested chemicals.

3.6 Field-based SSD and community sensitivity distribution as alternative ways for field validation of the PNECs derived from laboratory based approaches

Kenneth Mei Yee Leung

The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Hong Kong, China

The determination of PNECs and sediment quality guidelines (SQGs) of toxic chemicals in marine sediment is very crucial in ecological risk assessment, sediment quality management (e.g. mud disposal in the sea) and environmental remediation (e.g. dredging of contaminated mud). However, current methods of deriving sediment PNECs are primarily based on toxicity data generated from laboratory ecotoxicity bioassays that often lack ecological realism. To tackle this issue, we have developed 2 novel alternative approaches to scientifically derive site-specific SQGs by utilising field data of benthic biodiversity and contaminant concentration which are concurrently measured in sediment samples collected from the area of concern.

In this talk, I first described the principle of these field-based approaches. Secondly, I introduced the field-based species sensitivity distributions (f-SSDs) approach, which is based on the relationship between species abundance and contaminant level (Leung *et al*, 2005; Kwok *et al*, 2009). Since its establishment, f-SSDs have been utilised in different parts of the world such as Europe, Hong Kong, New Zealand and the United States. The Norwegian continental shelf and the marine environment of Hong Kong were taken as examples to illustrate the methodology. Thirdly, I presented the community sensitivity distributions (CSDs) approach which is founded on the relationship between species density and contaminant level, and makes use of Empirical Bayes methods (Gilbert *et al*, 2014). Overall, the field-data-derived SQGs appear to be more environmentally relevant and ecologically realistic. The f-SSD and CSD can be directly adopted as 'effect distributions' for probabilistic risk assessment. The field-data-derived SQGs can be employed as site-specific guidelines, and used to validate the current PNECs or SQGs derived from laboratory ecotoxicity data. Finally, the limitation of these field-based approaches were discussed, while their recent development and application in different countries were highlighted.

3.7 HC5 estimation in SSDs revisited

Tom Aldenberg
RIVM, the Netherlands

Species sensitivity distributions, in their basic form defined as univariate continuous statistical distributions over a logarithmic species sensitivity concentration axis for a particular chemical substance, can be applied in environmental risk assessment to estimate a PNEC for that toxicant. This PNEC is in many cases implemented as a statistical estimate of the log HC5 concentration. This minimalist model, originally due to Kooijman (Kooijman, 1987) and Van Straalen (Van Straalen and Denneman, 1989), needs extension to address a multitude of thinkable challenges, e.g. with regard to species selection, ecosystem representativeness / functioning, data quality, statistical model selection, and predictive evaluation of the SSD and its quantiles.

This presentation first reviewed how we handled the uncertainty of the log HC5 for the Logistic and Normal distribution, from a Bayesian viewpoint. Second, we developed the estimation of the so-called *predictive distribution* - formally the mean of the Bayesian spaghetti plot SSD - in order to pinpoint a single-curve SSD for a given statistical family. This leads to an *improved log HC5* - or other quantile - estimate, to better reflect uncertainty due to small sample size. Presently, we consider the ubiquitous median estimate log HC5 as being unrealistically insensitive to small sample size, hence risking lack of conservativeness. This is compounded by the 5^{th} and 95^{th} confidence limits of log HC5 uncertainty often not being reported. The Bayesian predictive distribution method spawns *a new table of extrapolation constants*, addressing both chronic and acute species sensitivity data, depending on the basic fraction affected. The sensitivity of these new extrapolation constants is evaluated in the light of the REACH-required samples sizes of 10, preferably 15. A recurring concern is the effect of log species toxicity *data uncertainty*. Operationally, this may derive from having multiple data for the same species, from *dose-response curve confidence limits*, from *QSAR-estimated toxicity data* with associated confidence, and possibly a host of other sources of uncertainty. Intuitively, one would expect data uncertainty to further lower old - as well as new - log HC5 estimates, but methods of hierarchical modelling reveal that the reverse is the case: the more variation has to be attributed to the individual species points, the less variation remains for the SSD itself. Surprisingly, theory, as well as numerical experiments, show that *the effect of data uncertainty is quite modest*, leading to the recommendation to take the mean of log data point uncertainty, and continue with the old, or updated, extrapolation methodology, *as if data were certain*. Averaging multiple species data was already recommended in REACH (EC, 2006). It follows that using such averages per species, or employing point estimates, i.e. expected values, through model-estimated species sensitivities, only leads to slightly increased conservative, i.e. lower estimates of PNEC values being pursued. New insights of predictive SSD and the effect of data uncertainty would both help to alleviate the need for assessment factors addressing these particular issues.

3.8 Assessment factors for deriving PNECs: Food for thought

Ad M.J. Ragas
Radboud University, the Netherlands

Within regulatory contexts such as REACH and the European Water Framework Directive (WFD), assessment factors are used to derive safe exposure levels for aquatic and terrestrial ecosystems from single species toxicity data. These safe exposure levels are also referred to as PNECs. If toxicity data are available for a limited set of aquatic species – e.g. an alga, a daphnia and a fish – the lowest value is typically divided by an assessment factor to arrive at the PNEC. The value of this assessment factor varies between 10 and 1000, depending on the number and the nature of the available data. If chronic NOECs are available for an extensive set of aquatic species (i.e. > 15 species covering at least 10 different taxonomic groups), the 5th percentile of the species sensitivity distribution is determined and an assessment factor of 1-5 is subsequently applied to arrive at the PNEC. The main aim of the current contribution is to formulate recommendations for improving the use of assessment factors in deriving PNECs. These recommendations are based on a statistical analysis of a large set of chronic toxicity data resulting from aquatic single species tests and mesocosm experiments.

A database with chronic single species NOECs on 20 different chemicals was compiled based on data reported in the open literature. Chronic mesocosm data were found for 6 of these substances and were also included in the database. For each of the substances in the database, the 5th percentile of the SSD (HC5) was determined. This HC5 was then compared with:

- the PNEC reported in the mesocosm experiments (if available);
- PNECs derived by applying a safety factor of 10 to the lowest value of a limited dataset of 3, 6 or 9 NOECs. These datasets were generated by parametric bootstrapping of the available single-species NOECs.

Mesocosm PNECs were generally lower than the HC5, with 2 notable exceptions, i.e. lindane and dimethoate, which can be explained by the limited set of species in the mesocosm. The HC5 is on average a factor of 2.0 lower than the PNEC derived from a set of 3 chronic NOECs. This difference increases to a factor of 4.5 and 7.2 for datasets with 6 and 9 chronic NOECs, respectively. Based on these results 2 general recommendations are formulated:

- The assessment factor of 10 that is currently being applied to the lowest value of small datasets (i.e. alga, daphnid and fish) should be differentiated depending on the number of available data, e.g. a factor of 20 if one value is available for each taxonomic group, but a value of 5 when 3 or more values are available for each taxonomic group.

- The default assessment factor of 2 is suggested for the HC5 of the SSD. This default value can be further refined based on the specific characteristics of the available toxicity data, i.e. representativeness, mode of action, interspecies variability and uncertainty.

3.9 Weight of evidence approaches for deriving HC5s

Sandrine Andres
INERIS, France

Experience gained in developing Quality Standards (e.g. PNEC) in the framework of the REACH Technical Guidance Document (ECHA, 2011) shows that only a few substances can benefit from the use of an species sensitivity distributions, even if substances appear after an initial assessment as data rich substances. The main drawback is the lack of validated studies for the additional taxa. Indeed, the level of standardisation for testing these additional taxa (such as mayfly, dragonfly, amphibian, rotifer, molluscs, etc...) is usually lower than for the regular algae/daphnid/fish simplified trophic chain. As a consequence, this additional information is often not used for the assessment.

In order to make the best use of all the information available, a Multi-Criteria Decision Analysis (MCDA) tool was developed in the framework of the research project AMORE (Multi-Criteria Analysis for the Development of a Decision Support Tool for the prevention of Environmental Risks). This tool is based on weight of evidence (WoE) methodology, which aims to improve the evaluation of ecotoxicological data, through the assessment of their relevance and reliability for the definition of SSDs. The methodology allows us to rank the acceptability of ecotoxicological data for further use in the risk assessment process and therefore optimise their influence in the production of reliable SSDs, through a weighted bootstrap modelling procedure for data resampling.

In this project, it was hypothesised that the SSD can be based on all available ecotoxicity data, which can be heterogeneous and often non comparable. These data can be obtained through different approaches (e.g. experimental or even modelling) and conditions, e.g. the protocol can be standardised or not; time duration can vary among experiments, leading to chronic or acute data; different physiological endpoints can be observed, e.g. mortality, growth, reproduction; statistics used for interpreting data can differ, e.g. leading to NOEC or ECx.

The methodology is based on the assessment of a hierarchically structured set of 57 criteria, which is used for assigning a quantitative score to every ecotoxicological datum and was created based on the review of the state of the art frameworks for the assessment of ecotoxicological data. The different endpoints are analysed based on their production method and specifically on 3 main aspects: the 'Experimental Reliability', the 'Statistical Reliability' and the 'Biological Relevance' of the experimental or modelling protocol used. This assessment has been developed with the contribution of an expert panel of scientists on ecotoxicology. Knowledge and preferences of experts have been gathered through a participatory process, and is used for the calculation of the aggregated reliability scores of data. The nature of the process mandates the use of Fuzzy Logic during the aggregation phase, for handling the inherent uncertainty which appears in the form of unreported information, as well as possible lack of knowledge of the experts.

This approach allows for a weighed use of the available information available in a weight of evidence perspective.

3.10 Sample size in PNEC derivation

Scott Dyer
Procter & Gamble, USA

SSDs have been used to develop water quality criteria (e.g. PNECs) and other protective environmental concentrations (e.g. HC5). These criteria typically require large datasets (e.g. US EPA ambient water quality criteria utilize at least 8 acute toxicity values from several taxa spanning 3 trophic levels, fish, invertebrates and plants) of measured toxicity values. However, there has been a considerable debate regarding the minimum requirements for establishing protective concentrations, such as the HC5, within the scientific and regulatory communities. For organizations needing to establish these criteria, questions remain whether the addition of taxa into the SSD will greatly change the criterion. Is it possible that the addition of taxa will not change the HC5 and thereafter the PNEC? Is there a law of diminishing returns for expanding the number of taxa incorporated into an SSD? If so, then understanding factors that dictate the lack of need for additional taxa would result in appropriate PNECs without undo cost and time. To explore this question we developed 3 distributions, each assuming normality: 1) wide (toxicity values ranged 4 orders of magnitude); narrow (values ranged by 1 order of magnitude; and 3) mixed (sensitive taxa corresponded to 1/3 of the distribution and more tolerant taxa the remaining 2/3). In each distribution the numbers of taxa sampled were varied as well as the number of replicate tests/taxa. Monte-Carlo was used to sample each distribution 1000-times. The following conclusions were noted: 1) the spread of the HC5 values were proportional to magnitude of the spread of the toxicity data per distribution. For example, the 5^{th} percentile and 95^{th} percentiles of the range of HC5s from the wide distribution were 4-orders of magnitude apart. 2) Increasing the number of taxa sampled per distribution increasingly approximated the 'true or ideal' HC5. 3) There appeared to be a law of diminishing returns with increasing the number of taxa to approximate the ideal HC5. Approximately 10 taxa were sufficient to within a factor of 2 of the ideal HC5 for predicting any distribution. Considering this, discussion is warranted in the cost versus benefits of obtaining more taxa to derive an ideal HC5. While not presented, we also found the number of replicates per taxa did not significantly change the HC5, though they did narrow the range of HC5s. The common occurrence of a mixed distribution (e.g. algae more sensitive than invertebrates and fish) did not change the conclusions from the narrow and wide distributions. The authors recognize that this exercise was theoretically-based, however, the findings simplify future discussions regarding how many taxa are needed to obtain an HC5 to derive a PNEC.

3.11 How to extrapolate across 100,000+ substances, sites and species with SSDs ?

Jan Hendriks
Radboud University, the Netherlands

Each second, 1 new chemical is added to the more than 65,000,000 already registered. In the EU, 100,000+ compounds are awaiting assessment while 1,500,000 contaminated sites potentially require clean-up. Worldwide, 8,000,000+ species, of which 10,000+ endangered, need protection (Hendriks, 2013).

At the same time, empirical research is severely limited by financial, practical and ethical constraints. Assessing 100,000+ substances at 100,000+ sites threatening 100,000+ species obviously cannot be achieved by toxicological testing only. As an alternative, I suggest that we focus on simple models. Instead of going for statistical regressions with the highest explained variability, we might attach more value to meaningful equations of which the coefficients and exponents can be interpreted physically.

We have derived and collected SSDs on toxic and non-toxic stressors to discern patterns across stressors, species and endpoints. In this contribution, some examples were discussed. We looked at (1) intra-species and inter-species variability, (2) the number of species included in an SSD, (3) SSDs across modes of action, (4) the combined use of SSDs for toxic and non-toxic stressors, (5) 'field'-based SSDs (PNOFs) and (6) *in vitro* biomarker SSDs to *in vivo* bioassay SSDs (Azevedo *et al*, 2014; De Hoop *et al*, 2011; Elshout *et al*, 2013; Fedorenkova *et al*, 2010, 2012, 2013; Golsteijn *et al*, 2012, 2013; Hendriks, 2013, Hendricks *et al*, 2013; Smit *et al*, 2009.

3.12 Interspecies correlation estimation (ICE) models predict supplemental toxicity data for SSDs

Sandy Raimondo
Environmental Protection Agency, USA

Species sensitivity distributions require a large number of toxicity values for a diversity of taxa to define a hazard level protective of multiple species. For most chemicals, measured toxicity data are limited to a few standard test species that are unlikely to adequately represent ecological communities. Interspecies correlation estimation (ICE) models are log-linear least squares regressions that predict the acute toxicity to untested taxa from known toxicity of a single surrogate species. A suite of ICE models is developed from a comprehensive, standardized dataset of acute toxicity with the goal of maximizing the number of potential species for which toxicity can be predicted while minimizing extraneous sources of variation in the models. The United States Environmental Protection Agency houses 3 ICE databases: aquatic animals (vertebrates and invertebrates; 5501 records, 180 species, 1266 chemicals), algae (1647 records, 69 species, 457 chemicals), and wildlife (birds and mammals; 4329 records, 156 species, 951 chemicals). Approximately 2400 models have been developed from these databases and made available through the Web-based Interspecies Correlation Estimation internet application (Web-ICE; http://epa.gov/ceampubl/fchain/webice/).

ICE models were validated using leave-one-out cross validation and sources of model uncertainty were evaluated. Toxicity predictions are most accurate for models with closely related taxa pairs, with over 90% of cross-validated values predicted within 5-fold of the measured value when the surrogate and predicted taxa are in the same family. Model mean square error and prediction confidence intervals should be considered when evaluating an ICE predicted value. Models built with a single mode of action (MOA) were often more robust than models built using toxicity values with multiple MOAs, and improve predictions among species pairs with large taxonomic distance (e.g. within phylum). SSDs developed solely from ICE-predicted toxicity values produce hazard levels with an average factor of 3.0 and 5.0 of those developed with all measured data for aquatic species and wildlife, respectively. For chemicals in which more measured data are available, ICE models may be used to augment datasets to increase species diversity in SSDs. Compared to SSDs developed from only measured data, the uncertainty of ICE model predictions contributes less variability to hazard levels than variance due to species composition. Through extensive study of ICE model evaluation and uncertainty and their application in developing SSDs, ICE generated toxicity values have been demonstrated to provide a statistically sound approach to supplementing datasets to generate SSD-based hazard levels applicable to ecological risk assessments.

3.13 HC5s from taxonomically structured hierarchical SSDs

Peter Craig
Department of Mathematical Sciences, Durham University, UK

One approach to deriving the predicted no-effect concentration for a chemical is to use a species sensitivity distribution model to estimate the hazardous concentration affecting p% of species (HCp), where p is usually 5. Many questions have been raised about both principles and application of SSDs but the concept has nevertheless been found to be useful.

Analysis of a database of acute aquatic toxicity test results reveals several features which should be addressed by SSD methodology, including: (a) inter-species correlation; (b) tendencies of particular species to one or other end of the sensitivity distribution; and (c) inter-test variation. In earlier work (Dyer *et al,* 2006, Craig *et al,* 2012, Hickey *et al,* 2012), each issue has been addressed on its own. Addressing them collectively requires multivariate statistical modelling.

We present a Bayesian hierarchical model of variability and uncertainty for: (i) sensitivities of species to a chemical undergoing assessment and (ii) a database of relevant test results for other chemicals (Craig, 2013). Bayesian statistical methodology has several advantages over traditional non-Bayesian methodology which is intended primarily for analysing experimental data. It can incorporate data, expert judgements and results of meta-analyses. It provides a collective description of uncertainty for all components of a model, a coherent mechanism for revising uncertainty when additional data become available, and a decision-making framework.

Our model generalises the Aldenberg and Jaworska (2000) single randomly-sampled-chemical log-normal SSD model and addresses issues (a)-(c). It models inter-species correlation by building species tendencies and sensitivities hierarchically, based on the taxonomic classification of species. The taxonomic structure seems natural and enables a better description of the available data but means that it is necessary also to specify an eco-taxonomic scenario: the taxonomic structure of the community being protected by the HCp. The HCp is then scenario-specific, being the p^{th} percentile of sensitivity to the chemical for species in the scenario. The model automatically delivers a quantitative assessment of uncertainty to accompany the HCp estimate.

The model is trained, using the same database as for the original data analysis, and is then ready for application to other chemicals. The trained model is available as software, known as hSSD, for application to test data for a new chemical in the user's chosen eco-taxonomic scenario. The workshop included a demonstration of hSSD which is one of the methodologies used in the workshop case studies.

As an illustration, we applied the trained model to a chemical for which a substantial number of test data are available. For the eco-taxonomic scenario, we took the Kent river scenario that was developed for the workshop case studies. We highlighted the prediction from the model for the true sensitivity of each species in the scenario; the predictive uncertainty is high for species which are taxonomically distant from all tested species and low for tested species.

3.14 Demonstration of the web-based interspecies correlation estimation (Web-ICE) modelling application

Sandy Raimondo

Environmental Protection Agency, USA

The Web-based Interspecies Correlation Estimation (Web-ICE) modeling application is available to the risk assessment community through a user-friendly internet platform (http://epa.gov/ceampubl/fchain/webice/). ICE models are log-linear least square regressions that predict acute toxicity (LC50/LD50) of a chemical to a species, genus, or family based on estimates of relative sensitivity between the taxon of interest and that of a surrogate species. Web-ICE v 3.2 includes over 1440 models for aquatic animal taxa, 100 models for algae, and 852 models for wildlife taxa. Web-ICE has modules that predict toxicity to one taxa of interest at a time while providing detailed information on model parameters. It also has species sensitivity distribution and endangered species modules that produce toxicity values to multiple species based on the number of surrogates entered. In the SSD module, a user can enter up to 20 surrogate species which are used to predict toxicity to all predicted taxa possible. The entered surrogate and predicted toxicity values are used to develop a log-logistic probability distribution and estimate a hazard level equivalent to either the 1^{st}, 5^{th} or 10^{th} percentile of the distribution. Users can also enter multiple surrogate toxicity values into the endangered species module, which are used to calculate predicted species, genus, and family level sensitivity for selected endangered species. Both the SSD and endangered species modules provide exportable data files of predicted results. A demonstration of the Web-ICE will familiarize participants with the functionality of the application and provide examples of its use for single taxon predictions, SSD generation, and development of endangered species toxicity reports.

3.15 Regulatory applications of SSDs in European regulations

Paul Whitehouse
Environment Agency, UK

Regulatory frameworks like the Water Framework Directive (EC, 2000), Marine Strategy Framework Directive (EC, 2008), Plant Protection Products Directive (EC, 2009 and REACH (EC, 2006) are far-reaching pieces of legislation that require us to identify and manage pressures on the environment, including toxic chemicals. Assessing the hazard posed by chemicals is central to chemical risk assessment and also to the derivation of Environmental Quality Standards (EQSs).These play a key role in identifying risks and helping manage emissions to ensure wildlife and human health are not adversely impacted.

This presentation was in 2 halves. In the first half, I compared the approaches to setting environmental thresholds for chemicals in several European regulatory regimes. The comparison paid particular attention to the use of species sensitivity distributions e.g. the data requirements, use of other lines of evidence (e.g. field and mesocosm data) and use of assessment factors for dealing with residual uncertainty that remains after addressing interspecies differences in sensitivity. The second half of the presentation focused on the derivation and role of EQSs in the Water Framework Directive. There have been important technical developments in the derivation and application of EQSs in recent years, some of which have been captured in EU Technical Guidance (EC, 2011). Whilst deterministic methods for deriving EQSs remain the only option in some cases, SSDs are now the method of choice, including standards for bioavailable metals. I briefly reviewed the experience of EU Member States in using such approaches, the sources of variability that can give rise to different outcomes when different jurisdictions derive EQSs, and how different lines of evidence can be combined to derive EQSs. Finally, I suggested where further development in EQS derivation would be welcomed. For example, can we use SSDs to calibrate EQSs to the biological thresholds used within WFD to classify water bodies?

3.16 Regulatory use of SSDs in Australia and New Zealand

Michael Warne

DSITIA Science Delivery, Australia

Author: Warne MStJ[1], Batley GE[2], Braga O[3], Chapman JC[4], Fox D[5], Hickey C[6], Stauber JL[2], and Van Dam R[7].

Affiliation:

[1.] Water Quality and Investigations, Environmental Monitoring and Assessment Science, Science Delivery, Department of Science, Information Technology, Innovation and the Arts, Brisbane, Queensland, Australia.
[2.] Centre for Environmental Contaminants Research, CSIRO Land and Water, Lucas Heights, NSW, Australia.
[3.] Department of Sustainability, Environment, Water, Population and Communities, Canberra, Australia.
[4.] Office of Environment & Heritage, Lidcombe, NSW, Australia.
[5.] Environmetrics, Melbourne, Victoria, Australia.
[6.] National Institute of Water and Atmospheric Research (NIWA), Hamilton, New Zealand.
[7.] Environmental Research Institute of the Supervising Scientist, GPO Box 461, Darwin, NT, Australia.

Australia and New Zealand, along with many other countries, use risk-based approaches to manage and regulate chemicals in the environment. A key component of the risk approach has been the use of species sensitivity distribution (SSD) methods. SSDs are central to the Australian and New Zealand approach to managing the quality of various environmental compartments (water, sediment and soil), of additives to soils (biosolids and mineral fertilisers) and in conducting environmental risk assessments. Australia and New Zealand developed a new SSD method called BurrliOZ (http://www.csiro.au/Outcomes/Environment/Australian-Landscapes/BurrliOZ.aspx) that uses the distribution from the Burr Type III family of statistical distributions that best fits the sensitivity data. This method can therefore provide a good fit to many more datasets than can SSD methods that use a single statistical distribution. The Australian and New Zealand Guidelines for Fresh and Marine Water Quality (http://www.environment.gov.au/resource/australian-and-new-zealand-guidelines-fresh-and-marine-water-quality-volume-1-guidelines) were released in 2000 and are currently undergoing a review. This is examining the framework used to derive the guidelines (called trigger values). Key recommendations arising from the review are: increasing the types and sources of data that can be used; working with industry to permit the use of commercial-in-confidence toxicity data; increasing data requirements; improving the software used to calculate trigger values; increasing the rigour of site-specific trigger values; improving the method for assessing the reliability of the trigger values; providing guidance of measures of toxicity and toxicological endpoints that may, in the near future, be appropriate for trigger value derivation. A new set of sediment quality guidelines and new trigger values for a number of existing metals will be derived. In addition, trigger values for a range of organic chemicals focussing on pesticides, pharmaceuticals and personal care products will be derived. Finally, a weight of evidence approach is being included into the guidelines. These changes will improve the number and quality of the trigger values that can be derived and will increase end-users' ability to understand and implement the guidelines in a scientifically rigorous manner.

The water quality guidelines are generic - a single value that applies to all waterways. The only exception being the trigger values of some metals that can be modified using hardness algorithms. In contrast, the Australian guidelines for contaminants in contaminated soils and in biosolids, are wherever possible, soil-specific. That is a matrix of guidelines are generated for each contaminant depending on the values of various soil physico-chemical properties known to modify toxicity. This presentation discussed the ways that SSDs are used in Australia and New Zealand and the proposed changes arising from the current review of the Australian and New Zealand water quality guidelines.

3.17 Use of SSD in China

Fengchang Wu
Chinese Research Academy of Environmental Sciences, China

Species sensitivity distributions are usually used in the development of water quality criteria (WQC) and require a large number of toxicity values to define a hazard level to protect the majority of species. In the present study, we introduced the specific use of SSD in the study of water quality criteria in China. As case studies, WQC for representative water-body pollutants in China using SSD were conducted. i.e. toxicological data for zinc (Zn), cadmium (Cd), hexavalent chromium (Cr (VI)), benzene, and nitrobenzene were collected from various databases, publications and experimental test data. These toxicological data were screened and constructed into SSD curves. WQC for protection of the freshwater aquatic life in China against 5 representative pollutants were then derived. The values derived in this study were compared with those issued by the US Environmental Protection Agency and the Chinese national environmental standard for surface water to identify factors underlying the differences. The results showed that the SSD curves for the 5 pollutants differed significantly, with the examined aquatic species being generally more sensitive to Zn, Cd, and Cr (VI) than benzene and nitrobenzene.

While SSDs based on measured toxicity values can provide a strong level of confidence for environmental protection, there is still some uncertainty in their applicability for untested species. Additionally, SSD development has been limited to a relatively few chemicals because of the requirement for toxicity data for a broad diversity of taxa. Interspecies correlation estimations (ICE) models may provide great assistance for addressing the development of WQC that are protective of species that cannot be tested. To address this need, we also tried to use ICE-based SSD in deriving WQC for zinc in China. Taken zinc for example, ICE-based-SSDs were generated using 3 surrogate species (common carp (*Cyprinus carpio*), rainbow trout (*Oncorhynchus mykiss*), and *Daphnia magna*) and compared with the measured-based SSD and corresponding HC5. The results showed that no significant differences were observed between the ICE- and the measured-based SSDs and HC5s. Given the similarity of SSD and HC5s for zinc, the use of ICE to derive potential water quality criteria for diverse chemicals in China is proposed. Further, a combination of measured and ICE-derived data will prove useful for assessing water quality and chemical risks in the near future. Above all, the comparative study of SSD in WQC studies may offer guideline values for future WQC studies in China.

3.18 Use of SSD to derive no-effect thresholds for water quality guidelines and ecological risk assessments in Canada

Anne Gosselin

Environment Canada, Canada

Authors: A. Gosselin[1], D.J. Spry[1], S. Dixit[1], S. Teed[2] and M. Bonnell[1]

Affiliations: [1]Environment Canada, Gatineau, Canada; [2]Intrinsik Environmental Sciences Inc., Ottawa, Canada

In Canada, species sensitivity distributions are used to derive 'no effect' thresholds that serve to determine water quality guidelines for aquatic life as well as PNECs in ecological risk assessments of chemicals. The Federal Water Quality Guidelines (FWQGs) are developed to meet the needs of risk assessment and risk management of chemicals under the Canadian Environmental Protection Act, 1999 (CEPA, 1999). In addition, Canadian WQGs are developed under the auspices of the Canadian Council of Ministers of the Environment (CCME) based on priorities identified by federal, provincial and territorial jurisdictions. Moreover, under CEPA 1999, regulatory ecological and human health risk assessments are conducted for substances identified as priorities on Canada's Domestic Substances List.

The FWQGs, CCME guidelines and PNECs used in ecological risk assessments all identify thresholds for aquatic ecosystems that are intended to protect all forms of aquatic life and all life stages for indefinite exposure periods. The methodology used to derive these thresholds is the "Protocol for the Derivation of Water Quality Guidelines for the Protection of Aquatic Life" (CCME, 2007). SSD is the preferred approach for FWQGs, CCME guidelines and PNECs. It follows these steps: toxicity data collection, evaluation and selection, SSD plotting, verification of statistical assumptions including determination of the goodness-of-fit (i.e. selection of the model), and determination of the FWQG, CCME guideline and/or PNEC. They are set at the 5^{th} percentile of the SSD, which may, in the case of the PNEC used for risk assessment, be divided by an assessment factor if deemed necessary.

Examples of the use of SSDs in Canada were presented, including the Federal Water Quality Guidelines and risk assessment for metals (vanadium and uranium) and the antimicrobial triclosan.

3.19 Use of SSDs in the USA – endangered species and water quality criteria

Mace Barron

Environmental Protection Agency, USA

Species sensitivity distributions are used in the United States (US) in the development of national ambient water quality criteria (AWQC), with site-specific and numeric modifications to protect sensitive taxa including threatened and endangered species. The US Environmental Protection Agency (EPA) first used SSDs constructed of acute toxicity values in 1978, with formal guidance issued in 1985 for computing 5th percentile hazard concentrations (HC5) from SSDs constructed of at least 8 families with acceptable toxicity data. Additional minimum data requirements (MDRs) include acceptance of only North American species and specific taxa diversity requirements that have limited the development of AWQC to only 47 chemicals. EPA is currently considering alternative approaches for developing SSD-based AWQC, with the recognition that species composition appears to affect HC5 estimates for aquatic species more than differences in geography or habitat of the assemblage. The protectiveness of SSD hazard concentrations used in endangered species risk assessment remains a concern because of uncertainty in sensitivity compared to standard test species. In a recent study, the relative sensitivities of US federally listed and non-listed aquatic species were compared for a broad range of chemicals. The SSD, HC5s and HC1s were lower than 97 and 99.5% of all endangered species mean acute LC50s, indicating that the use of SSDs as distribution-based risk assessment and criteria development approaches can be generally protective of listed species. A recent US National Academy of Sciences report suggested SSDs should be applied in endangered species risk assessments as an alternative to general uncertainty factors. This presentation gave an overview of US applications of SSDs in AWQC development and listed species assessment, and included perspectives on modifying MDRs and adopting new approaches to meet taxa diversity requirements.

4. SYNDICATE SESSIONS

4.1 Syndicate Session 1: Ecological Considerations

Group 1A

Moderator: L. Maltby
Rapporteur: M. Hamer
T. Aldenberg
T. Barber
P. Craig
P. de Vries
C. Feng
G. Kon Kam King
K. Leung
A. Peters

The following questions / concerns were discussed:

1. *Are we making ecologically relevant assessments? Are regulatory protection goals explicit and clear? Are they set in relation to environmental quality? How do prospective and retrospective approaches differ?*

It was considered that the correct place to start was with the protection goals. Within the legislation, there are generally no explicit statements of what we are trying to protect. Rather broad statements, such as "the need to protect ecosystem health", or "have no unacceptable effects" are used. In addition, there is no such thing as 'the' ecosystem. Therefore, it is necessary to make these protection goals more specific. The use of ecosystem services allows different protection goals to be set based on a cost/benefit basis rather than just protecting everything everywhere all of the time. There is the possibility to set protection goals based on ecosystem structure and/or function, with the assumption that by protecting ecosystem structure, function and other ecosystem services would be protected.

It was apparent that there were different protection goals between different legislative frameworks. Examples of goals included:

- preservation of ecosystem structure and function based on measured endpoints;
- no to minimal impacts accepted (Sijm *et al*, 2002)
- acceptability of some effects (EC, 2009)
- the need to "protect aquatic life", through the ecosystem structure, with essentially no effects acceptable (EC, 2000).

Examples were given of protection goals, mostly based on population, community or ecosystem structure/function, acceptable/no effects, with little quantification of the scale of effect. However for the US Clean Water Act, the stated protection goal is 95% of all species and this is achieved through setting a quality standard based on an HC5 from an SSD.

There was recognition of the limitations of the ecological relevance of SSDs. They are a collection of single species laboratory tests (or extrapolations), limited to direct effects, without interspecies interactions and so do not represent communities or ecosystems. Despite the fact that SSD-predictions are limited to predicting direct effects, indirect effects are more likely to be associated with concentrations to the right of the SSD curve (HC50+), since indirect effects cannot occur when direct effects are absent – so that indirect effects are unlikely at an HC5-level that is really protective. So do these limitations matter, provided there is empirical evidence that the risk assessment procedure is effective and ecosystems are protected? Certainly, the relatively limited available evidence does tend to suggest that communities are adequately protected by thresholds derived from SSDs compiled from single species laboratory tests.

Prospective and retrospective approaches to risk assessment and the application of SSDs to achieve their stated aims are very different. The aim of using an SSD in a prospective risk assessment is to derive a threshold which is protective, whilst SSDs in retrospective analysis are used in a predictive way, to assess the likely degree of impact on local communities. For prospective risk assessment a single parameter such as an HC5 may be all that is required from an SSD, whereas other characteristics of the SSD will be important for retrospective analysis, such as the relative sensitivity of species, and the slope.

2. *Are all species of equal importance, or are there keystone species that are more important than others?*

Participants at the workshop were in general agreement that all species are not of equal importance, there can be differences ecologically, economically or aesthetically, for instance. This question brings the discussion back to the protection goals, and deciding what it is we are trying to protect. If the protection goal is to preserve ecosystem structure, i.e. no effects, then it does not matter where the species are on the SSD. If using an SSD as a diagnostic tool, site specific protection goals (and testing of particular species) may be appropriate.

Protection goals may impact the species tested for an SSD, such as a site specific assessment, but there are normally other considerations such as availability, ability to culture, endangered species concerns or ethical considerations (vertebrates). Furthermore, not all tests are performed in light of protection goals. All of this can create a bias in the species that are being tested.

3. *Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness or should we develop representative assemblages/communities (archetypes) to represent different typologies? Should protection goals account for local community composition?*

4. *How does aquatic community sensitivity vary with species composition?*

The above 2 questions were tackled together. There was much discussion over the need to consider different communities taking into account geographical and climatic factors. Limited evidence exists to support whether or not these factors can be deemed important for the majority of organic chemicals, and evidence is therefore insufficient to draw general conclusions. For metals it is known that differences in water chemistry can affect not only community sensitivity, but also the relative sensitivity of different organisms within an SSD. Similarly for freshwater and marine communities, whilst acknowledging that these communities can be very different structurally, aside from chemical considerations affecting toxicity, there is little evidence of differences between the inherent sensitivity of the communities. When using SSDs in a prospective manner with no effects as the protection goal a single threshold value derived from a generic SSD is all that is needed, provided it is sufficiently protective. Specific community composition may be important when some effects are allowed as both direct and indirect effects may occur and recovery is included. Again, when using SSDs retrospectively in a site specific assessment the need to understand particular assemblages may be important. Approaches which are able to predict community composition in the absence of pressures can provide a valuable tool in assessing whether or not especially sensitive organisms would be likely to be present.

5. How can knowledge of chemical MoA help construct SSDs for HC5 estimation?

When chemicals have a non-specific, or unknown, mode of action, which species are tested should not make a difference when using SSDs in a predictive, protective manner. In these cases including a diverse taxonomic range of relevant organisms is likely to be the most appropriate approach. However when there is a specific mode of action, this needs to be taken into consideration when selecting species for testing and combining in an SSD.

6. What are the research needs?

Further validation of thresholds derived using SSD based approaches to adequately protect communities and ecosystems from adverse responses to predicted chemical exposures.

Group 1B

Moderator: L. Posthuma
Rapporteur: P. Whitehouse
S. Belanger
C. Collin-Hansen
C. Eadsforth
M. Galay Burgos
J-P. Gosling
M. Junghans
M. Warne
R. Wenning

The following questions / concerns were discussed:

1. *Are we making ecologically relevant assessments? Are regulatory protection goals explicit and clear? Are they set in relation to environmental quality? How do prospective and retrospective approaches differ?*

2. *Are all species of equal importance, or are there keystone species that are more important than others?*

3. *Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness or should we develop representative assemblages/communities (archetypes) to represent different typologies? Should protection goals account for local community composition?*

4. *How does aquatic community sensitivity vary with species composition?*

5. *How can knowledge of chemical MoA help construct SSDs for HC5 estimation?*

6. *What are the research needs?*

Due to emerging views, the discussion followed the questions only loosely, instead taking a 'high level' view of the issues being raised.

A theme of the group's discussions was that there was no such thing as 'the' ecosystem and consequently no single benchmark that would apply in all circumstances, in the sense of predicting accurately between 'safe' and 'impacted' conditions. It is reasonable to expect that assemblages that may be exposed to stressors like chemicals will vary in composition from place to place and over time. Examples of the more extreme ecosystem types might be communities associated with polar regions (characterised by low diversity but large numbers of organisms) or sub-tropical seas (where biodiversity is typically very high). It is also reasonable to suppose that these different ecosystem types will show a range of sensitivities to stressors, and while some may exhibit high sensitivity to a particular set of stressors (e.g. eutrophication), others may be more sensitive to the presence of other stressors (e.g. toxic chemicals).

Often, there is a requirement to set a generic criterion - one that is intended to apply across a large area (perhaps even a continent) - and is independent of environmental variables. Ideally, the most sensitive ecosystem is protected by the generic criterion. Logically, when this protects the most sensitive of a suite of systems (ecotypes) then other systems which are less sensitive would be protected too. That would imply a margin of safety when such a criterion is applied to less sensitive systems. In order to identify risk management measures, we need to acknowledge the variability between ecosystems, and try to understand the normal operating range of traits/species abundance in these different ecosystem types. This is very important to ensure that risk managers do not jump to the wrong conclusion and take action to address a less important pressure (false positive), or fail to fix a problem that really does need attention (false negative). For making more ecologically relevant assessments, we could envisage a distribution of ecosystem sensitivities. The group speculated whether it might be possible to place such ecosystem types on a scale of sensitivity to a particular stressor, and to extrapolate from ones where we have a good understanding of chemical sensitivity to ones where we do not.

Taking stock of current approaches, the group felt that most HC5s seem to be accurate (based on limited corroboration from field and mesocosm studies) in the sense that HC5 values often correspond to the absence of visible or measurable biodiversity changes in field or mesocosm studies. But is that merely a coincidence? We can see no underlying reason why ecosystems should 'tolerate' risks (NOEC-exceedance) to a maximum of 5% of species without measurable adverse changes to structure or function. As presented by Posthuma (section 3.1), there is no ecology (yet) in SSDs, nor in the definition of the 'safe' level. On the other hand, requirements of SSDs to capture a diverse spectrum taxonomic representation does suggest that practitioners broadly view ecological and taxonomic considerations as relevant and important.

The group also agreed that functional aspects of ecosystems are more resilient than structural aspects (e.g. primary production could go unaffected even if several species of algae were impacted by a stressor). This implies that a focus on protection against impacts on ecosystem structure would generally protect major ecosystem functional aspects too. However, it must be recognised that some species have specific value, because they provide important ecological services, they are charismatic, or are rare. However, we have little appreciation of a relationship between scarcity and sensitivity to stressors (are they rare because they are sensitive?). We seek to protect structure, and thereby function, in a generic (non-ecological) way and, based on the corroboration from field evidence, generally seem to succeed. However, we should be careful that the proportion of species at risk does not lead to functional ecosystem change e.g. if pollinators were in the affected fraction of species. Scenarios that look at specific receptors need be systematically considered but this is much more challenging when the aim is to develop a generic criterion, as opposed to a well-defined ecosystem or habitat .

In the current regulatory paradigm, the (largely unknown) variability between different ecotypes is dealt with by focussing on structural protection in our thresholds. We have been seeking to protect all but the most sensitive 5% of species by setting an HC5-NOEC based threshold criterion, based on clean water laboratory studies and most sensitive endpoints, in the expectation that such a threshold will provide adequate protection to a wide range of communities. With hindsight, where SSDs are used to estimate a threshold for a single stressor, the approach appears to be protective, but the approach may be too simplistic: we usually rely on information about the sensitivity of a sample of closely related individuals of a species to the stressor over a fixed period of exposure. We are not making effective use of the wider insights

that are now available about (1) ecological, (2) chemical, (3) exposure, or (4) toxicological influences on risk. The following examples illustrate this over-simplicity.

(1) Ecological aspects: this includes an understanding of interdependencies between species through food webs (system level), or the relationship between sensitivity and traits such as reproductive strategy and feeding behaviour (species level). Instead of describing communities of organisms in terms of the species they contain, they could be described in terms of the traits they exhibit, or their dependencies on each other. Some aspects of food web architecture may be common to many ecosystem types, so this is where our effort should focus initially. Past studies for example have already looked into QSSRs (Quantitative Species Sensitivity Relationships), where not only chemical sensitivity but also relationships with body size and so on were studied. Trait-related studies are on-going in this respect.

    It might be expected that the responses of organisms in extreme environments (e.g. highly saline) or where there is adaptation to stressors e.g. in metal mine tailings) are linked to the physiological adaptations needed for those environments.

(2) Chemical behaviour: There is now a much better understanding of the importance of water chemistry on bioavailability of metals and hence their toxicity to aquatic organisms. Bioavailability is now explicitly incorporated into SSDs for some metals. The influence of pH on weak acids could readily be accommodated in a similar way, giving rise to more environmentally relevant estimates of risk. The effect should be to lead to more accurate assessments of risk so that any remediation is directed to where it is really needed.

(3) Chemical exposure: The exposure profiles of many substances are known to affect the response of organisms, and these can be characterised. For example, many household chemicals typically give rise to low-level chronic exposure from point sources, whilst some insecticides applied to arable and tree crops are more likely to give rise to short episodes of exposure such as following accidental overspray or run-off shortly after application. Sessile organisms in estuaries are likely to experience diurnal variations in exposure to chemicals with tidal ebb and flow.

(4) Toxic mode of action: Information about a chemical's mechanism of action is important in helping to identify taxa (or perhaps traits) that are likely to be particularly sensitive. Existing guidance in the EU Common Implementation Strategy 'CIS' Technical Guidance on EQS Derivation (EC, 2011) acknowledges this and suggests that information about mode of action may be used to adjust assessment factors. Alternatively, an understanding of mode of action can help focus attention on critical data gaps that may be filled by testing particular species, or some of the non-testing alternatives suggested below.

How might such higher level thinking be incorporated into our hazard and risk assessments? One approach may be to use existing and emerging tools for generating information about sensitivity of different species to chemicals (e.g. QSARs, 'read across' tools, Web-ICE, dynamic energy budget tools [DEBTOX], to simulate different ecosystem types, and the consequences for thresholds if we were to simulate exposure of a community dominated by certain taxa or trophic level e.g. primary producers. That is, there is a need to distinguish between the protective success of our conventional generic methods (deriving one criterion to

cover all eventualities), and the need to tailor the generic approach to specific circumstances when this is required (e.g. site-specific thresholds).

In a tiered system, the generic criterion serves as a starting point, which protects all systems. This is intended to be protective, but provides different margins of safety to different ecosystems. At a higher tier of assessment, SSDs for specific scenario (a specific area, species composition, or community exposed to with other stresses, etc.) may utilise information drawn from the 4 fields mentioned above. The aim would be to refine the SSD output to the system of interest. As a strategy, the group felt that there is ample opportunity to create 'what if' scenarios with SSDs, so the assessor can decide whether the criterion is, or is not, sensitive to adding scientific insights and data. By incorporating some of the principles mentioned in (1) – (4), it may be possible to simulate the effects of, for example, changing the set of tested species data in the SSDs to mirror the ecosystem under study. Low impacts of such simulations would suggest that the criterion is robust, whilst high impacts suggest specific attention for the factors causing the change, which may feed into risk management activities, or help guide the generation of new data.

The more intensive approach to hazard and risk assessment we are advocating could add a lot more work. How do we know when enough is enough? How can we decide when a next tier is necessary, and when we should stop? The group felt that the alternative scenario outputs, and the confidence estimates around an HCx provide a useful prompt for more (or less) scrutiny. A large uncertainty should be a major trigger for the sorts of systems thinking advocated above. Field data may also have a stronger role to play as a line of evidence in defining and using thresholds. The EU Technical Guidance on EQS derivation (EC, 2011) already refers to the use of field data in informing the size of assessment factors to be used.

In summary, the group felt the time was right to move away from using SSDs as a purely statistical construct applied to poorly understood species sensitivity data to one in which SSDs provide the framework for a more process-based approach. We envisage a statistics-related approach will remain at the core of our assessments but it can be enriched with fundamental insights of the kinds stated in (1) – (4). This would be greatly assisted by some sort of over-arching guidance (not a 'rule book') that would prompt the assessor to think about some of the factors that might be important (1-4 above) and the options for pursuing them further. The guidance might usefully adopt a hierarchical structure (as suggested above) that takes the assessor through a series of 'things to think about' in some logical sequence. The aim should be to stimulate a broader approach to the assessment and not to 'fossilise' the science by being over-prescriptive or wedded to particular tools. Any over-arching guidance should, however, be clear about what tools are available, along with their strengths and limitations.

Group 1C

Moderator: S. Duquesne
Rapporteur: S. Dyer
M. Barron
J-L. Dorne
A. Gosselin
M. Habekost
J. Hendriks
C. Michel
A. Ragas
F. Wu

The following questions / concerns were discussed:

1. *Are all species of equal importance, or are there keystone species that are more important than others?*

There was a clear consensus that not all species should be considered as equal to each other within an ecosystem. Species sensitivities to toxicants and other stressors clearly illustrate that there are sensitive taxa as well as tolerant taxa. A commonly held ideal is that in ecosystems there are keystone species. However, the concept of a keystone species is context dependent. For example: 1) a keystone species may refer to exquisite sensitivity to toxicants and/or other stressors; 2) a rare species, worth protecting via regulations (e.g. nature conservation); 3) a species that provides a valued service (e.g. salmon – human food); and 4) species that are so interdependent upon each other (e.g. snail kite and the apple snail, found in the Everglades, FL and in South America) that the larger ecosystem is dependent upon their continued survival. The role of a keystone species may be most impactful in simple food webs as compared to highly robust and redundant ecosystems. An example discussed within Syndicate 1C was that the elimination or reduced abundance of one taxa in an arctic ecosystem, where food webs are short, would be highly impactful to the entire system compared to highly complicated webs.

2. *Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness or should we develop representative assemblages/communities (archetypes) to represent different typologies? Should protection goals account for local community composition?*

There was a consensus that derivation of a PNEC should follow a tiered process as is commonly done in current risk assessments. Protection goals should be based upon what is to be protected and how it is related to environmental exposures. For some exposure scenarios, the generation species sensitivity distributions may not be necessary to generate species sensitivity distributions as the use of generic safety factors may be sufficient. For example, a chemical that is not toxic and used at low levels may not need a sophisticated assessment. However, with the availability of Interspecies Correlations Estimations (ICE) it is possible to estimate screening level HC5s based on data from a few laboratory tested species. The species present in the screening level SSDs provide a suite of traits worthy of protection. The development of traits-based SSDs and HC5s are considered as a generic manner of addressing different types of ecological sensitivities. However as SSDs are usually based on an array of single species toxicity test results, the

ecological sensitivity will thus not include the variations due to indirect effects such as community effects (e.g. competition for resources, predation).

3. How does aquatic community sensitivity vary with species composition?

Detecting ecological sensitivity to chemical and/or physical stressors is dependent upon what are considered 'reference conditions'. In general, reference conditions refer to ecological states which humans have not significantly altered. Detecting differences from the reference condition may be dependent upon how specious the ecological community is. For instance, highly diverse and specious ecological states may have traits and functions that are redundant among some taxa, hence – resiliency. Losing taxa in such situations may be more difficult to ascertain than say reference conditions in which there are few species (e.g. arctic systems or nutrient poor situations). In more simple systems the loss of taxa may have large ecological consequences.

There is a great need to provide a mechanistic argument for the use of traits. That is, there is a need to illustrate the links between biochemical responses to stressors and morphological, physiological and ecological consequences. If SSDs are to be accepted by the ecologically-scientifically minded community, links of species traits to community – level traits will be needed. A simple example includes r versus K (quantity versus quality of offspring) reproduction strategies of different species and how they lead to ecological community resilience to stress. Currently, SSDs do not provide such information, but should in the near future.

4. How can knowledge of chemical MoA help construct SSDs for HC5 estimation?

A chemical's mode of action (MoA) and its potential effect in the construction of SSDs needs to take into account its intended as well as non-intended effects. MoAs are typically defined in terms of acute toxicity. As such, there are a limited number of MoAs. Further, they are often develop for and thus tied to specific taxa. For instance, insecticides are obviously more potent to arthropods than plants, however, fish may also be sensitive. The advent of adverse outcome pathways are an excellent tool to illustrate the probable causal relationships between the dose of a chemical at a subcellular level to the series of events leading to impairment of a population. However, chronic exposures may greatly change the potential for adverse ecological effects. Long-term exposures may affect more metabolic life stages of diverse taxa and thus provide a different array of species sensitivities. Traits such as those related to r or K reproductive strategies, body size, and accessibility to direct exposure to the chemical may be very different depending on taxa.

Tools, such as QSARs, 'Omics, physiologically based pharmacokinetic (PBPK) and energy-based models are very useful in classifying the potential effects of a chemical's MoA. Some caution needs to be exercised here, however, as for most taxa, such models do not exist. Indeed, reasonable data sets, that can tie a chemical's MoA with biological effects exist for only a few taxa (e.g. zebrafish).

5. What are the research needs?

There are 2 key areas in which more research is needed regarding the use and generation of SSDs for environmental protection: 1) inclusion of adverse outcome pathways (AOPs) or mechanisms of toxicity and 2) linking mechanisms to species traits. Such research would provide knowledge on mechanistically-based outcome pathways and on determining taxa possessing these pathways. Establishing links between these aspects would eventually show a continuum between exposure and the propagation of effects. Based on such information, a SSD could then become more ecologically informative.

Group 1D

Moderator: K. Solomon
Rapporteur: S. Marshall
S. Andres
P. Chapman
D. de Zwart
A. Hart
A. Macken
Y. Pan
S. Raimondo
H. Sanderson
Z. Yan

The following questions / concerns were discussed:

1. *Are we making ecologically relevant assessments? Are regulatory protection goals explicit and clear? Are they set in relation to environmental quality? How do prospective and retrospective approaches differ?*

The group did not consider this question.

2. *Are all species of equal importance, or are there keystone species that are more important than others?*

This depends on the ecological role of the species and whether this needs to be explicitly protected (is there functional redundancy?). Most keystone species are known and can be protected appropriately.

Some keystone species are actually tropho-species, i.e. assemblages of species with an ecological role, e.g. Krill in polar marine environments.

Keystone species should be protected and, therefore, ideally would also be included in the SSD. However, the likelihood of having test data or being able to generate it is dependent on other, more practical factors.

Charismatic species often do not have ecological importance (compared to keystone species) but can be important to human society, e.g. pandas.

3. *Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness or should we develop representative assemblages/communities (archetypes) to represent different typologies? Should protection goals account for local community composition?*

It seems likely that the species data available to fit an SSD for prospective estimation of an HC5 would be unrepresentative because they tend to be those species conducive to laboratory testing regimes and methods. However, this does not mean that the assumed statistical distribution of the SSD would be

inappropriate. The group did not decide whether representative assemblages would make a substantial difference to the generic approach. However, the group did consider that local community composition should be considered in setting protection goals. Setting protection goals from policies informed by science is key to good management of chemicals. Once protection goals are set, scientists can design risk assessment strategies as necessary. This can benefit from use of SSDs.

The group noted that toxicity data generated from studies with laboratory cultured organisms may not be representative of toxicity in field organisms where the field organisms have developed a degree of tolerance to some chemicals, e.g. metals. This raises the question of whether laboratory tests should involve organisms pre-exposed to the test chemical. Of course, for chemicals not yet present in the environment, this would be inappropriate.

4. How does aquatic community sensitivity vary with species composition?

The work of Professor Maltby (section 3.2) suggests that differences in community structure do not relate to clear trends in sensitivity. Although there may be data gaps as in the case of chronic data for EPT (the Ephemeroptera–Plecoptera–Trichoptera taxa, considered as relatively sensitive taxa), does comparison of PNECs derived from HC5s with field biomonitoring data suggest such taxa are protected? It follows that there would be little value in developing generic community scenarios/archetypes to represent different community sensitivities.

At an organism-sensitivity level, taxonomic distance starts to be important at higher levels of organisation, i.e. order or above. This implies that it is more important to include a range of broadly different taxonomic species than to add more species that are taxonomically close to species already included in the distribution.

Differences in sensitivity between freshwater and saltwater organisms can vary more than between different freshwater communities, but this is likely to be due to differences in bioavailability/chemistry.

There may be specialised organisms with specific adaptations to local conditions that could lead to higher or lower sensitivity. For example, cold water adapted species, e.g. arctic cod, may adopt different excretion routes for some chemicals (via liver rather than kidney). If exposed to a pollutant that is metabolised in the liver, this could influence toxico-kinetics and, therefore, sensitivity compared to other fish. Of course, other factors such as slower rates of degradation will influence exposures and also must be taken in consideration when undertaking risk assessments in cold water marine environments.

5. How can knowledge of chemical MoA help construct SSDs for HC5 estimation?

The slope of an SSD, i.e. its cumulative distribution function, is indicative of a chemical's MoA. This seems to work for distinguishing narcotics, where the slope is usually uni-modal and steeper, from chemicals with specific MoAs, where the distribution is usually flatter and multi-modal. Comparison of the slope of a chemical SSD with unknown MoA with SSDs for existing chemicals (with known MoAs) can also be useful.

For pesticides, which often have wide-ranging potencies to different taxonomic groups, the SSD should be constructed from the more sensitive taxa when this can be demonstrated or separate SSDs constructed for sensitive and less-sensitive tax.

Care is needed to ensure responses used in the SSD are comparable - responses for plants tend to be measures of growth inhibition whereas invertebrate tests measure binary endpoints such as mortality. These should not be mixed in the same SSD although they often are. If there are enough data to make a SSD for fish, invertebrates, and algae, it is useful to consider if the different taxa data overlap to decide if they should be combined to represent the 'ecosystem'. Colour coding species in the SSD plot helps visualise sensitive groups.

US EPA may also look at the most sensitive 4 or 5 species in the SSD to estimate a conservative HC5. This clearly provides a strong conservative bias although it should be noted that the statistical endpoint, HC5, is influenced by the span of the sensitivity distribution. Using SSDs to set WQSs can be applied differently in different regulatory jurisdictions.

Knowledge of the mode or mechanism of action is useful to decide if taxa should be in one SSD for one MoA. MoA needs to be described at an appropriate level. The group thought mode of action was more practical than mechanism of action.

Knowledge of MoA could help decide which species to test, e.g. herbicides and algae/plants. However, this may be a problem when guidelines for test methods are not available for some taxonomic groups, e.g. molluscs.

6. What are the research needs?

Sensitivity of cold water communities, e.g. arctic.

Better application of toxicological data in SSDs, e.g. using more chronic data, mechanistic understanding.

More use of predictive modelling to overcome limited data sets.

4.2 Syndicate Session 2: Statistical Considerations

Group 2A

Moderator: K. Leung
Rapporteur: P. Craig
T. Aldenberg
T. Barber
P. de Vries
C. Feng
M. Hamer
G. Kon Kam King
L. Maltby
A. Peters

The following questions / concerns were discussed:

1. *Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.*

The group considered a number of tools in turn, trying to evaluate the strengths and weaknesses of each. The tools considered were Web-ICE (a web-based interspecies correlation estimation tool), hSSD (a hierarchical SSD modelling tool), ETX 2.0 (a program for calculating estimates of the HC5 and potential fraction affected and confidence intervals for the estimates, based on a log-normal SSD model), MOSAIC_SSD (a web interface which calculates an estimate of the HCx for any x and provides a bootstrap confidence interval for the estimate) and the R software developed by Procter & Gamble and presented by Scott Belanger in this workshop (section 3.5). The group were aware of the existence of some other tools (e.g. BurrliOZ software used in Australia) but there was insufficient time to evaluate them.

The hierarchical SSD model (hSSD) was considered to be an experimental tool and still under development, whereas the other tools were available for use to construct SSDs and compute HCx estimates and confidence intervals.

The tools divided naturally into 2 groups:

Web-ICE and hSSD

Features in common: Both tools make use of taxonomic structure and have the potential to address data gaps by predicting toxicity for species which have not been tested. This was felt to be generally useful but especially in the context of reduced animal testing. It was noted that Web-ICE has a specific feature for prediction of the chemical sensitivity of endangered species based on simple regression models. Both tools were built using only data from acute tests and this was felt to be a significant limitation, especially in regulatory settings requiring chronic data. Both tools made no distinction between classes of chemical and it was felt that mode of toxic action is an important consideration which should be taken into account.

Contrasting features: The Web-ICE user can select or reject individual toxicity predictions whereas the hSSD user cannot. The hSSD programme accepts censored toxicity data whereas Web-ICE does not.

ETX, MOSAIC_SSD and R-SB

Features in common: None of these 3 tools has the ability to address data gaps although they can be used to indicate the influence of adding additional data points to the HC5. All of these tools would therefore require a toxicity dataset which is regarded as complete in terms of regulatory requirements for taxonomic representation, but the dataset might be supplemented with values predicted with/extrapolated from other models. However, all of these tools could be used with either acute or chronic data.

Contrasting features: Mosaic and R-SB both accept censored data whereas ETX does not. ETX is restricted to log-normal models whereas the others can fit other parametric distributions, in particular the log-logistic family. Notably, both MOSAIC_SSD and R-SB provide a feature to compute bootstrapped approximate confidence intervals for the HCx value.

In addition to this general division into 2 groups, some specific features were identified:

Web-ICE:

Some of the members in the syndicate group felt that the capacity to reduce data gaps would lead to reduced uncertainty attached to the resulting HC5 estimate while others felt that it was not clear that it would do so, as known uncertainty caused by a small sample size is replaced by uncertainty originating from modelled/extrapolated toxicity values, which may not be completely quantifiable. Members also considered that the method by which Web-ICE computes a confidence interval for the HC5 was not transparent and clear enough; some members questioned the validity of the method.

It was felt to be ecologically biased, in particular towards fish, in terms of the pairs of species for which a sufficient dataset was available to exploit the interspecies correlation. It was noted also that it is built mainly using species primarily from North America. Since the method is simply based on a simple linear regression model using toxicity endpoint data from the well-studied standard test species and those from a group of rarely tested species for various chemicals, there are foreseeable uncertainties due to the limited data, highly variable combination of chemicals and unknown mode of toxic actions of the chemicals being included in the regression.

In terms of validation of individual predictions, it was felt by some of our members that, although quality of prediction had been assessed by a cross-validation approach, this was still essentially a validation internal to the data used rather than a validation against data from a truly external source.

hSSD:

The hSSD model is highly parameterised; some of our members felt that this was beneficial, but others worried that it was over-parameterised.

The hSSD requires the user to specify an ecological scenario. This was felt to have some potential benefits, especially in terms of ability to test hypotheses about effect of the chemical on the site-specific community structure.

Up to now, the hSSD has not been validated against laboratory and/or field data.

ETX:

The ETX tool has regulatory acceptance in Europe since it is easy to use and understand.

At the end of the detailed discussion on the pros and cons of the various tools, there was a short discussion on their implications. It was agreed that this remains an area in which progress is on-going and it is not clear which tool is the most useful nor is it clear whether a single approach should be used. It was suggested that the way in which any tool is used may be just as important as the choice of the tool.

2. *As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?*

The group devoted limited time to this question.

It was agreed that knowledge of mode of toxic action should be taken into account first. Resulting predictable differences between taxonomic groups are a key driver in ecological risk assessment.

Overall it was felt that the best choice of methodology and software tool will depend on the question to be answered. Consequently, it is very important for risk managers to specify clear protection goals and questions to be answered by the risk assessment. For example, one should be clear about purpose of the HC5; is it intended to protect 95% of species or 95% of ecosystems?

The question of the need to continue with testing of taxonomically diverse species was raised. A distinction was made between prospective and retrospective assessments and between biocides and other chemicals. For prospective assessments, it was suggested that for biocides, one may often know the mode of toxic action and therefore be able to target testing appropriately (e.g. insects and arthropods should be emphasised in SSDs for insecticides; plants and algae should be the focus of SSDs for herbicides) whereas this would probably not be so for an industrial chemical without clear *a priori* information on its mode of toxic action. For retrospective assessments, there appeared to be a greater need for taxonomically diverse data, especially for general biocides.

3. *What are the research needs?*

The group's discussion was limited to a brief list of topics:

- Further validation for extrapolations in relevant models (i.e. hSSD and Web-ICE) and of consequences for HC5 uncertainty.

- Extending software tools to add the capacity to predict chronic toxicity.
- Validation of hSSD scenario-specific HC5s relative to the field and/or mesocosm studies.
- Uses of SSDs for purposes other than estimating the HC5 (e.g. using the entire SSD for probabilistic risk assessment and deriving other values (say HC50) for trigger management action).
- The role of SSDs in risk assessment for mixtures is still at an exploratory stage. It is likely to be dependent on the questions to be addressed and the rationale for using SSDs.
- Possible inclusion of microorganisms in SSDs to protect ecosystem functions was debated. For instance, when assessing the ecological risk of fungicides, we never consider including various fungal species in the test battery and incorporating their data into the SSD; if our management goal is to protect the ecosystem functions and services, we should try to protect the fungi as well. As such, microorganisms should be considered in the HCx derivation. Nonetheless, such a development is currently hindered by the lack of available approved testing procedures for different groups of microorganisms.

Group 2B

Moderator: R. Wenning
Rapporteur: J-P. Gosling
S. Belanger
C. Eadsforth
M. Galay Burgos
M. Junghans
L. Posthuma
M. Warne
P. Whitehouse

The group was asked to consider the following set of issues covering statistical aspects of SSD use in the order presented (order not according to importance):

1. *Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.*
2. *As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?*
3. *Do models based on prior knowledge provide advantages over other methods?*
4. *Are current modelling success criteria, such as those identified in the REACH TGD, sufficient, overly prescriptive or insufficient?*
5. *What are the research needs?*

Although the group was guided by considering these issues, the discussion was not structured using the questions, and, ultimately, the group decided to report back about 3 main themes. Those themes yielded over-arching notions, valuable to answer the detailed questions and the application of SSDs and criteria. Building on the presentations and discussions of the previous workshop day regarding ecological aspects and validity, for example, it was made clear and discussed that various ecological aspects are currently being addressed in SSD-related risk assessments. Further, SSDs have various uses beyond criteria setting.

Initial Reactions and Thoughts from the Syndicate

At the start of the discussions, there was an opportunity for each member of group to highlight what they thought was the key issue surrounding the statistical models being discussed at the workshop. Here is the list of comments and questions that individuals thought were important for us to discuss:

- What are the limitations of the statistical models? We need more experience with the models to understand the limitations and the differences. The influence of several data aspects on different models are uncertain. Factors include, but are not limited to, amount of data, confounding factors, and sensitivity to natural variability.

- What are the possibilities for using the models for extrapolation outside the existing data? More experience is needed in understanding how chemicals and aquatic organisms differ in different climatic regimes and aquatic environments. It is uncertain how biological functions might differ between species and how structure might be different in various food webs.
- What extra ecological knowledge needs to be included in the models to add further value? The connection between mathematical models and ecology is tenuous. While the goal is to mimic biological response mathematically, it is important to understand the influential biological and ecological factors that need to be measured and accounted for in deriving models tailored to the problem. There are pathways to address ecological information in SSDs, like via hSSD, tailoring the model to a site or water body or system. Questions that arise then include when and how to do that and how to handle spatial differentiation in modelling results (here: criteria)? Would tiering apply to this (i.e. a generic model for 'the' ecosystem and tailored models for 'this' ecosystem?). Improving the connection between statistical models and ecological relevance will help to reduce uncertainty and elevate confidence in the protectiveness of threshold values derived from the models.
- Is the use of SSDs curves the best approach; are there other approaches? Tailoring and tiering also applies here. The regulatory community has prescribed the statistical methods behind SSD curves, and that appears to be one of the main reasons why SSD curves are being used, next to versatility. It would be useful to understand if sufficient effort has been invested to explore other options that provide equal or better confidence in the threshold values derived using the SSD approach. It would also add confidence to the current approach and perhaps point in new directions for improvements.
- What biological, chemical, and ecological aspects are important and, perhaps, unaddressed in the current SSD approach? We should not throw away what is known in terms of mode of action, species differences, chemical interactions in mixtures, exposure scenarios, food webs etc. We do not understand well how these aspects translate into a statistical model, though hSSD is an example of incorporating novel data and concepts within the classical 'flat' SSD modelling which is based on ecotoxicological test data and model choice.
- Are taxonomic similarities and differences important when extrapolating between species or adopting only available biota-response data (as opposed to a prescribed mix of different taxonomy or species)? There seems to be strong evidence that taxonomic closeness implies similar toxicological responses. It would be worth investigating similar relationships for mode of action or exposure scenarios. For mode of action, there expectedly is a stronger numerical effect of specific modes of action on SSD-modelling outcomes then there may be of narcotic action.
- Has sufficient attention been given to whether SSD curves answer the important questions that underlie this work? There should be careful consideration of the modelling assumptions. In particular, regulators and scientists should ask themselves what biota or ecology are they trying to protect?
- Does the use of SSD curves replace or compliment risk assessment? Using current models seems to be tinkering on the edge of what we expect from risk assessment. Are we trying to too hard to re-package risk assessment and its elements of exposure, dose-response, ecotoxicity and uncertainty into a different statistical method?

In conclusion, the statistical questions posed are of actual relevance, but gain perspective when considered in relation to the wider perspective of scientific developments on e.g. mode of action, ecological aspects linked to SSDs, and so forth. The group summarised the above discussions and perspectives into 3 main themes: (1) *thoughtful decision processes*, (2) *inclusivity with regards to available data,* and (3) *interpretation of the uncertainty estimates from different models* and took a step back from the detailed questions to consider the context for applying SSD curves, the information used to populate an SSD curve, and the interpretation of results generated by the model.

(1) Thoughtful decision processes

A decision tree (or other formal decision process) could be used to help structure the prior thinking and the subsequent steps in the hazard or risk assessment. Risk assessments do have very different problem definitions, and do imply different data availabilities, while asking for either generic or specific answers. This context suggests a formal decision process. SSDs could then be contained within such a defined process, and their role, which may be small, will be appropriate for the risk assessment problem in hand. Given the emergence of new chemicals every day, a thoughtful decision process can help to anticipate key chemical and ecological characteristics that may be more or less important relative to the evaluation of other chemicals in the same or similar aquatic systems.

It was noted, for example, that the REACH Technical Guidance Document (ECHA, 2011) is probably the most well documented example of how a decision process can guide the framework for conducting a risk assessment that yields proposals for criteria (PNECs). The same approach to a decision process might be useful to guide the development of SSD curves, as well, and other similar tools useful to understanding biological/ecological responses to chemicals in the environment. In the context of SSD curves and similar models, the decision process could be extended further to incorporate recommended approaches for handling old data, new data and different types of knowledge. A decision process will help risk assessors to plan, implement and evaluate how to apply models and decide what data to use. In fact, decision processes can help us to understand the value of the model and data. A proper decision process yields approaches that are best tailored to the problem definitions that exist, and harbour (thus) contextual flexibility (which question is answered, which approach is chosen) as well as transparent consistency (given a chosen method, there is a clear way how to do it in that context).

Also, a well-documented decision process will add transparency which is needed to improve current practices. At present, the decision process used to select information for populating an SSD curve is known only to the extent that the developer has openly identified the assumptions used to judge the quality of available data and to select certain studies or aquatic species and not others.

As part of developing useful decision trees or data evaluation processes, regulators and scientists need to ask the following questions: what are the problem definitions for which the SSD model is applied, where are current processes recorded and do individual organisations have different assessment procedures when using SSD results. For example, RIVM refers to technical guidance and certain rules for what procedure(s) to follow depending on the circumstances of the risk assessment. However, the danger of over-prescription should be avoided. Guidance and recommendations are preferable to hard do – and – don't rules. There is a worry that guidance can rapidly evolve to become rules, which can lead to fossilisation of statistical models and approaches, or major communication problems when the SSD model is used in a different context (e.g.

disaster management as opposed to a generic risk assessment for which the SSD was derived) when the rules set for criteria derivation are assumed valid. Therefore, any guidance that is produced must also look to future proofing.

Lastly, there is a concern that overly prescriptive models and data assessment methodologies might stop regulators and scientists from thinking about the assumptions of the models and their best use for individual risk assessments. Coupled with this is the concern that regulation might constrict the process because transparency and uncertainty are difficult aspects to accept. The key for any process (and associated statistical methodology) is that it must be fit-for-purpose. We do not wish to be regimented in our assessment approaches, given the context of various applications of risk assessment.

(2) Inclusivity with regards to available data

Undoubtedly, scientists and regulators wish to use as much data as possible in the risk assessment process, including the populating of SSD curves. A formal weighting process is one approach that can help to qualify all available data and prioritise the importance and value of data.

To achieve this, there is a need for a more formal data evaluation process. There are often several good reasons for excluding data from an analysis (e.g. chemical purity issues, exposure and biological issues, and poor reporting of the testing procedures). However, attention should be given to whether so-called discarded/rejected data might have some utility in the evaluation of model results. Early exploration of models using high- and low- quality data might be a logical first step as part of early data exploration, and before settling on a formal and final model and analysis. The early consideration of all available data offers the opportunity for insight on chemical and ecological attributes that might be missed when certain data sets are removed from the risk assessment. Care should be given though that the less strict approach to data evaluation for SSD derivation is not flawed by creating a bias in the SSD, as would be expected to occur, if for example, a very volatile substance is tested in open vessels and the test concentration is not verified.

The importance of a principled weight-of-evidence approach would address the final outcome. The practice of including and excluding data sets in a systematic manner to explore the influence of different data sets used to develop an SSD curve and perform a risk assessment could be particularly valuable in a weight-of-evidence scheme. Such a scheme should be included with the presentation of results of modelling and risk assessment.

Data are not available for all possible risk assessment scenarios. It was felt that improvement is needed in the both the quality and breadth (in terms of taxonomic diversity, mode of actions etc.) of data used in assessments. For instance, Web-ICE is currently based upon a data repository that is fit for US risk assessments. It would be useful to know if Web-ICE and similar approaches could be fit for risk assessments in other regulatory arenas.

(3) Interpretation of uncertainty estimates

It was evident to the group that the uncertainties associated with the HC5 results reported using each of the 3 primary statistical models discussed in the plenary meeting (i.e. ETX/R, Web-ICE and hSSD) are not the same in kind, representing different aspects of underlying variability, though numerically (partly) overlapping or (often) in the same order of magnitude. The different approaches to choice of data and data interpretation generate different types of uncertainties. The attributes of the uncertainties must be communicated properly.

For example, the confidence intervals reported using the ETX method stem from uncertainty in the fitted parameters of the underlying statistical distribution. The confidence intervals from the Web-ICE method attempt to capture uncertainty caused only by cross-species extrapolation. The credible intervals reported in the hSSD method stem from the characterisation of uncertainty about the underlying biological-response data and taxonomic differences. Each of these models is highlighting different model limitations.

This itself highlights the importance of data and model transparency when interpreting SSD curves and risk assessment models. There should be no blind application of statistical models (for instance, we should be concerned if the data underpinning the SSD is showing multi-modal behaviour and we are fitting a unimodal distribution due to habit or procedural prescription). Because uncertainty influences assessment factor specifications in some regulatory arenas, care must be taken in the interpretation of the uncertainty. Characterising and interpreting uncertainty correctly could influence the interpretation of SSD curves and risk assessments that encourage maximum insight from available data.

In addition to the uncertainty associated with the statistical modelling approach, there is a more general concern about ecological relevance and the interpretation of the models and the associated uncertainties. SSD curves as they are produced may have a fundamentally flawed misfit to the ecology and exposure conditions of the exposed ecosystem(s) of concern, and neither knowledge of the ecosystems nor the SSD-model itself may be flexible enough to capture that variability in nature. By their nature, SSDs are statistical models, which can only to a limited extent be expected to incorporate ecological information in them. They are, and will probably remain, lower-tier approaches in terms of addressing ecological. This issue raises questions about the predictive accuracy of statistical models based entirely or predominantly on data extrapolation. We also need to be clear about whether the results of SSD curves should be correctly reported in terms of the HC5, and whether the upper and lower bounds of uncertainty should also be reported and considered in regulatory decision-making.

Research questions

Throughout the discussion, the group identified research questions and paths for new or additional work that could help to improve modelling and risk assessment.

- What are the limitations of the models and are they fit for purpose?
- What are viable methods for incorporating all relevant data?
- Is it possible to treat mode of action in the statistical models in the same way taxonomic distance is being used? (In particular, is this feasible for Web-ICE and hSSD?)

- Can a formal decision tree approach that is inclusive of the available data and is transparent be defined?
- What additional ecological knowledge needs to be included to add value for the risk assessors?

Group 2C

Moderator: A. Ragas
Rapporteur: M. Barron
J-L. Dorne
S. Duquesne
S. Dyer
A. Gosselin
M. Habekost
J. Hendriks
C. Michel
F. Wu

The following questions / concerns were discussed:

1. *Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.*

Statistical aspects considered by the group included SSD-based extrapolation methods as well as alternative approaches. 2 SSD-based extrapolation tools were discussed: U.S. EPA's Web-ICE (http://www.epa.gov/ceampubl/fchain/webice/) and a recent tool developed by Peter Craig (Craig, 2013). Both tools were limited to acute toxicity data, incorporated taxonomic distance, and provided similar outputs including HC5 estimates using limited data. Web-ICE was considered to have more defined user rules, but was a less statistically rigorous SSD generator than hSSD. Also, the level of confidence with Web-ICE is lower when extrapolating over large taxonomic distance. Outputs of the hSSD tool were considered to be user dependent, which could result in substantially varying results between users. Of additional concern was that a high degree of ecological community expertise or knowledge was necessary to provide reliable estimates. Both tools were considered better alternatives to the use of generic safety factors. Both tools also require the availability of appropriate datasets, including standardised toxicity values and relevant exposure metrics (e.g. dissolved metals).

A variety of alternative approaches to SSD development were discussed, including trait-based SSDs, chemoinformatic methods (e.g. QSAR, read across), and determination of protective levels by just focusing on sensitive species. While trait-based SSDs were considered to have potential utility, questions on what traits should be considered (e.g. ecological, physiological, etc.) remained. It was unclear if there was sufficient knowledge for a sensitive species approach that would ensure protection of multiple aquatic systems. Overall, there was no consensus recommendation for clear alternatives to the SSD-based extrapolation methods above, and additional research would be needed.

2. *As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?*

There was general consensus that taxonomic closeness can be important in extrapolation of sensitivity across species. Species sensitivity may be considered highly correlated at the Family level. Understanding sensitive taxa, such as to Family level, could be used to ensure SSDs are representative of aquatic communities. Uncertainties remain regarding the need to alter SSD composition for different aquatic communities, including fresh versus saltwater species, large versus small assemblages, and sensitive versus robust systems. The proportion of invertebrates and fish in the SSD was noted as probably important in the estimation of HC5 of compounds that can have large differences in species sensitivity such as insecticides. The need to integrate water column species with other compartments such as sediment and terrestrial systems was also noted.

3. *Do models based on prior knowledge provide advantages over other methods?*

Prior knowledge can provide significant advantages to both constructing and interpreting SSDs. Important aspects include knowledge of MOA, taxon sensitivity, composition of the aquatic community being assessed, physiological and ecological species traits, and physico-chemical properties of the chemical. Relationships between taxon sensitivity and MOA are important because some chemicals will show large taxon specific differences in toxicity, such as herbicide sensitivity of plants versus fish, and acetylcholinesterase inhibitor toxicity to invertebrates versus fish. SSD development should consider this information, such as including plant species in a herbicide SSD and consider the proportion of invertebrates in constructing insecticide SSDs. Knowledge of chemical properties such as solubility are important in understanding maximum values to include in SSDs (i.e. should not exceed the solubility cut off).

4. *Are current modelling success criteria, such as those identified in the REACH TGD, sufficient, overly prescriptive or insufficient?*

The current REACH criteria for SSD composition includes 10 species of 8 taxonomic groups. Overall, these requirements seem reasonable and are consistent with the 8 family minimum data requirement in U.S. EPA guidelines for developing U.S. Ambient Water Quality Criteria for aquatic life. However, these criteria may be hard to meet because of limited data for the number of substances. There was general consensus that additional research was needed on minimum datasets and taxa diversity requirements, and the use of extrapolation methods to fill species sensitivity data gaps. The question whether current modelling success criteria are sufficient should also be considered relative to alternative approaches. If the alternative is the use of assessment factors (AF), it has clearly been shown that they provide an inconsistent method, i.e. the method is more conservative for large data sets ($n > 6$) than for small data sets ($n < 6$). In this context, optimising the method by making the AF dependent of the number of available toxicity data or allowing the application of the SSD for smaller sample sizes should be considered. There was general consensus that the current modelling criteria of REACH TGD are a guideline and that motivated deviation of these guidelines, based on a solid scientific justification, should always be possible on a case-by-case basis.

5. What are the research needs?

A variety of research and development needs were discussed, including methods validation, developing alternative estimation approaches, incorporating knowledge of chemical properties and exposure, and peer review and engaging with stakeholders. One identified research need was to compare trait-based SSDs with traditional strictly taxonomic-based SSDs, and to define what traits are most relevant to SSD generation. Alternative approaches should be explored, including focusing on sensitive taxa rather than broadly populating an SSD. However, there is uncertainty of what the sensitive taxa will be for many substances. A sensitive species approach may require novel methods development, including integrating chemical structure, genomic, traits and MOA information. An additional research question was whether critical body residue (CBR)-based SSDs could be developed by incorporating bioconcentration factors into the SSD generation. There was general consensus that extrapolation approaches and minimal dataset SSDs are better alternatives to generic safety factors, but validation against field and mesocosm data is required. MOA was considered to be an important determinant of species sensitivity and research is needed to determine linkages between MOA and SSD composition requirements. SSD research and development has been focused on acute toxicity data for water column organisms. The development and validation of chronic toxicity extrapolation methods and approaches applicable to other environmental compartments (such as sediment, soil and air) both remain significant research needs. How to best leverage knowledge of the compounds chemical properties, behaviour and environmental exposure scenarios should be explored. Finally, stakeholders and others in the scientific community should be engaged to assist in peer review, validation and tool improvement, and to facilitate communication of uncertainties and the value of research investment.

Group 2D

Moderator: P. Chapman
Rapporteur: S. Raimondo
S. Andres
D. de Zwart
A. Hart
A. Macken
S. Marshall
Y. Pan
H. Sanderson
K. Solomon
Z. Yan

The syndicate discussed the questions in the order they were presented and spent 50% of the allotted time on the first question.

1. *Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.*

Assumptions about frequency distributions:

- In practice a number of different distributions are fitted to SSDs.
- In practice, even for large numbers of data values, it is difficult to distinguish (e.g. via a statistical test) 2 similar distributions (such as a log normal and a log logistic) but they may give different estimates of HC5 because of differences in the fit to the data in the tail.
- Different subsets of the data will give rise to different estimates of the HC5. This raises the question of whether the objective is to protect a certain group of sensitive species, in which case testing may need to be targeted, or whether the objective is more general.
- The distribution of importance for risk management is the distribution of sensitivities for the community you're trying to protect (not tested species).
- Tools: Aldenberg and Jaworska (2000) and Aldenberg and Slob (1993): These and other methods are statistically rigorous but not necessarily ecologically rigorous. Some of the methods use a Bayesian approach. They involve fitting a log-normal or log-logistic distribution or similar to measures of toxicity. The species measured are assumed to be a random selection of species in the community of interest (or exchangeable using Bayesian terminology). Under these assumptions, estimates of HC5 and confidence intervals are statistically sound. These methods were the first to be proposed and have been extensively used. Software, such as ETX (van Vlaardingen *et al*, 2004) and SSD Master (CCME, 2013) are readily available and are easy to set up from an Excel spreadsheet.

- WebICE[1] (Raimondo *et al*, 2013): This method makes use of the historic database of toxicity values in ICE[1]. First a community of relevant species is identified (e.g. aquatic or wildlife species), and then toxicity levels for absent species (predicted values) are estimated using measured toxicities (surrogate observations) and interspecies correlations (or regressions). A complex set of filters can be used to exclude predicted data both prior to and during the fitting process. Each surrogate results in different values to the same predicted species (where models are available) but Web-ICE includes only one value for each species in the SSD, so values predicted by multiple surrogates are evaluated to ensure the most robust prediction is included. This process results in a set of toxicities, some of which are measured and some of which are predicted. Finally an HC5 is computed from the mixed set of toxicities using a log-logistic distribution. The sample of species is again assumed to be a random selection (exchangeable) from a community or population. Confidence intervals for HC5 are not computed according to sound statistical principles, so the application can give odd results for intervals. The method was developed by US EPA who have built an easy to use online tool that is backed up by an extensive historic database. The method has been described in several published peer-reviewed papers and a user manual[2]. US colleagues have considerable experience of using WebICE but it has not been widely used in Europe. The historic database is regularly updated, which can change the model set used to predict to species. HC5 estimates obtained today may be somewhat different to those obtained in future if an updated suite of models yields additional species for the SSD.
- The hSSD concept: This method is based on a Bayesian hierarchical model. It is statistically rigorous and does not assume that measured species are a random sample from a community or population. It is currently in the prototype stage, very few people have experience using it, and it needs to be evaluated more widely and more thoroughly. Effective evaluation requires knowledge of communities of species actually found in the field so it cannot be evaluated from a purely statistical point of view. It makes use of an historic database of toxicity values provided by RIVM and, whilst there might be overlaps, the data set is not the same as that used in WebICE.

2. *As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?*

    - It is important to first consider context and scoping.
    - We can have more diagnostic settings for screening resources for developing SSDs.
    - Taxonomic closeness can say something about communities, but not ecosystems.
    - Goal: need a protective HC5 with as little testing as possible.

[1] http://www.epa.gov/ceampubl/fchain/webice/

[2] http://www.epa.gov/ceampubl/fchain/webice/iceManual.html

The existence of taxonomic patterns (consistent sensitivity relationships) means that we must be cautious when extrapolating over large taxonomic distances, and also means that we will get better estimates of what we are interested in if we take patterns of sensitivity into consideration. Taxonomic structure of the community we are trying to protect needs to be taken into account in risk assessment and consideration of related differences in sensitivity can be useful in setting guidelines for protection of structure and function. Where sufficient data are available separate SSDs should be constructed for taxonomic- or sensitivity-groups to allow more ecological information to the incorporated into the assessment process.

3. *Do models based on prior knowledge provide advantages over other methods?*

Yes. The more that is known, the better the prediction will be. Methods that use prior knowledge will be better than methods that do not.

Prior knowledge can include what taxonomic groups might be more sensitive to a chemical class (e.g. molluscs, metals). Having this knowledge prior to developing SSDs can guide assessors to ensure that representatives of the sensitive taxonomic group are included in the SSD.

4. *Are current modelling success criteria, such as those identified in the REACH TGD, sufficient, overly prescriptive or insufficient?*

- The guidelines and criteria are fine, but it is important to define context. It is also important to distinguish between populations, communities, and ecosystems.
- Question: Can SSDs be used when there are fewer than 10 tested species?
- Question: Is it better to prescribe a criterion that has acceptable confidence intervals rather than a prescribed number of data species.
    - If we can show that confidence intervals and HC5 estimates obtained from 5 species are not materially different from those obtained from 10 species, can we assume that fewer species would be reliable enough for regulation purposes? Confidence intervals should indicate how well a method performs. This should be caveated with the discussion point above regarding *a priori* knowledge of sensitive taxa. If the 5 data points do not include the most sensitive taxa, but have robust confidence intervals, is it protective even if statistically sound?
    - Can/should existing criteria be replaced by confidence interval criteria? Should the criteria require either a confidence interval of a given size or use a specific list of taxa.
    - Can uncertainty factors be applied depending on amount of data used?
- It is better to have more information than less. But if datasets have a large number of common taxa (e.g. fish), then it might be difficult to characterise impacts to less represented species such as amphibians.
- We need to be sure to capture taxonomic diversity.

5. What are the research needs?

There is a need to:

- determine whether traits are meaningful in development of SSDS.
- evaluate SSDs against high quality mesocosm studies.
- develop criteria for an acceptable confidence interval for SSDs and HC5s.
- develop a model that takes account of the number and type of species in a community and that shows you the consequences/reliability of what you get. Validity criteria – what do we practically need?
- be able to extrapolate better to all ecosystems. There is no strong science based evidence that an SSD based on example criteria is protective for ecosystems, however, this argument also applies to the simplistic use of the toxicity value for the most sensitive species tested.
- Agree how confident we want to be? Back calculate how confident assessments are given current criteria.

4.3 Syndicate Session 3: Regulatory Considerations

Group 3A

Moderator: A. Peters
Rapporteur: M. Hamer
T. Barber
P. Craig
P. de Vries
G. Kon Kam King
K. Leung
L. Maltby

The following questions / concerns were discussed:

1. *Would the methods reviewed in this workshop be accepted for use in regulatory assessments under current guidance? If not, what steps would be needed to facilitate their acceptance in the future? What are the opportunities to update technical guidance?*

The use of SSDs is already widespread in regulation for data rich substances (e.g. for water quality standard derivation in various regulatory regimes), less common is the use of interspecies correlation approaches such as Web Ice and SSD when limited data are available. Some statistical instruments, for compiling SSDs from toxicity data and deriving threshold values, are already routinely in use for these purposes.

A possible limitation to the current application of SSD approaches within regulatory processes is the limited guidance available in some areas, although the availability of toxicity test data which fulfils the taxonomic diversity requirements established under several regulatory regimes is also an important potential limitation. Whilst the minimum data requirements for the use of an SSD differ between different regulatory regimes several European regimes require a minimum of 10 species, representing at least 8 different taxa and 5 different phyla. There is a need for balance in regulatory guidance between prescriptive approaches, which give more consistent outcomes, and flexible approaches based on best scientific practice. It is especially important that where professional or expert judgement is used the justification for the approach taken must be scientifically defensible and clearly documented.

The focus is on fulfilling the required number of species required for an SSD for it to be acceptable. Some flexibility around the number of species/taxa might be better with more emphasis on the uncertainty associated with the derived HC5. Additional guidance is required on the most appropriate ways to derive the confidence interval around the HC5. Furthermore, some attention should also be given to the uncertainty in HC5 that are hard or impossible to quantify (e.g. uncertainties associated with model assumptions). Currently the focus is on those sources of uncertainty that can be quantified, and more work is required to better address those areas of uncertainty which cannot be readily addressed statistically at present.

Extrapolation and estimation techniques are used within some regulatory areas, and some limited general guidance exists for establishing their validity and applicability from the OECD. Again there is a need to understand the uncertainties introduced through the use of extrapolated data.

2. *Should current guidance on the use of SSDs be revised in the light of the issues and approaches discussed in this workshop, e.g. number of species?*

Some aspects of current guidance could be reviewed, particularly in light of the practical experience gained through the more widespread application of SSDs in regulation. Whilst the taxonomic diversity criteria defined by the London Workshop (EC, 2001[*]) could be updated it is important to recognise that different applications of SSD have different requirements. It is likely that where there is an existing requirement for taxonomic diversity this would still be maintained. Knowledge concerning the mechanism of toxic action should be used when evaluating the appropriate number of species/taxa required for an SSD.

Guidance in the form of a decision tree would assist in identifying the most appropriate approach for any particular situation. There should be a principle of including as much information as possible, in an intelligent manner so that less reliable or relevant information makes a smaller overall contribution to the overall weight of evidence.

The assessment can be an iterative process but this will not always be the case.

3. *What implications are there for the interpretation of SSDs and HC5s in risk assessment and risk management?*

Flexibility requires an intelligent approach, which should be fully documented with the supporting scientific justification for any decisions taken.

A distinction needs to be made between protective and predictive applications of SSDs, as this can have implications for the data requirements of different models and approaches. The use of an SSD to derive a PNEC, or similar threshold, for a substance is typically a protective application, and aims to derive an HC5 for the overall community. The use of an SSD to assess impacts at a contaminated site undergoing risk management, where the Potentially Affected Fraction of the overall community might be derived, would typically be a predictive application of an SSD.

SSDs should ideally be applied within approaches which employ multiple lines of evidence, and are updated to include new information as it becomes available.

[*] EC. 2001. *Report of the expert consultation workshop on statistical extrapolation techniques for environmental effects assessment. London Workshop 18-19 January 2001.* Despite extensive efforts to track this report, we have only been able to locate a draft copy thanks to some participants. This draft is available by contacting the ECETOC Secretariat.

4. What are the research needs?

The applicability of toxicity extrapolation methods should be further validated for acute effects, and should also be evaluated for chronic effects.

There is a need to better understand the uncertainties within the assessment which are currently unquantifiable.

Further validation of SSDs derived from laboratory data against field and mesocosm studies is required, as is guidance on the different approaches (including their limitation) which can be taken.

Group 3B

Moderator: M. Warne
Rapporteur: M. Junghans
S. Belanger
C. Collin-Hansen
C. Eadsforth
M. Galay Burgos
J-P.Gosling
L. Posthuma
R. Wenning
P. Whitehouse

Whilst the original questions developed before the workshop were addressed in the other syndicates, group 3B choose to address another set of questions which were proposed by the workshop organisers immediately prior to this syndicate session. The new questions were:

1. *What can we do to improve regulatory use of species sensitivity distribution methods?*

2. *How can we achieve this?*

Syndicate session report

The 3 recommendations that arose from the syndicates discussions were to:

1. develop a compendium of SSD best practice;

2. use uncertainty to steer future research;

3. improve communication.

1. A compendium of SSD best practices

It was agreed, that the SSD methodology is a valuable regulatory and management tool since it can give more insight into the potential ecological effects than the assessment factor method (enabling better problem definitions) and it yields more generalisable results than a mesocosm-based methodology.

It was felt that a compendium of current best practices, the state of the science and answers to frequently asked questions would facilitate acceptance of SSDs by regulators and risk managers and their implementation in regulation and management. The compendium should be a technical document aimed at users with knowledge of SSDs and ecosystems. However, this would limit the usefulness of the compendium and therefore another document suitable for a general audience is also necessary.

During the workshop it was shown that SSDs are being derived differently by different jurisdictions, e.g. they have different minimal data requirements to sufficiently represent ecosystems of concern. Despite this, analyses of HC5 values (the SSD output used for standards or thresholds setting), showed that SSDs give robust results based on a relatively low number of input data, if the data are distributed uni-modally and the most sensitive taxonomic groups are included. Less robust HC5 values have to be expected, if a very sensitive

taxonomic group is overlooked or if the data are patchy and seem to have multiple-modes. This variability supports the idea of deriving a compendium that will highlight to the risk assessor and regulator when such considerations are important. Within such a compendium, decision trees to guide professional judgement were seen as a good way to avoid overly strict use of data requirements and derivation methods while ensuring clear identification of situations where the application of strict requirements are necessary, e.g. as laid down in the REACH Technical Guidance Document (ECHA, 2011). For example, while for some herbicides missing insect data might not have a severe impact on the accuracy of the SSD, the example of the chronic SSDs for triclosan shown by Anne Gosselin (section 3.18) underlines the general usefulness of falling back on a broader set of data requirements. Although the REACH criteria were pretty much fulfilled in the case of triclosan, the identification of the most relevant data was difficult, resulting in HC5 values, derived by considering different groups, ranging over 1 order of magnitude. In any case, the compendium should be flexible enough to allow the risk assessor to tailor the use of SSDs to the actual ecological question being considered.

To make the best use of the already existing SSD guidelines and methods, it was also proposed to promote databases to increase the availability of toxicity data and to reduce duplication of effort. Together with more research on the question of how *in vitro* and *in silico* approaches can be used within a compendium of best practices for use of SSDs in risk assessment, greater availability of such databases may boost the use of SSDs as a versatile, lower-tier approach within current environmental risk assessment schemes.

When using SSDs for predicting effects in the field, knowledge of effects of non-chemical stressors should be incorporated where available, to promote a multi-stress ecotoxicology/ecology analysis. This is very important to ensure risk managers or regulators do not 'jump to the wrong conclusion' and take action to fix a less important pressure, or fail to fix a problem that really does need attention. This multi-stress analysis may be based on existing ecological knowledge on optimal population growth conditions, as well as existing knowledge on ecosystem modelling. Improved liaising to ecology is indeed possible within SSD-derivation and interpretation, as shown in various presentations.

The compendium should also answer 'frequently asked questions' such as whether the use of an SSD partially or totally based on species from regionally or climatically different ecosystems would be scientifically sound, and if not – which options are suggested.

Finally the compendium would be an important document that will facilitate international harmonisation of the use of SSDs. It will not be possible to have a single internationally agreed method for deriving water quality guidelines/limits/standards. However, by presenting the state of the science it should be possible to harmonise individual components of the overall methodology. For example, agreement could be reached on the types of toxicity data (measures and endpoints) that can be used, or methods for assessing the quality of toxicity data. By providing a common platform the compendium could be used to establish international peer groups that could provide guidance on the appropriateness of decisions based on professional (expert) judgement. Such peer review conducted during the derivation of Environmental Quality Standards (EQSs) for specific pollutants under the EU Water Framework Directive has proven valuable and could help promote consistency when standards for the same substance are derived by different authorities. A compendium could also facilitate the implementation of SSDs in newly established environmental risk assessment

schemes in other countries, both for deriving criteria and for evaluation of risk management scenarios for contaminated ecosystems.

2. Uncertainty driven research

Throughout the workshop and in all 3 sessions from syndicate B, uncertainty was identified as an important and recurring issue. Studies should be conducted to identify the magnitude of the uncertainty of various components of the SSD methodology. Uncertainty may be related to lack of data, (non)representativity of data, mode of action considerations, and many other aspects of real exposure situations. An understanding of the mathematical magnitude of uncertainty alone may not be enough as it is possible that large sources of error may have little ecological importance, and *vice-versa*. Research should then be focussed on reducing the uncertainty of the most important sources uncertainty in the SSD methodology. The group felt that uncertainty-driven research would be an important means to improve SSDs and maximise their usefulness in a cost-efficient manner. An uncertainty driven research agenda is also likely to increase uptake of the other methods that can be used in combination with SSDs e.g. QSARs, Web-ICE.

A simple example of uncertainty-driven research would be the selection of chemicals (or species) to be used in ecotoxicity tests. If the toxicity of a chemical to a large number of species belonging to different taxonomic groups has been determined then the need for further research for that chemical may be low compared to a chemical that has been the subject of no or minimal toxicity testing. Another example is that very few SSDs have been conducted for non-chemical stressors (e.g. temperature, salinity) or the combined action of chemical and non-chemical stressors. Conducting such research could dramatically reduce uncertainty in the ecological relevance of single chemical SSDs, and place the risks posed by chemicals into a more meaningful context that addresses all possible pressures.

3. Communication

Communicating the success and limitations of the SSD methodology was felt to be essential. 2 targets of communication were identified: (1) regulators and stakeholders and (2) users and potential users, representing passive (results) and active (analysers) users, respectively. For the first group some basic communication gaps need to be bridged such as how the SSD method works, its underlying assumptions as well as the magnitude of the uncertainties of SSD-based risk assessments. It was also felt important to explain the implications of the fact that regulatory decisions are often based on single numerical values, which ignores the underlying uncertainties in the estimate and model assumptions. The compendium (the first of our proposals) would certainly help address some of these issues, but is likely to be quite technical and not appropriate for all users and potential SSD- result users. Hence, a way must be found to also simply communicate the boundaries of certainty around the predicted HC estimate. A clear and easy to understand communication strategy is needed including ‘success stories’ of the SSD method. Such communication should assist stakeholder acceptance of risk management measures and hence be an important step in improving environmental quality by regulatory means. Proposals for communication with the second group (active users) focussed mainly around the establishment of communities of practice – whereby users were constantly informing and educating colleagues of the latest developments and the state of the science. This could also be of great benefit to scientists and regulators in developing countries who are just beginning the task of environmental regulation and management of chemicals.

Group 3C

Moderator: A. Gosselin
Rapporteur: M. Barron
J-L. Dorne
S. Duquesne
M. Habekost
C. Michel
A. Ragas

The following questions / concerns were discussed:

1. *Would the methods reviewed in this workshop be accepted for use in regulatory assessments under current guidance? If not, what steps would be needed to facilitate their acceptance in the future? What are the opportunities to update technical guidance?*

Two general categories of methods were evaluated: fitting methods and tools, and extrapolation approaches. SSD distribution tools overviewed at the workshop included R and BurrliOz. R is a command-line based statistical programming software, which can be used to implement methods for deriving SSDs/HC5 analyses. BurrliOZ is a software specifically designed to derive SSDs/HC5. What BurrliOZ does could be also done in R. As such R is a more general software tool which, in its current form, is far less user friendly than BurrliOZ. Yet, R can be used to implement methods that can then be made available for more general use by incorporating them e.g. in a graphical user interface. BurrliOz was specifically developed to fit SSD models to data using multiple distribution types. Both provide approaches for quantifying uncertainty in HC5 estimations through rigorous statistical bases. The consensus recommendation was that both tools require more in depth evaluation and peer review before general acceptance.

Two SSD extrapolation tools were evaluated: Web-ICE and hSSD. ICE was considered to be a valuable tool for toxicity estimation to individual species, including identification of sensitive species. However, there were statistical concerns with the use of the ICE estimates within SSDs because of the potential correlation between the estimated values (e.g. how does the SSD generator deal with the correlation structure in toxicity data) that should be addressed before broad application. The hSSD tool required model ecosystem selection, potentially resulting in uncertainty in HC5 estimation and management/policy concerns. However, selecting the individual species of an ecosystem may conflict with the results of other studies that show that ecosystem sensitivity is not very dependent on ecosystem composition. The general consensus was that the hSSD tool required validation against measured SSD HC5 values and field results, and that the development of standardised model ecosystems should be considered. Both Web-ICE and hSSD were viewed as an opportunity to reduce the use of generic assessment factors.

2. *Should current guidance on the use of SSDs be revised in the light of the issues and approaches discussed in this workshop, e.g. number of species?*

SSDs should be the preferred alternative rather than using generic assessment factors. Their utility may increase if the 10 species/8 taxon group requirements could be relaxed with an acceptable level of uncertainty in HC5 estimation. Extrapolation tools (ICE, hSSD) and additional distribution fitting methods should be considered in the advancement of SSD regulatory applications.

3. *What implications are there for the interpretation of SSDs and HC5s in risk assessment and risk management?*

There is a need for balance between prescriptive guidance and user flexibility in terms of data quality, taxonomic and species number requirements, allowable extrapolation tools, and statistical approaches to HC5 estimation. Use of *a priori* knowledge of MOA and potentially exposed communities is recommended for determining SSD requirements. The group was uncertain on how best to apply and interpret protective values and quantitative protection goals, and what level of conservatism and subjectivity is reasonable. Peer review and uncertainty/sensitivity analyses by outside experts may facilitate an understanding of the degree of subjectivity in SSD generation. There was consensus that the interpretation of SSDs and HC5s in risk assessment and management should not follow a predefined recipe. It should be a case-by-case assessment in which all available data and knowledge are considered by experts in the field of ecological risk assessment.

4. *What are the research needs?*

A variety of research and development needs were considered that could improve future regulatory applications of SSDs. Incorporating dose-response or L(E)C50 confidence limits, rather than only point estimates of toxicity, could have value in representing the range of uncertainties in an SSD. A consistent theme was the need to compare SSD-based approaches to the use of generic AF values under different scenarios of data richness, and the need to explore uncertainty in relaxed (10 species/8 taxa group) requirements versus AF uncertainty and conservatism. Determination of the ecology and composition of representative ecosystems should inform requirements for taxa composition in SSDs. SSD-based estimates determined from various approaches and data richness scenarios should be compared to field data, and field monitoring should be performed to verify SSD-based predictions of community level effects. Research is also needed to determine how best to use available data (e.g. strict standardisation criteria with resulting loss of species diversity or use weighting based on data quality). The focus of SSD development has been on acute toxicity data, and chronic toxicity estimation approaches will need the same level of evaluation (e.g. minimum data sets, acute to chronic ratio estimation, lowest toxicity value approaches). Finally, there is a growing amount of information about chemicals that could be used to inform SSD development, application, and interpretation, including knowledge of ‘omics, mechanisms, chemical properties, and exposure scenarios.

Group 3D

Moderator: D. De Zwart
Rapporteur: A. Hart
S. Andres
P. Chapman
A. Macken
S. Marshall
Y. Pan
S. Raimondo
H. Sanderson
K. Solomon
Z. Yan

The following questions / concerns were discussed:

1. Would the methods reviewed in this workshop be accepted for use in regulatory assessments under current guidance? If not, what steps would be needed to facilitate their acceptance in the future? What are the opportunities to update technical guidance?

The syndicate group discussed and concluded that some uses of SSDs are already accepted in some jurisdictions, e.g. ETx in Europe, Canada, Australia and WebICE in USA. The group identified that the type of tools that are most appropriately applicable is strongly depending on the regulatory setting:

- Evaluation of water quality.
- Site-specific risk assessment.
- REACH.
- Retrospective assessment and assignment of causality to effects in ecosystems.
- The protection of endangered species.
- Non-regulatory settings, e.g. internal business assessments.

It was concluded that SSDs could in principle be applied in appropriate ways to all settings. However, the group mentioned that there was more confidence in applying SSDs in prospective applications, e.g. ETx where the statistical methodology is well developed (compared to retrospective/diagnostic applications). A remark was that the discussion could be better structured by addressing precision, accuracy and domain for different policy uses (hereunder retro- and prospective risk assessments) in the use of SSDs, and the specific research needs to increase the applicability and acceptance of SSDs.

Regulatory acceptance is considered to require:

- Retrospective analyses showing the reliability of the proposed approaches.
- The availability of guidance and decision-trees on how to use SSDs in each setting and how to overcome limitations.
- Confidence needs to be built about the extrapolation to untested species.

- The concept of the dependency of the slope of the SSD on toxic mode of action urgently needs to be validated before it can be used for purposes of extrapolation.
- New approaches should be developed in collaboration with regulators taking account of their needs, including preference for rules on what to use and how in each context.
- Benefits of the use of SSDs need to be demonstrated – this should include the benefits to the regulators themselves.
- Clear communication of approaches, results and limitations is needed.
- Formalised but appropriately conservative criteria are to be set in Tier-1, more complex use of weight of evidence and expert judgement is needed in higher tiers?

The group does not expect a major contribution from SSD to the reduction of animal testing. Coping with the trend for less testing in some jurisdictions will be a challenge for the SSD approach.

2. Should current guidance on the use of SSDs be revised in the light of the issues and approaches discussed in this workshop, e.g. number of species?

The group concluded that more guidance is needed, mainly with respect to:

- The number and nature of tested species required for the construction of a valid SSD.
- The same holds for the number and nature of taxonomical groups to be tested.

Lengthy discussions lead to the conclusion that data quality is a major concern:

- Firm and consistent criteria for data-quality should be formulated and standardised for all uses of SSDs – lack of this now is considered a problem.
- The group is concerned about the possibilities for manipulating the SSD output by the selection of input data (cherry picking).
- The group was uncertain about the benefit of adding more data with weighting methods for reliability - more research is needed before this can be recommended.
- Lack of established testing guidelines for non-standard species is a potential problem.

3. What implications are there for the interpretation of SSDs and HC5s in risk assessment and risk management?

The final topic addressed by the syndicate group on the interpretation value of SSDs leads to a very short conclusion: Garbage input will automatically lead to garbage output.

4. What are the research needs?

The syndicate group identified the following research needs:

- Guidelines need to be developed on how to deal with data quality.
- Guidance needs to be formulated for the use of non-standard test species.
- Guidance should be developed on which methods and tools can be used to generate SSDs – this requires sensitivity analysis, identification of causes of differences, etc.
- Methods need to be developed to include censored input data, such as greater-than values for toxicity endpoints. These are addressed in SSD-Master.
- Methods may be developed to expand on data availability by adding less strictly selected input data and putting less weight on their inclusion, based on reliability of data.

5. Main overall conclusions

- SSDs can, in principle, be applied to all regulatory settings if appropriately done as suggested in the following points.
- Regulatory acceptance may require the formulation of a challenging set of arguments and proof of concept.
- A strong focus on data quality and desire for strict standardisation of approaches is needed.
- There is a need to demonstrate when and where criteria of acceptance of data and/or requirements can be relaxed.

4.4 Feedback from plenary sessions

After the opening presentation in which Leo Posthuma covered the broader aspects of using SSDs in environmental protection and management, emphasising their origins and utility, the focus shifted to ecological considerations. How can a small set of toxicity data from a limited number of species, be used to assess risks or potential impacts in real world situations? Ecosystems differ, covering a range of different habitats, geographic locations and contain differing assemblages of species with complex, maybe unknown interactions. It is not unreasonable to assume that individuals, species, communities and ecosystems will all differ in their sensitivities to stressors, including chemicals.

The many ecological considerations raised with respect to SSDs were covered by the various presentations and highlighted in the questions posed and the responses given in the Syndicate Sessions. There was considerable discussion about the importance of species sensitivities and traits, particularly reproductive strategies, community structure and resilience, ecological redundancy, warm, cold, salt and freshwaters and the importance of considering modes of action of chemicals.

It is recognised that establishing protection goals is the important first step in the process. SSDs are used in both prospective risk- and retrospective impact assessment of chemicals. Prior to the registration of a particular chemical use and consequent environmental exposure, a prospective risk assessment needs to establish that its use will not cause unacceptable risk. In contrast, retrospective impact assessment uses diagnostic tools to identify the cause of existing adverse effects, including chemical thresholds, to quantify expected impacts. Prospective assessments need to be protective, generally deriving a single value from an SSD using the inverse method – establishing the concentration at which there is a tolerably low risk of an unacceptable effect. This is done when setting Environmental Quality Standards (EQS) for water quality, deriving Predicted No Effect Concentrations (PNECs) under REACH or Regulatory Acceptable Concentrations (RACs) under EU Plant Protection Product legislation, with similar procedures oriented on deriving and using quality standards in other jurisdictions.

It is important that the estimate or prediction is protective; perhaps with the proviso that it should not be over-protective and restrict safe chemical use. However, it is less important than, for example, correctly positioning each species within the distribution. There was some difference of opinion at the workshop about the potential for deriving *valid* generic PNEC values, i.e. a concentration that assures protection under all circumstances. However, limited data were presented to suggest that provided the protection goal is to achieve no worse than negligible effects, it is possible to derive a generic PNEC from an HCx derived from a generic SSD by applying appropriate assessment factors. As discussed in syndicate group 1B, a consequence of using a generic SSD to protect all ecosystems (including the most sensitive) implies that the margin of safety would vary when considering less-sensitive systems, and there is the potential to be overprotective, leaving scope for refinement in decision-making at higher tiers of assessment.

SSDs can have a role as a diagnostic tool in a site specific, retrospective assessment where they can help assign causality (‘reasons for poor status’) or perhaps in site remediation. In these situations, an SSD is used in a predictive and quantitative manner, quantifying expected impacts (as the fraction of species potentially affected when exposed to a contaminated sample). In this sort of application, the SSD needs to reflect reality for the ecosystem of concern and a site or scenario specific assessment is more likely to be representative.

Whilst SSDs have been, and continue to be applied across different sectors, there is still much to understand to allow them to be used more widely and confidently in taking environmental management decisions. Undoubtedly, questions remain about further field validation. Further insight into some of the considerations explored within the syndicate groups will allow the use of SSDs in a more flexible, perhaps less prescriptive manner when needed, increasing their versatility as one of the available tools for ERA.

The workshop has been effective in drawing together a range of highly relevant skills and experience from around the world. The focus on regulatory applications of SSDs is both rewarding and refreshing, as there are multiple practical and societally valued uses whilst there is a clear view on method limitations too.

Regulators are placing increasing reliance on the use of SSDs because it is thought that they make better use of the available data, have more relevance to threshold effects in the real world than traditional deterministic methods, and encourage all interested parties to think about the relationship between chemical exposure and risks to wildlife, and the environment on which the services we use depend. The regulatory communities around the world are starting to see the application of SSDs to other stressors (e.g. radionuclides) and to help understand risks from mixtures, and to situations of site-specific contamination where we need to account for the many other factors that can influence decision making. This is a very welcome development.

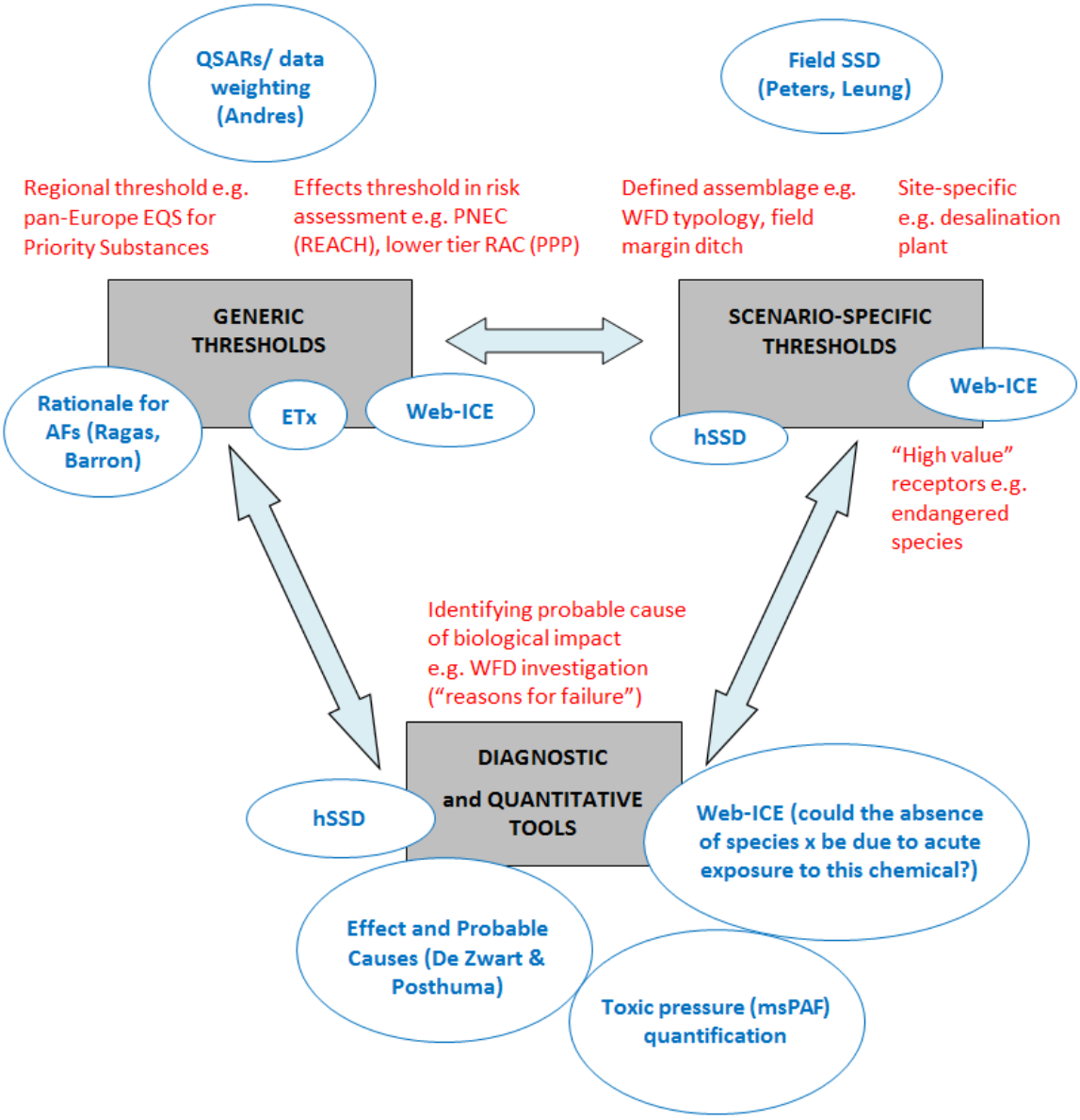
The wider uptake of SSDs is prevented by a perception that the approach is 'data hungry' so that only a small proportion of substances satisfy the minimum criteria that were set 3 decades ago for United States Ambient Water Quality Criteria and a decade ago for REACH in the context of deriving methods for transparent and reproducible derivation of protective quality criteria using NOECs as test endpoints. It is interesting to learn that, under some circumstances, useful insights can be gained and progress can be made with quite small datasets, and/or with data sets collected in the context of a novel problem, such as the use of EC50-based SSDs for ranking sanitation priorities. The development of non-testing approaches for augmenting test data to be fed into SSDs, or making better use of the data we already have, is particularly exciting. The prospect of a more rational approach to the selection of assessment factors is also encouraging because this will make the process of EQS derivation more transparent and reproducible. Some of these techniques can be applied in a regulatory context almost immediately. Regulators are seeing a greater interest in the ecological relevance of the way SSDs are constructed and/or interpreted. This is exemplified by studies where ecosystem differences and characteristics are actively taken into account in deriving and using SSDs; a better understanding of protection goals and the ecological significance of our risk management decisions is of considerable regulatory relevance.

To help appreciate how the tools and approaches discussed at the workshop could be applied to different regulatory questions, they were subsequently mapped onto the 3 distinct regulatory activities identified at the workshop (Figure 1):

1. The derivation of generic boundaries like an HC5 (which is used as the basis for generic criteria such as EQSs and PNECs in international risk assessments) that need to be applied to many different locations, perhaps over very large geographical regions. These are assumed to offer sufficient protection everywhere, even in the most sensitive systems.

2. The derivation of scenario-specific thresholds that more closely reflect local conditions but which may not be transferable from one place to another.
3. Identifying the causes of biological impact ('diagnosis') or expected impact magnitudes of existing (mixture) contamination, so that regulators can make sure that any remedial action focuses on the correct pressure and at a time of pressure on resources, effort is focussed on the sites of highest risks and impacts first.

Figure 1: Mapping tools (blue font) onto risk/effects assessment scenarios (red font)



Perhaps this sort of thinking will help identify where certain tools have a particular role to play. This could be further enhanced by thinking about the particular strengths of the various tools and techniques. Figure 2 was developed after the workshop but illustrates how the various tools and techniques now available could help deal with some of the challenges faced by risk assessors and regulators.

Figure 2: Linking tools with the challenges faced by risk assessors

Strong benefit | **Possible benefit** | **Uncertain benefit/not relevant**

Challenge	Tool					
	ETx	***hSSD***	***Web-ICE***	***Rationale for AFs***	***QSARs / weighting***	***Field and community SSDs***
Adequacy of taxonomic 'spread' in data for SSD		Receptor community defined	Better taxonomic representation		Includes data that would otherwise not be used	
Choice of AF applied to HC5						
Ecological relevance		Receptor community defined	Better taxonomic representation		Better taxonomic representation	
Consistency in estimation of thresholds	Standardised method with good supporting documentation	'Bespoke' receptor community		Reduces variability and bias		
SSDs for other stressors	In theory		In theory, could be developed			
Precision of conclusions	CIs around HC5 stabilise		Add species until CIs around HC5 stabilise			

Even when regulators are able to call upon a wider range of techniques (as is now probably the case), there will be a need for value judgement and this could give rise to different interpretations even when assessors are presented with the same information. Sometimes this will not matter but, for some questions, such as the development of pan-continental generic thresholds, this can be a problem. The preferred response is not to further restrict flexibility in technical guidance because this will only 'fossilise' the science and innovation. Instead, the regulatory authorities support the idea of devising guidance that prompts the assessor to think carefully about the protection goal (e.g. generic protection, local protection, contaminated site assessments), what is known about the particular case (e.g. local water chemistry conditions), the uncertainties being dealt with, and the growing range of tools now available.

There was general consensus that scientifically sound extrapolation approaches and even relatively minimal dataset SSDs can be better alternatives to deriving toxicity threshold concentrations/PNECs than the application of generic assessment factors to simple aquatic toxicity tests, However, continued validation against field and mesocosm data is required to ensure that a threshold like an EQS or PNEC has ecological relevance. Furthermore, the results of extrapolation from SSDs should be critically assessed using all available knowledge on the substance and related substances. It was agreed that the SSD methodology is a valuable regulatory and management tool since it can give more insight into the potential ecological effects than the assessment factor method (enabling better problem definitions) and it yields more generalisable results than a mesocosm-based methodology. However, it is important to validate predictions of risk or thresholds such as EQSs against field and mesocosm data.

It was proposed that a compendium of current best practices, the state of the science and answers to frequently asked questions would facilitate acceptance of SSDs by regulators and risk managers and their implementation in regulation and management. The compendium should be a technical document aimed at users with knowledge of SSDs and ecosystems. However, this could limit the usefulness of the compendium and therefore another document suitable for a general audience is also necessary.

There are a range of statistical SSD tools in use which can vary with regional regulatory application. For example the ETX tool is accepted for use in regulatory risk assessments in Europe. ETX uses data from acute or chronic toxicity tests as does the BurrliOZ tool used by Australian and New Zealand regulators although the latter applies the Burr family of distributions. Other variations of ETX-type tools are used in other jurisdictions such as Canada and China. A rather different tool, Web-ICE is available for regulatory use in the USA. Web-ICE uses both measured toxicity test data for a test chemical as well as estimated toxicity values based on interspecies correlations. There is a helpful guide for users to avoid inappropriate use of the tool, e.g. deriving HC5 values outside reasonable statistical limits.

Other examples of available tools include Mosaic and a tool developed by P&G using the R software. Both of these represent variations or extensions on the ETX approach. The hSSD tool developed by Peter Craig at the University of Durham uses taxonomic patterns of sensitivity and generates SSDs for specified communities. This prototype tool was considered to be statistically rigorous but requires more evaluation to determine its applicability in risk assessment.

During the workshop discussions of SSD tools and their applications several emerging themes developed. These themes should be considered in future development and application of SSD tools and include the following:

- There is need to specify the protection goals more precisely. This can be important both in prospective and retrospective applications of SSD tools.
- Taxonomic distance is important, e.g. a fish species is likely to have more similar sensitivity to a given chemical to that of other fish species compared with more distantly related taxa such as algae, molluscs, or insects.
- Including prior information is useful, e.g. Web-ICE, hSSD.
- MOA can be important in deciding if particular taxonomic groups are expected to differ in their sensitivity to a chemical compared to the broader community, e.g. algal and macrophyte sensitivity to herbicides.
- Choice of statistical distribution does not seem to have a strong influence on the derived HC5.
- The current REACH guidance/criteria on the use of SSDs for deriving HC5/PNEC values were considered to be basically reasonable (the requirement to test a diverse range of taxa was understood and accepted), but experience gained so far indicated that it was also considered to be over-prescriptive and not flexible enough for certain situations, especially for deriving expected impacts (PAF) given an ambient exposure level. As a response, it was suggested a compendium of best practices be compiled. Using SSDs requires a thoughtful decision process and should not be over prescriptive following a regimented check-list mentality. Particular challenges include the need to be protective, but not overly so, with minimum new testing, e.g. can SSDs be used with less than 10 tested species?

Research needs for SSD tool development include the following considerations:

- Tools for regulatory decision making should be given high priority with particular focus on i) SSDs for chronic toxicity, ii) validating HC5s with mesocosms and real ecosystems and iii) maximising the use of available data, e.g. by applying weighting criteria.
- Further development of tools for assessing mixtures of chemicals.
- Trait-based SSDs appear to offer advantages over conventional taxonomic based approaches, but there is currently no practical application.
- SSDs for more taxa including plants and, possibly, micro-organisms.
- Cheminformatic approaches.
- Focus on sensitive groups.
- The usefulness/applicability of SSDs for defined communities.
- Internal dose (CBR)-based approaches have potential to incorporate mechanistic toxicokinetic/toxicodynamic modelling approaches that could help explain sensitivity differences between taxa/traits.
- Quantifying uncertainty as an alternative to standard assessment factors.
- What level of confidence do current criteria provide.

Given the various uses of SSDs discussed at the workshop, and the use of SSDs in decision support situations ranging from generic to specific, it is evident that application of expert knowledge can improve decision making when doing practical assessments of problems with chemical in the environment.

5. CONCLUSIONS AND RECOMMENDATIONS

The aim of the workshop was to discuss and report current thinking on when and how species sensitivity distributions (SSDs) should be used and how the methodology could be further developed to improve the quality and usefulness of decision making in environmental protection and management of chemicals.

The workshop covered the broader aspects of the use of SSDs in environmental protection and management, recent developments and specific case studies. In addition, there were sessions which focused on ecological considerations, statistical considerations and regulatory considerations. There was general consensus that, where data permit, the SSD approach should provide a more useful and transparent assessment of hazard thresholds than a deterministic approach using generic assessment factors. However, validation against field and mesocosm data is required where data permit, the SSD approach should provide a more useful and transparent assessment of risks than a deterministic approach using generic assessment factors. The ability to quantify uncertainty is important but could be used more explicitly in decision-making. There is also a need for better validation against field and mesocosm data. It was further agreed, that the SSD methodology is a valuable regulatory and management tool since it can give more insight into the potential ecological effects than the assessment factor method (enabling better problem definitions) and it yields more generalisable results than a mesocosm-based methodology.

For the future, it was proposed that a compendium of current best practices, the state of the science and answers to frequently asked questions would facilitate acceptance of SSDs by regulators and risk managers and their implementation in regulation and management. The compendium should be a technical document aimed at users with knowledge of SSDs and ecosystems. However, this would limit the usefulness of the compendium and therefore another document suitable for a general audience is also necessary.

In line with current uses in decision making, various research areas were identified to improve the usefulness and validity of output generated with SSDs to solve the array of problems encountered. The research areas identified in the various syndicate sessions have been listed in Table 1. Although not given a priority during the workshop, the report authors will seek an indication of priority for the work from the workshop attendees.

The table below collates the research ideas mentioned in the Syndicate sessions, and thereafter collated and sorted into subgroups. In some cases, similar suggestions were merged. The subgroups are, first, the use of SSDs in various decision contexts (protection, quantitative assessment, diagnosis). Secondly, in any decision context, the output of SSDs should be relevant for the ecosystem situation considered; this encompasses various research needs. Thirdly, guidelines should be adapted to accommodate standardisation for criteria setting under novel scientific insights, as well as novel uses of SSDs in other context. The fourth area addresses the underlying improvements that can be made in modelling as well as in data used for assessments. Again, various options are given. Attention for accommodating further knowledge sources, such as mode of action and body burdens is foreseen. Finally, decision making with SSDs requires attention for uncertainties, their types and origins and the options for reducing uncertainty.

Table 1: Identified research areas

Research area	Description
Uses of SSD	Collate and review the uses of SSDs for purposes other than estimating the HC5 (e.g. using the entire SSD for probabilistic risk assessment and deriving other values (say HC50) for trigger management action).
Ecology	Investigate whether an approach which allows better extrapolate to all ecosystems is viable.
Ecology	Compare trait-based SSDs with traditional strictly taxonomic-based SSDs, and to define what traits are most relevant to SSD generation. Alternative approaches should be explored, including focusing on sensitive taxa rather than broadly populating an SSD. However, there is uncertainty of what the sensitive taxa will be for many substances. A sensitive species approach may require novel methods development, including integrating chemical structure, genomic, traits and MOA information.
Ecology	Compare SSD-based approaches to the use of generic AF values under different scenarios of data richness, and the need to explore uncertainty in relaxed (10 species/8 taxa group) requirements versus AF uncertainty and conservatism. Determination of the ecology and composition of representative ecosystems should inform requirements for taxa composition in SSDs. SSD-based estimates determined from various approaches and data richness scenarios should be compared to field data, and field monitoring should be performed to verify SSD-based predictions of community level effects.
Ecology	(Further) Develop a model that takes account of the number and type of species in a community and that shows the consequences/reliability of the results. Establish what validity criteria are needed.
Ecology	Determine what additional ecological knowledge needs to be included to add value for the risk assessors.
Guidelines	Develop a formal and transparent decision tree approach that is inclusive of the available data, and that considers the generic or specific use of SSDs in environmental protection and management.
Guidelines	Develop guidelines on how to deal with data quality (of the input data on species sensitivities, or sometimes functions sensitivities).
Guidelines	Develop guidance on the use of non-standard test species.
Guidelines	Develop guidance on which methods and tools can be used to generate SSDs – this requires sensitivity analysis, identification of causes of differences, etc.
Model development and validation	Investigate the limitations of the models and whether they are fit for the purpose for which they are used.
Model development and validation	Evaluate the viable methods for incorporating all relevant data in SSDs

Research area	Description
Model development and validation	Further validation of SSDs derived from laboratory data against field and mesocosm studies is required, as is guidance on the different approaches (including their limitations) that can be taken.
Model development and validation	Further validation for extrapolations that are in relevant models (i.e. hSSD and Web-ICE) and of consequences for HC5 uncertainty.
Model development and validation	Validation of hSSD scenario-specific HC5s relative to the field and/or mesocosm studies.
Model development and validation	Critically review whether any of the growing amount of information types about chemicals and their impacts that is now available should be used to inform SSD development, application, and interpretation, including for example knowledge of omics, mechanisms, chemical properties, and exposure scenarios.
Toxicity data	Research is needed to determine how best to use available data (e.g. strict standardisation criteria with resulting loss of species diversity or use weighting based on data quality). The focus of SSD development has been on acute toxicity data, and chronic toxicity estimation approaches will need the same level of evaluation (e.g. minimum data sets, acute to chronic ratio estimation, lowest toxicity value approaches). Develop better application of toxicological data in SSDs, e.g. using more chronic data, mechanistic understanding. Develop methods to expand on data availability by adding less strictly selected input data and putting less weight on their inclusion, based on reliability of data.
Toxicity data	Develop methodology to improve the use of predictive modelling to overcome limited data sets. The applicability of toxicity extrapolation method should be further validated for acute effects, and should also be evaluated for chronic effects. Develop and extend software tools to add the capacity to predict chronic toxicity and approaches applicable to other environmental compartments (such as sediment, soil and air) both remain significant research needs.
Toxicity data	Investigate the value of including microorganisms in SSDs to protect ecosystem functions e.g. when assessing the ecological risk of fungicides, investigate the effects of including various fungal species in the test battery and incorporating their data into the SSD; Microorganisms should be considered in the HCx derivation but development is currently hindered by the lack of available approved testing procedures for different groups of microorganisms.
Critical body burden	Investigate whether critical body residue (CBR)-based SSDs could be developed.
Mode of action	MOA is an important determinant of species sensitivity. Research is needed to determine linkages between MOA and SSD composition requirements. Investigate whether it is possible to treat MoA in the statistical models in the same way taxonomic distance is being used? (In particular, is this feasible for Web-ICE and hSSD?)

Research area	Description
Uncertainty	There is a need to better understand the uncertainties within the assessment which are currently unquantifiable. Studies should be conducted to identify the magnitude of the uncertainty of various components of the SSD methodology. Uncertainty may be related to lack of data, (non)representativity of data, mode of action considerations, and many other aspects of real exposure situations. An understanding of the mathematical magnitude of uncertainty alone may not be enough as it is possible that large sources of error may have little ecological importance, and *vice-versa*. Research should then be focussed on reducing the uncertainty of the most important sources uncertainty in the SSD methodology. The group felt that uncertainty-driven research would be an important means to improve SSDs and maximise their usefulness in a cost-efficient manner. An uncertainty driven research agenda is also likely to increase uptake of the other methods that can be used in combination with SSDs e.g. QSARs, Web-ICE.
Uncertainty	A simple example of uncertainty-driven research would be the selection of chemicals (or species) to be used in ecotoxicity tests. If the toxicity of a chemical to a large number of species belonging to different taxonomic groups has been determined then the need for further research for that chemical may be low compared to a chemical that has been the subject of no or minimal toxicity testing. Another example is that very few SSDs have been conducted for non-chemical stressors (e.g. temperature, salinity) or the combined action of chemical and non-chemical stressors. Conducting such research could dramatically reduce uncertainty in the ecological relevance of single chemical SSDs, and place the risks posed by chemicals into a more meaningful context that addresses all possible pressures.

ABBREVIATIONS

AF	Assessment factor
AOP	Adverse outcome pathway
AWQC	Ambient water quality criteria
CBB	Critical body burden
CBR	Critical body residue
CCME	(Canadian) Council of Ministers of the Environment
CSD	Community sensitivity distributions
EQS	Environmental quality standards
EPA	(US) Environmental Protection Agency
EPT	Ephemeroptera–plecoptera–trichoptera taxa
ERA	Environmental risk assessment
f-SSD	field-based species sensitivity distribution
FWQG	(US) Federal water quality guideline
HCp	Hazardous Concentration for p% of species (where p is usually 5)
hSSD	Hierarchical species sensitivity distribution
ICE	Interspecies correlation estimation
LAS	Linear alkylbenzene sulfonate
LC50	Lethal concentration for 50% of test population
LD50	Lethal dose for 50% of test population

LLHC5	Lower limit hazardous concentration for 5% of species
MCDA	Multi-criteria decision analysis
MDR	Minimum data requirement
MOA	Mode of action
MSS	Mode-specific sensitivity
NOEC	No observed effect concentration
PBPK	Physiologically based pharmacokinetic
PEC	Predicted environmental concentration
PNEC	Predicted no effect concentration
PNOF	Potentially not occurring fractions
QSAR	Quantitative structure activity relationship
QSSR	Quantitative species sensitivity relationship
RAC	Regulatory acceptable concentration
REACH	EU regulatory framework for the Registration, Evaluation and Authorisation of Chemicals
SSD	Species sensitivity distribution
SQG	Sediment quality guidelines
TGD	Technical guidance document
WFD	European Water Framework Directive
WoE	Weight of evidence
WQC	Water quality criteria

BIBLIOGRAPHY

Aldenberg T, Jaworska JS. 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicol Environ Saf* 46(1):1-18.

Aldenberg T, Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicol Environ Saf* 25(1):48-63.

Azevedo LB, Van Zelm R, Elshout PMF, Hendriks AJ, Leuven RSEW, Struijs J, De Zwart D, Huijbregts MAJ. 2014. Species richness – phosphorus relationships for lakes and streams worldwide. *Global Ecology and Biogeography*, in press.

CCME. 2007. A Protocol for the Derivation of Water Quality Guidelines for the Protection of Aquatic Life 2007. http://www.ec.gc.ca

CCME. 2013. Determination of Hazardous Concentrations with Species Sensitivity Distributions, SSD Master. Ottawa, ON, Canada: Canadian Council of Ministers of the Environment. Report 38.

CEPA. 1999. Canadian Environmental Protection Act 1999. http://www.ec.gc.ca

Craig PS, Hickey GL, Luttik R, Hart A. 2012. On species non-exchangeability in probabilistic ecological risk assessment. *J Roy Stat Soc Stat Soc A* 175:243-262.

Craig, PS. 2013. Exploring novel ways of using species sensitivity distributions to establish PNECs for industrial chemicals. Technical Report. http://dro.dur.ac.uk/13383/

De Hoop L, Smit M, Huijbregts MAJ, Leuven RSEW, Schipper AM, Hendriks AJ. 2011. Sensitivity of arctic species to oil and other contaminants in comparison to other species. *Environ Sci Technol* 45:9017-9023.

Dyer SD, Versteeg DJ, Belanger SE, Chaney JG, Mayer FL. 2006. Interspecies correlation estimates predict protective environmental concentrations. *Environ Sci Technol* 40:3102-3111.

EC. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (Water Framework Directive or WFD).

EC. 2006. Regulation (EC) no 1907/2006 of the European Parliament and Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency. (The latest consolidated version of REACH(December, 2011) is available on the legislation page of the European Chemicals Agency website.

EC. 2008. The marine strategy framework directive (MSFD) 2008/56/EC on establishing a framework for community action in the field of marine environmental policy adopted in July 2008.

EC. 2009. Regulation (EC) No 1107/2009 of the European Parliament and Council of 18 December 2006 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC Official Journal of the European Union L 309/1 from 24.11.2009

EC. 2011. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) Technical Report - 2011 - 055 Guidance Document No. 27 Technical Guidance For Deriving Environmental Quality Standards. European Commission.

ECHA. 2011. Guidance on information requirements and chemical safety assessment. Part B: Hazard assessment. http://echa.europa.eu/documents/10162/13643/information_requirements_part_b_en.pdf

ECETOC. 2008. Probabilistic approaches to marine hazard assessment. Workshop Report No 15. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

Elshout PMF, Dionisio Pires LM, Leuven RSEW, Wendelaar Bonga SE, Hendriks AJ. 2013. Low oxygen tolerance of different life stages of temperate freshwater fish species. *J Fish Biology* 83:190-206.

Fedorenkova A, Lenders HJR, Ouborg J, Breure AM, Hendriks AJ. 2010. Ecotoxicogenomic: bridging the gap from genes to population, *Environ Sci Technol* 44:4328-4333.

Fedorenkova A, Vonk JA, Lenders HJR, Creemers R, Breure AM, Hendriks AJ. 2012. Ranking ecological risks of multiple chemical stressors on amphibians, *Environ Toxicol Chem* 31:1-6.

Fedorenkova A, Vonk JA, Breure AM, Hendriks AJ, Leuven RSEW. 2013. Tolerance of native and non-native fish species to chemical stress: a case study for the river Rhine. *Aquatic Invasions* 8:231–241.

Gilbert CS, Lui, Bjorgasaeter A, Leung KMY. 2014. Deriving field based sediment quality guidelines from the relationship between species density and contaminant level using a novel nonparametric empirical Baysian approach. *Environ Sci Pollution Res* 21:177-192

Golsteijn L, Van Zelm R, Hendriks AJ, Huijbregts MAJ. 2013. Statistical uncertainty in hazardous terrestrial concentrations estimated with aquatic ecotoxicity data. *Chemosphere* 93:366-372.

Golsteijn L, Van Zelm R, Veltman K, Musters G, Hendriks AJ, Huijbregts MAJ. 2012. Including ecotoxic impacts on warm-blooded predators in Life Cycle Impact Assessment, Integrated Environmental Assessment and Management 8:372–378.

Hendriks AJ. 2013. How to deal with 100,000+ substances, sites, and species: Overarching principles in environmental risk assessment. *Environ Sci Technol* 47:3546-3547.

Hendriks AJ, Awkerman JA, De Zwart D, Huijbregts MAJ. 2013. Sensitivity of species to chemicals: differences between test types (LC50, LD50), cold-blooded and warm-blooded species and modes of action. *Ecotoxicol Environ Saf* 97:10-16.

Hickey GL, Craig PS, Luttik R, De Zwart D. 2012. On the quantification of interest variability in ecotoxicity data with application to species sensitivity distributions. *Environ Toxicol Chem* 31(8):1903-1910.
Kooijman S. 1987. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Research* 21(3):269-276

Kwok KWH, Bjorgasaeter A, Leung KMY, Gilbert CS Lui, Gray JS, Shin PKS, Lam PKS. 2009. Deriving site-specific sediment quality guidelines in Hong Kong marine environments using field based species sensitivity distributions. *Environ Toxicol Chem* 27:226-234.

Leung KMY, Bjorgesater A, Gray JS, Li WK, Lui GCS, Wang Y, Lam PKS. 2005. Deriving sediment quality guidelines from field-based species sensitivity distributions. *Environ Sci Technol* 39:5148-5156.

Maltby L, Blake N, Brock TCM, van den Brink PJ. 2005. Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems. *Environ Toxicol Chem* 24:379-388

Maltby L, Brock TCM, van den Brink, PJ. 2009. Fungicide risk assessment for aquatic ecosystems: Importance of interspecific variation, toxic mode of action and exposure regime. *Environ Sci Technol* 43:7556-7563.

Sijm DTHM, Van Wezel AP, Crommentuijn T. 2002. Environmental risk limits in the Netherlands. In Posthuma L, Suter GW, II, Traas TP, eds, Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, FL, USA, pp 221-253.

Smit MGD, Bechman RK, Hendriks AJ, Skadsheim A, Larssen BK, Baussant T, Bamber S, Sanni S. 2009. Relating biomarkers to whole-organism effects using species sensitivity distributions: a pilot study for marine species exposed to oil. *Environ Toxicol Chem* 28:1104-1109.

Van den Brink PJ, Blake N, Brock TCM, Maltby L. 2006. Predictive values of species sensitivity distributions for effects of herbicides in water. *Human Ecol Risk Assessment* 12:645-674.

Van Straalen NM, Denneman CAJ. 1989. Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety* 18:241-251

van Vlaardingen TP, Traas TP, Wintersen AM, Aldenberg T. 2004. *ETX* 2.0: A Program to Calculate Hazardous Concentrations and Fraction Affected, Based on Normally Distributed Toxicity Data. RIVM Report 601501028/2004.

APPENDIX A: LIST OF PARTICIPANTS

Name		*Affiliation*	*E-mail*
Tom	**Aldenberg**	RIVM, NL	tom.aldenberg@rivm.nl
Sandrine	**Andres**	INERIS, France	sandrine.andres@ineris.fr
Timothy	**Barber**	Environ, USA	tbarber@environcorp.com
Mace	**Barron**	US Environmental Protection Agency	barron.mace@epa.gov
Scott	**Belanger**	Procter & Gamble, USA	belanger.se@pg.com
Peter	**Chapman**	Unilever, UK	pe51ter.chapman@gmail.com
Christian	**Collin-Hansen**	Statoil, Norway	chrc@statoil.com
Peter	**Craig**	University of Durham, UK	p.s.craig@durham.ac.uk
Pepijn	**de Vries**	IMARES, NL	pepijn.devries@wur.nl
Dick	**de Zwart**	RIVM, NL	dick.de.zwart@rivm.nl
Jean Lou	**Dorne**	EFSA, Italy	jean-lou.dorne@efsa.europa.eu
Sabine	**Duquesne**	UFZ, Germany	sabine.duquesne@uba.de
Scott	**Dyer**	Procter & Gamble, USA	dyer.sd@pg.com
Charles	**Eadsforth**	Shell, UK	charles.eadsforth@shell.com
Chenglian	**Feng**	Chinese Research Academy	fengcl@craes.org.cn
Malyka	**Galay Burgos**	ECETOC, Belgium	malyka.galay-burgos@ecetoc.org
John Paul	**Gosling**	University of Leeds, UK	j.p.gosling@leeds.ac.uk
Anne	**Gosselin**	Environment Canada, Canada	anne.gosselin@ec.gc.ca
Maike	**Habekost**	BASF, Germany	maike.habekost@basf.com
Mick	**Hamer**	Syngenta, UK	mick.hamer@syngenta.com
Andy	**Hart**	FERA, UK	andy.hart@fera.gsi.gov.uk
Jan	**Hendriks**	Radboud University, NL	a.j.hendriks@science.ru.nl
Marion	**Junghans**	EAWAG, Switzerland	marion.junghans@oekotoxzentrum.ch
Guillaume	**Kon Kam King**	Université Claude Bernard, France	guillaume.kon-kam-king@univ-lyon1.fr
Kenneth	**Leung**	University of Hong Kong, China	kmyleung@hku.hk
Ailbhe	**Macken**	NIVA, Norway	ama@niva.no
Lorraine	**Maltby**	University of Sheffield, UK	l.maltby@sheffield.ac.uk
Stuart	**Marshall**	Unilever, UK	stuart.marshall@unilever.com

Christian	**Michel**	Eawag, Switzerland	christian.michel@eawag.ch
Yuan	**Pan**	University of Sheffield, UK	ypan8@sheffield.ac.uk
Adam	**Peters**	WCA Environment, UK	adam.peters@wca-environment.com
Leo	**Posthuma**	RIVM, NL	leo.posthuma@rivm.nl
Ad	**Ragas**	Radboud University, NL	a.ragas@science.ru.nl
Sandy	**Raimondo**	US Environmental Protection Agency	raimondo.sandy@epa.gov
Hans	**Sanderson**	University of Aarhus, Denmark	hasa@dmu.dk
Keith	**Solomon**	University of Guelph, Canada	ksolomon@uoguelph.ca
Paul	**Van den Brink ***	Alterra, Wageningen University, NL	paul.vandenbrink@wur.nl
Michael	**Warne**	DSITIA, Australia	michael.warne@science.dsitia.qld.gov.au
Richard	**Wenning**	Environ, USA	rjwenning@environcorp.com
Paul	**Whitehouse**	Environment Agency, England	paul.whitehouse@environment-agency.gov.uk
Fengchang	**Wu**	Chinese Research Academy	wufengchang@vip.skleg.cn
Zhen-guang	**Yan**	Chinese Research Academy	zgyan@craes.org.cn

*Cancelled due to ill health

APPENDIX B: WORKSHOP PROGRAMME

Tuesday 11 February 2014

08:00 - 09.00	*Registration and coffee*	
09:00 - 09.10	**Welcome and introductory remarks**	Organising Committee
09:10 - 09.40	**Sense, simplicity and successes of SSDs in environmental protection, assessment and management**	Leo Posthuma RIVM, The Netherlands
	What is the ecological relevance of an SSD?	Chair: Scott Belanger P&G, USA
09:40 - 10:10	**Ecological limitations of SSDs**	Lorraine Maltby University of Sheffield, UK
10:10 - 10:40	**How do species traits influence sensitivity and herewith species sensitivity distributions?** (Cancelled due to ill health)	Paul van den Brink Alterra, The Netherlands
10:40 - 11:00	*Coffee break*	
11:00 - 11:30	**Field validation of species sensitivity distributions**	Adam Peters WCA Environment, UK
11:30 - 12:00	**Derivation of toxicity thresholds for LAS – integration of QSARs, SSDs, mesocosms, and field data**	Scott Belanger P&G, USA
12:00 – 12:30	**Field-based species sensitivity distribution and community sensitivity distribution as alternative ways for field validation of the PNECs derived from laboratory based approaches**	Kenneth Leung University of Hong Kong
12:30 - 13:30	*Lunch*	
13:30 - 15:00	**Syndicate Session 1: Ecological considerations**	Chair: Scott Belanger P&G, USA

Group:	**1A**	**1B**	**1C**	**1D**
Moderator:	L Maltby	L Posthuma	S Duquesne	K Solomon
Rapporteur:	*M Hamer*	*P Whitehouse*	*S Dyer*	*S Marshall*

- Are we making ecologically relevant assessments?

- Are regulatory protection goals explicit and clear?

- Are they set in relation to environmental quality?

- How do prospective and retrospective approaches differ?

- Are all species of equal importance, or are there keystone species that are more important than others?

- Is a generic PNEC derived from an SSD overly simplistic in terms of ecological representativeness or should we develop representative assemblages/communities (archetypes) to represent different typologies?

- Should protection goals account for local community composition?

- How does aquatic community sensitivity vary with species composition?

- How can knowledge of chemical MoA help construct SSDs for HC5 estimation?

- What are the research needs?

15:00 - 16:00 **Plenary feedback & discussion with panel**

Chair: Scott Belanger and Mick Hamer

Breakouts report back (5-10 minutes each)
Identify key points, consensus and research needs

16:00 - 16:30 *Coffee break*

What SSD statistical models are available for deriving toxic thresholds (HC5/PNEC) for aquatic communities?

Chair: Peter Craig

16:30 - 16:50 **HC5 estimation in SSDs revisited** — Tom Aldenberg, RIVM, The Netherlands

16:50 - 17:10 **Assessment factors for deriving PNECs: food for thought** — Ad Ragas, Radboud University, The Netherlands

17:10 - 17:30 **Weight of evidence approaches for deriving HC5s** — Sandrine Andres, INERIS, France

17:30 – 17:50 **Sample size in PNEC derivation** — Scott Dyer, P&G, USA

17:50 – 18:10 **How to extrapolate across 100,000+ substances, sites and species with SSDs?** — Jan Hendriks, Radboud University, The Netherlands

Close of first day

Wednesday 12 February 2014

What SSD statistical models are available for deriving toxic thresholds (HC5/PNEC) for aquatic communities?

Chair: Andy Hart
FERA, UK

09:00 - 09:30 **Interspecies correlation estimation (ICE) models predict supplemental toxicity data for SSDs** — Sandy Raimondo, US EPA, USA

09:30 - 10:00 **HC5s from taxonomically structured hierarchical species sensitivity distributions** — Peter Craig, University of Durham, UK

10:00 - 10.30 *Coffee break*

10:30 - 12:00 **Demonstration of the web-based interspecies correlation estimation (web-ICE) modelling application** — Peter Craig/ Mace Baron/Sandy Raimondo

12:00 - 13:00 *Lunch*

13:00 - 14:00 **Case studies Session** — Stuart Marshall, Mick Hamer, Scott Belanger and Peter Craig

- 2 case studies will be described and discussed using a surfactant LAS and a pesticide, chlorpyrifos.
- For each chemical, HC5s will be derived with available data using a range of SSD methods/tools.
- Different ecological scenarios will be assessed: stream, pond, marine.

14:00 - 15:30 **Syndicate Session 2: Statistical considerations**

Chair: Andy Hart
FERA, UK

Group:	**2A**	**2B**	**2C**	**2D**
Moderator:	K Leung	R Wenning	A Ragas	P Chapman
Rapporteur:	*P Craig*	*JP Gosling*	*M Barron*	*S Raimondo*

- Review current tools and key (statistical) methodology, including assumptions about distributions of sensitivity, use of hierarchical models, interspecies correlations. Identify where there are important differences and what the implications of these could be.

- As sensitivity to chemical stress seems to be related to taxonomic closeness, how could this be used in the construction and interpretation of SSDs?

- Do models based on prior knowledge provide advantages over other methods?

15:30 - 16:00 *Coffee break*

16:00 - 17:00 **Plenary: feedback & discussion with panel**

Chair: Andy Hart/Peter Craig

- Breakouts report back (5-10 minutes each)

- Identify key points, consensus and research needs

Regulatory Applications

Chair: Mace Barron
US EPA, USA

17:00 - 17:30 **Regulatory application of SSDs in European regulations** — Paul Whitehouse, Environment Agency, England

17:30 – 18:00 **Regulatory use of SSDs in Australia and New Zealand** — Michael Warne, DSITIA Science Delivery, Australia

Close of second day

Thursday 13 February 2014

Regulatory Applications

Chair: Paul Whitehouse
Environment Agency, England

08:30 - 09:00 **Use of SSD in China** — Fengchang Wu, Chinese Research Academy of Environmental Sciences

09:00 - 09:30 **Use of SSD to derive no-effect thresholds for water quality guidelines and ecological risk assessment in Canada** — Anne Gosselin, Environment Canada, Canada

09:30 - 10:00 **Use of SSDs in the USA – endangered species and water quality criteria** — Mace Barron, US EPA, USA

10:00 – 10:30 *Coffee break*

10:30 - 11:30 **SYNDICATE SESSION 3: REGULATORY CONSIDERATIONS**

Chair: Paul Whitehouse

Group:	**3A**	**3B**	**3C**	**3D**
Moderator:	A Peters	M Warne	A Gosselin	D de Zwart
Rapporteur:	*M Hamer*	*S Belanger*	*M Barron*	*A Hart*

- Would the methods reviewed in this workshop be accepted for use in regulatory assessments under current guidance? If not, what steps would be needed to facilitate their acceptance in the future? What are the opportunities to update technical guidance?

- Should current guidance on the use of SSDs be revised in the light of the issues and approaches discussed in this workshop, e.g. number of species?

- What implications are there for the interpretation of SSDs and HC5s in risk assessment and risk management?

- What are the research needs?

11:30 -12:30 **Plenary: feedback & discussion with panel**

Chair: Paul Whitehouse/Mace Barron

Breakouts report back (5-10 minutes each)
Identify key points, consensus and research needs

12:30 - 13:30 **Final Plenary discussion: synthesis of key points and research needs from the 3 sessions**

Chair: Mick Hamer/Andy Hart/Paul Whitehouse

Identify key points and consensus
What are the research needs?
Next steps

13:30 – 14:30 *Adjourn and lunch*

Close of Workshop

APPENDIX C: ORGANISING COMMITTEE

Scott Belanger (Chairman)
Procter & Gamble
Miami Valley Laboratories
Cincinnati 45253-8707, USA

Peter Craig
Durham University
South Road
Durham DH1 3LE, UK

Scott Dyer
Procter & Gamble
Miami Valley Laboratories
Cincinnati 45253-8707, USA

Malyka Galay Burgos
ECETOC
Avenue E. Van Nieuwenhuyse, 2
B - 1160 Brussels, Belgium

Mick Hamer
Syngenta
Jealotts Hill Research Station
Bracknell RG42 6EY, UK

Andy Hart
Food and Environment Research Agency
Sand Hutton
York YO41 1LZ, UK

Stuart Marshall
Unilever
Colworth Science Park
Bedford MK44 1LQ, UK

Leo Posthuma
National Institute of Public Health and the Environment (RIVM)
PO Box 1, NL-3720 BA
Bilthoven, The Netherlands

Paul Whitehouse
Environment Agency
Isis House
Howbury Park
Wallingford, Oxon OX10 8BD, UK

ECETOC PUBLISHED REPORTS

The full catalogue of ECETOC publications can be found on the ECETOC website:
http://www.ecetoc.org/publications

Responsible Editor:
Dr. Alan Poole
ECETOC AISBL
Av. E. Van Nieuwenhuyse 2 (bte. 8)
B-1160 Brussels, Belgium
VAT: BE 0418344469
www.ecetoc.org
D-2014-3001-238

Established in 1978, ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals) is Europe's leading industry association for developing and promoting top quality science in human and environmental risk assessment of chemicals. Members include the main companies with interests in the manufacture and use of chemicals, biomaterials and pharmaceuticals, and organisations active in these fields. ECETOC is the scientific forum where member company experts meet and co-operate with government and academic scientists, to evaluate and assess the available data, identify gaps in knowledge and recommend research, and publish critical reviews on the ecotoxicology and toxicology of chemicals, biomaterials and pharmaceuticals.