EnviroTox Platform

EnviroTox Database
PNEC Calculator
ecoTTC
Chemical Toxicity Distribution (CTD)

USER GUIDE

October 2021

EnviroTox Database Team

User Guide Authors

Amy Beasley (The Dow Chemical Company)
Scott Belanger (Procter & Gamble)
Kristin Connors (Procter & Gamble)
Michelle Embry (HESI)
Ryan Otter (Middle Tennessee State University)

Project Team

Mace Barron (USEPA)
Adriana Bejarano (Shell)
Mark Bonnell (ECCC)
Jessica Brill (Procter & Gamble)
Kristin Connors (Procter & Gamble)
Dick DeZwart (DdZ Ecotox)
Kellie Fay (USEPA)
Robert Hoke (DuPont)
Sarah Hughes (Shell)
Cristina Inglis (ECCC)

Jessie Krailler (Procter & Gamble)
Connie Mitchell (HESI)
Teresa Norberg-King (USEPA)
Ryan Otter (MTSU)
Joshua Phillips (MTSU)
Raechel Puglisi (HESI)
Alan Samel (FMC)
Hans Sanderson (Aarhus University)
Peter Wilson (Sanofi)

For further information, contact:

Health and Environmental Sciences Institute (HESI)
740 15th Street, NW, Suite 600
Washington, DC 20005 USA
Email: membry@hesiglobal.org
Website: www.hesiglobal.org

Acknowledgements

HESI Animal Alternatives in ERA Committee

This user guide and associated EnviroTox Database website and tools are a work product of the Health and Environmental Sciences Institute (HESI) Animal Alternatives in Environmental Risk Assessment Technical Committee. The Committee's mission is:

> *To ensure the development of a sound technical basis for alternative test methods as a means to reduce, refine, or replace standard ecotoxicity test procedures around the globe.*

> *The committee aims to provide a forum to coordinate the debates and best emerging practices of the alternatives and animal model development sciences to meet existing hazard assessment, effluent assessment, risk assessment, classification and labeling, and other regulatory needs.*

Information regarding the Technical Committee and HESI can be found at http://hesiglobal.org/. This committee is one of several active programs managed by a tripartite consortium of members representing academia, government, and industry in a public-private partnership with global representation.

Supporting Organizations

The following organizations have provided support for this project since conceptualized (both direct financial resources and in-kind):

Chevron
Environment and Climate Change Canada
European Commission, Joint Research Centre
ExxonMobil
FMC Corporation
L'Oréal Corporation
Middle Tennessee State University
Procter & Gamble Company
Research Institute for Fragrance Materials (RIFM)
SC Johnson

Sanofi
Shell Chemicals, Ltd.
Texas Christian University
The Dow Chemical Company
UK Home Office
University of Aarhus
University of Bern
University of Guelph
University of Heidelberg
University of Miami, Ohio
US Environmental Protection Agency

2017 Workshop Participants

We would like to thank the participants from the September 2017 International Workshop on "Establishment of the eco-TTC approach for environmental risk assessment of chemicals" for their valuable insight and expertise to improve this effort.

Disclaimer

The Health and Environmental Sciences Institute (HESI) and all associated collaborators do not guarantee, warrant, or make representations, either expressed or implied, regarding the use, or the results of the use of the materials provided with regards to reliability, accuracy, correctness, or otherwise. There are no warranty rights granted to users of EnviroTox Database or the associated analysis tools. Users assume the entire risk as to the results and performance of the tools and database. HESI and all associate collaborators are not liable under any circumstances, for any damages whatsoever, arising out of the use, or the inability to use, the tools and database provided, even if advised of the possibility of such damages.

Table of Contents

List of Figures

List of Tables

Terminology / Acronyms

AiiDA	Aquatic Impact Indicator Database
CAS	Chemical abstracts service
CTD	Chemical toxicity distribution
DNEL	Derived negligible effect level
EC_{50}	Half maximal effective concentration
ECCC	Environment and Climate Change Canada
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ecoTTC	Ecological threshold of toxicological concern
FET	Fish embryo test
HC_5	5th percentile hazardous concentration
HESI	Health and Environmental Sciences Institute
HPV	High Production Volume
ICCA	International Council of Chemical Associations
METI	Japanese Ministry of the Environment
MOA	Mode of action
NAM	New approach methodology
OECD	Organization for Economic Cooperation and Development
PNEC	Predicted no effect concentration
QSAR	Quantitative structure activity relationship
REACH	Registration, evaluation, and authorization of chemicals
SSD	Species sensitivity distribution
TTC	Threshold of toxicological concern
USEPA	United States Environmental Protection Agency
USFDA	United States Food and Drug Administration
USGS	United States Geological Survey

Preface

The information described in the User's Guide can also be found at: www.envirotoxdatabase.org.
- The EnviroTox Database: development, content, structure, and search functions
- The PNEC calculator tool
- The ecoTTC distribution tool
- The CTD tool

Publications

The following publications are available (all Open Access) directly related to the EnviroTox Database and associated tools:

- Barron MG, Otter RR, Connors KA, Kienzler A, Embry MR. (2021). Ecological thresholds of toxicological concern: A review. *Frontiers in Toxicology* 3(7); 10.3389/ftox.2021.640183.

- Belanger SE, Beasley A, Brill JL, Krailler J, Connors KA, Carr GJ, Embry MR, Barron MG, Otter R, Kienzler A. (2021). Comparisons of PNEC derivation logic flows under example regulatory schemes and implications for ecoTTC. *Regulatory Toxicology & Pharmacology.* 123: 104933.

- Belanger SE, Sanderson H, Embry MR, Coady K, DeZwart D, Farr BA, Gutsell S, Halder M, Sternberg R, Wilson P. (2015). It is time to develop ecological thresholds of toxicological concern to assist environmental hazard assessment. *Environmental Toxicology & Chemistry* 34: 2864-2869.

- Brill JL, Belanger SE, Barron MG, Beasley A, Connors KA, Embry M, Carr GJ. 2021. Derivation of algal acute to chronic ratios for use in chemical toxicity extrapolations. *Chemosphere* 263: 127804.

- Connors KA, Beasley A, Barron MG, Belanger SE, Bonnell M, Brill JL, deZwart D, Kienzler A, Kraller J, Otter R, Phillips JL, Embry MR. (2019). Creation of a curated aquatic toxicology database: EnviroTox. *Environmental Toxicology & Chemistry* 38(5): 1062-1073.

- Kienzler A, Connors KA, Bonnell M, Barron MG, Beasley A, Inglis CG, Norberg-King TJ, Martin T, Sanderson H, Vallotton N, Wilson P, Embry MR. (2019). Mode of action (MOA) classifications in the EnviroTox database: Development and implementation of a consensus MOA classification. *Environmental Toxicology & Chemistry* 38(10): 2294-2304.

- Kienzler A, Bopp S, Halder M, Embry M, Worth A. (2019). Application of new statistical distribution approaches for environmental risk assessment: A case study. *Science of the Total Environment* 693: 133510.

- Kienzler A, Barron MG, Belanger SE, Beasley A, Embry MR. (2017). Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. *Environ Sci Technol.* 51:10203-10211.

Introduction

Need for alternative approaches

The need for rapid, predictive methods to address aquatic ecological hazards of diverse substances remains essential, as the chemical universe remains largely untested. Flexible approaches that do not require the use of large numbers of vertebrate test animals (fish, amphibians, birds, etc.) are needed to address broad animal welfare concerns.   To appropriately develop new approaches methodologies (NAMs) and non-testing approaches, existing information must be made available via integrated and curated datasets. Increasing regulatory requirements have laid the foundation for the development of more standardized and extensive data sets for a broader range of chemicals. Regulatory programs such as REACH; (EC, 2007), ICCA (International Council of Chemical Association) High Production Volume

(HPV), Chemicals Challenge (ICCA, 2018), and Canada's Domestic Substance List (ECCC, 2018) have helped to create unprecedented levels of available toxicity data, along with continuing investigations of hazards of substances to aquatic life published in the scientific literature.

ecoTTC concept

Risk assessment of chemicals inherently involves an assessment of toxicity, exposure and the resulting likelihood or probability of observing an adverse response. Further, it requires ethical and resource consideration as to how much data is attainable and should be derived (e.g., via use of animal testing) versus what is considered an acceptable level of extrapolation (Belanger et al., 2015). A recent addition to the methodology is the concept of the Threshold for Toxicological Concern, or TTC. The TTC establishes an exposure level for chemicals, below which no appreciable risk to human health or the environment is expected based upon a de minimis value for toxicity identified for many chemicals (USFDA, 1995;Kroes et al., 2004). This level can then be compared to an estimate of the likely exposure to a chemical to complete a screening level safety assessment for a given route of exposure or environmental compartment/species of concern. The TTC concept is well-established for assessing human safety of indirect food-contact substances and has been reapplied for a variety of endpoints including carcinogenicity, teratogenicity, and reproductive toxicity. TTCs have benefits for screening-level risk assessments, including the potential for rapid decision-making, fully utilizing existing knowledge, reasonable conservativeness for chemicals used in lower volumes, and reduction or elimination of unnecessary animal tests.

TTC approaches have only recently been explored in environmental assessments. It is not the intent to review the literature here for the purpose of this User's Guide, but assessors can refer to a series of papers for background information (de Wolf et al., 2005;Gross et al., 2010;Williams et al., 2011;Hendriks et al., 2013;Mons et al., 2013;Gutsell et al., 2015). Examples can be found therein that are varied in approach, but which are near derivatives of the concept termed ecological Thresholds of Toxicologic Concern, or ecoTTCs, as used here. Aspects of information management, tracking of test species, use of categorization principles, and applications of statistical analyses to identify thresholds are common among the historical efforts (see Table 1 of Belanger et al., 2015). Further discussion of the ecoTTC approach and its application can be found in the publications listed above (Kienzler et al., 2017a;Connors et al., 2019;Kienzler et al., 2019a;Kienzler et al., 2019b;Barron et al., 2021;Belanger et al., 2021;Brill et al., 2021).

EcoTTCs summarize the distribution of a large array of species level toxicity data as ecosystem PNECs (Predicted No Effect Concentrations). PNECs are derived for a chemical category or mode of action and project a conservative prediction for similar, but untested chemicals. In human safety, the TTCs are typically set as a 5th percentile value DNEL (Derived Negligible Effect Level) from a statistical distribution of similarly acting chemicals. EcoTTCs are defined here as the 5th percentile value derived from a statistical distribution of PNECs of similarly acting chemicals. The categorization of "similar acting" compounds can be based on Mode of Action (*sensu* Kienzler et al. 2017), a formed chemical category, or for a functional use of a chemical (e.g., pharmaceutical, detergent surfactant, etc.). Because PNECs are regionally based in their development, reflecting local attitudes around hazard conservatism and local Application Factors, PNEC distributions are expected to vary by region (Hahn et al., 2014). There may be situations where a researcher or assessor is interested in the toxicity distributions, without the added conservatism introduced by the assessment factors or the regional overtones of their application. A

chemical toxicity distribution (CTD) can be used to perform this type of analysis, and a tool to evaluate this has been added as a component of this work.

EnviroTox Platform – Overview

Architecture

The overall structure of the database and tools housed in the EnviroTox platform, along with how this information is to be used, is depicted in **Error! Reference source not found.**. Three broad categories of data are housed within the EnviroTox database and include physical-chemical information (including mode of action, or MoA, assignments), information on test species, and lastly ecotoxicological information. The database is subject to query using the application interface and is described below. Queries are constructed based on the user's specific questions and interest. An output file containing data that matches the user query and can be subjected to further analyses both ad hoc, outside the application, and within the application. The content of files, query, and outputs will be described in greater detail below.
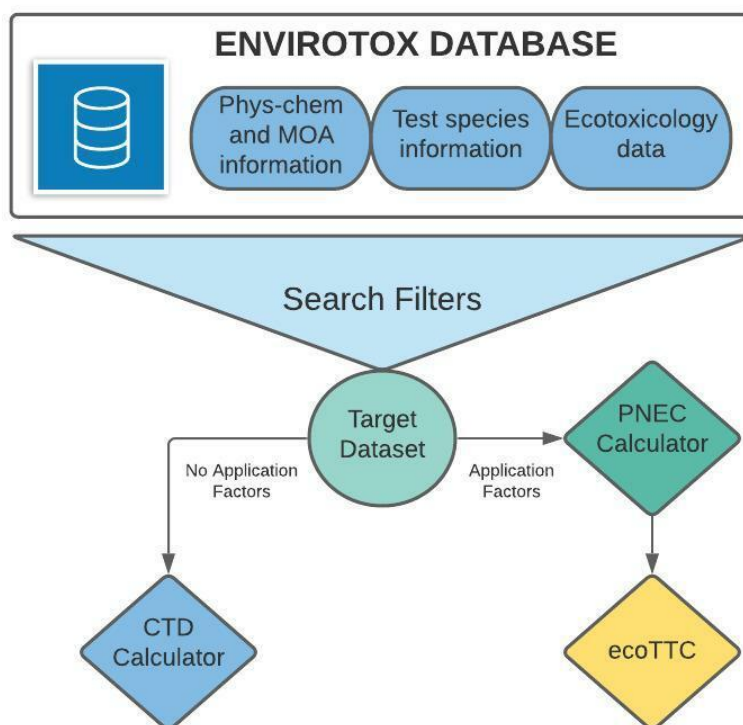

Figure 1. Overall architecture of the EnviroTox web platform

In the following sections, the tabs of the web tool and their various functionalities are described. A schematic overview is provided in **Error! Reference source not found.**. The user can search the database to construct and customize the base ecotoxicological information from which an ecoTTC could be calculated. Analysis includes choosing a geographic region for deriving chemical specific PNECs (different regions use different processes). Once PNECs are derived the ecoTTC calculation can proceed with additional customization that is also user-defined such as including or excluding PNECs that are supported with less or more ecotoxicological data. Chemicals with less data are extrapolated to

the PNEC with larger uncertainty or application factors than those with more complete information. Finally, once the PNEC distributions are determined, the outputs are provided in EXCEL and graphical formats. At this point users can do further analyses or evaluate outputs directly for the purposes that were chosen. This is a highly genericized diagram of information flow, and many others are possible. There may be situations where a researcher or assessor is interested in the toxicity distributions, without the added conservatism introduced by the assessment factors or the regional overtones of their application. A chemical toxicity distribution (CTD) can be used to perform this type of analysis.
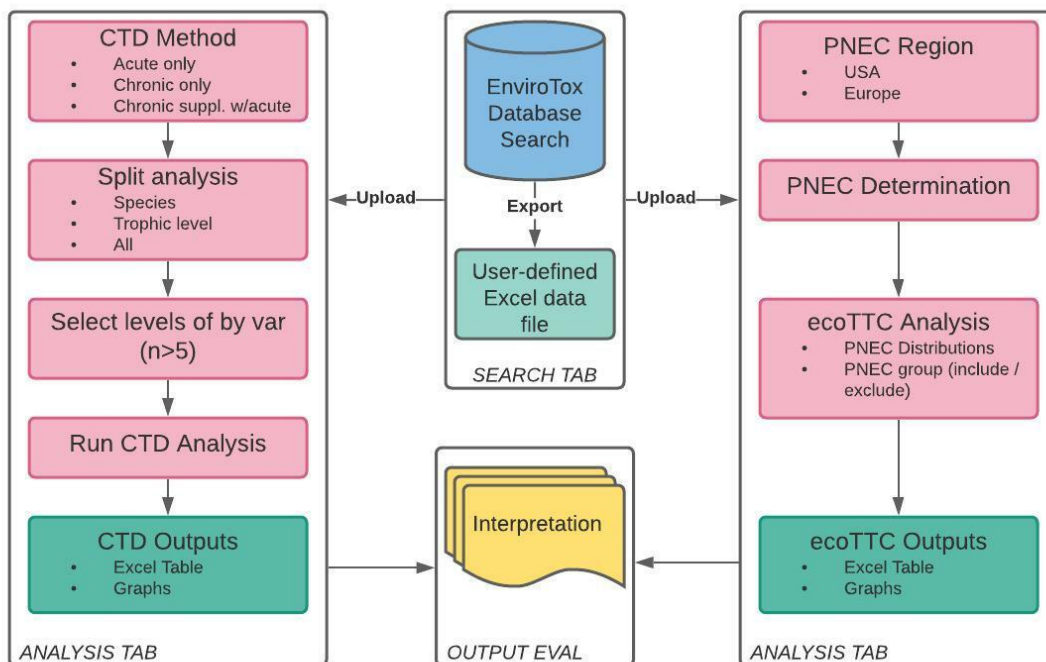

Figure 2. Schematic of EnviroTox platform tools and functionality

EnviroTox Database

Summary

Version 2.0.0 of the EnviroTox database contains 80,912 records representing 1,641 species, and 4,267 unique chemical CAS. This latest update was performed in September 2021 and provides updated records from the USEPA's Ecotox Database through December 2020. Chemical-specific information is also linked to each record and includes physical chemical information, chemical descriptors, and mode of action classifications (Kienzler et al., 2017b). Taxonomic descriptions of test species (phylogeny, trophic level, etc.) are also included and all records include the original source citations. Toxicity data is associated with the physical chemistry data, mode of action classifications, and curated taxonomic information for the organisms tested. The database also includes a systematic process for including acute and chronic effects, as well as computing a predicted no effect concentration (PNEC) for exposed ecosystems based on depth and breadth of data included in the statistical computation. Additional discussion of the development of this database is available in (Connors et al., 2019). Information included in the database is included in the sections below.

Data collection

Ecotoxicological information was gathered from a wide variety of sources listed in **Error! Reference source not found.** (discussed in Connors et al., 2019). Information was compiled by associating individual CAS numbers with ecotoxicological data. Each individual data point within a study was considered as a separate entity. The potential to include a data point or study was based on the SIFT methodology where predefined inclusion criteria is used to address relevance, validity and acceptability of data (Beasley et al., 2015) (**Error! Reference source not found.**).

Table 1. Sources of aquatic ecotoxicology data for the EnviroTox Database

Data source	Description
ECHA (REACH)	Obtained by query of the REACH data from eChemPortal database of publicly available substance data, submitted to ECHA (European Chemicals Agency) under the REACH (Registration, Evaluation and Authorisation of Chemicals) regulations. (OECD 2018). http://echemportal.org
USEPA ECOTOX	Obtained by query of the USEPA's ECOTOX Knowledgebase, including EPA-generated test data and data from the public literature. (USEPA 2018a). https://cfpub.epa.gov
Peer-reviewed literature	Original dataset foundational to Species Sensitivity Distribution work by (De Zwart, 2002) and colleagues, personal communication to the HESI Technical Committee, containing data and metadata stripped from peer-reviewed literature
ECETOC OASIS	Aquatic toxicity results from a variety of sources, available via the OECD QSAR Toolbox. http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm
AiiDA	Aquatic Impact Indicator DAtabase; contains data sourced from ECHA, ECOTOX and others. Queried to supplement for data not found in REACH. aiida.tools4env.com
METI	Summary of aquatic toxicity test results from OECD guideline tests conducted by the Japanese Ministry of the Environment (METI). Some data publicly available via the OECD Toolbox. Also known as the NITE-CHRIP database. http://www.nite.go.jp/en/chem/chrip/chrip_search/srhInput
FET	Dataset of acute aquatic toxicity test results from the OECD validation study to evaluate the reproducibility of the Zebrafish Embryo Test (ZFET). (Belanger et al. 2013; Busquet et al. 2014)
USGS Columbia	Summary dataset of acute aquatic toxicity tests conducted by the USGS Columbia Environmental Research Center, including (Mayer and R. Ellersieck, 1986). http://www.cerc.usgs.gov
Pharmaceuticals	Summary of acute and chronic aquatic toxicity data for active pharmaceutical ingredients. Provided by Sanofi S.A. and detailed in (Vestel et al., 2016).
ECOSAR training set	Set of aquatic toxicity data used to train the computational QSAR tool ECOSAR (ECOlogical Structure Activity Relationship) developed by the U.S. Environmental Protection Agency for hazard estimation; sourced from the Help files for the ECOSAR program (USEPA 2012) https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model
EPA Pesticide Data	Pesticide Ecotoxicity Database (formerly the Ecological Effects Database); aquatic toxicity data provided by the U.S. Environmental Protection Agency's Office of Pesticide Programs Environmental Fate and Effects Division. http://www.epa.gov/pesticides
OECD QSAR Toolbox	Queried to supplement aquatic toxicity data from ECOTOX and ECHA. Contents include data from Aquatic ECETOC and Aquatic Japan MoE. http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm

Table 2. SIFT criteria used to ascertain inclusion of ecotoxicological data in the EnviroTox database

Step	Criteria	Specifics
0: Purpose	Aquatic toxicity data and metadata	Initial pull of available information from databases listed in Table 1.
1: Relevance	Trophic designations	Fish, amphibian, invertebrate, algae
2: Validity	CAS	CAS present
	Required fields	Effect value/units, duration, test statistic, effect measured, source present
	Qualifiers	Exclude effect values with qualifiers (e.g., <>)
	Effect	Specific effect measurement (e.g., EC50)
3: Acceptability	Duration	≥ 24h
	Test Statistic	≥5% and ≤70% effect measure (e.g., IC10, LC50), N OEC, LOEC, MATC
	Effect	Abundance, biomass, cells, chlorophyll, emergence, filtration rate, gross primary productivity, growth, hatchability, intoxication, mortality, nitrogen fixation, population growth, population reduction, population change, primary production, regeneration, reproduction, shell deposition, teratogenesis Focus is on endpoints of regulatory significance and known use in decision-making
4: Additional criteria	CAS, Chemical name, SMILES	Harmonized. Database trimmed to only contain validated chemicals
	Metals	Inorganic compounds were collapsed to a 'dummy metal ion' CAS
	Identification of duplicates	Removed records that were full duplicates (e.g., Citation, species, test duration, test statistics, measured effect, effect level)
	Removal of extreme outliers	Solubility, effect concentrations

Physical-Chemical Information

Several compound identifiers were included in the database as shown in

. Information on specific chemicals is associated with Chemical Abstracts Service registry number (CAS). CAS numbers in the database are absent spaces or dashes. Chemical CAS, chemical name, and SMILES were systematically verified. This process involved first running all CAS through the USEPA Chemistry Dashboard (comptox.epa.gov/dashboard); if chemicals had a CAS and chemical name match through this tool, they were considered validated, and the corresponding SMILES was extracted. For those chemicals where there was not a match through the USEPA Chemistry Dashboard, the chemical CAS was run through SciFinder and checked against several chemical identification tools to determine the name and CAS, then the SMILES were extracted.

Table 3. Description of information included in the physical-chemical file for the EnviroTox database

CAS	Chemical Abstracts Service (CAS) number, no dashes, or spaces
Chemical descriptors	
Chemical name	Commonly employed chemical name.
SMILES	Unified SMILES (Simplified Molecular Input Line Entry Specification) code associated with the chemical and CAS.
Desalted canonical SMILES	Open Babel (Open Babel 2018) was used to generate desalted and canonicalized SMILES for subsequent modeling and chemical categorization.
Molecular weight	Molecular weight in g/mol; generated from desalted SMILES using EpiSuite DermWin (USEPA 2018b)
Log K_{ow}	Octanol-water partition coefficient; unitless; EpiSuite KOWWIN (USEPA 2018b) used to populate Log K_{ow} from desalted SMILES. Experimental used if available; modeled if no experimental available
Water Solubility	Solubility of the chemical in pure water (25°C, 1 atmosphere) in mg/L; EpiSuite WSKOW (USEPA 2018b) used to populate water solubility from desalted canonical SMILES. Experimental used if available; modeled if no experimental available. Effect values that are greater than 5x of the water solubility level were flagged but not removed.
ECOSAR Classification	Assignment of chemical class based on desalted, canonical SMILES input to OECD QSAR Toolbox ((https://www.qsartoolbox.org/home). Note that compounds may be assigned to multiple ECOSAR groups depending on types of substitutions.
ECOSAR Classification – collapsed	For chemicals where multiple classifications were generated by ECOSAR, the first reported was used. These categories were further collapsed into 46 more general categories. The complete list of ECOSAR classification collapsed assignments is available as Supplementary Information
USEPA New Chemical Categories	Original categories cited in the document "TSCA New Chemicals Program (NCP)/ Chemical Categories" (USEPA 2010).
MOA Classifications	
Verhaar	Verhaar classes obtained via OECD QSAR Toolbox (https://www.qsartoolbox.org/home).
TEST	Toxicity Estimation Software Tool (TEST) based on the MOAtox broad assignments as described by (Barron et al., 2015).
OASIS	OASIS acute aquatic toxicity MOA obtained via OECD QSAR Toolbox. (https://www.qsartoolbox.org/home)
ASTER	ASTER (ASsessment Tool for Evaluating Risk) is a rule-based expert system and is operated on a proprietary basis by US EPA based on the MOA categories in (Russom et al., 1997).
Consensus MOA assignment	A 'consensus' MOA assignment of narcotic (N), specifically-acting (S), or unknown (U) was assigned to each chemical based on a consensus from the 4 classification schemes.
Chemical Categories	Determined from SMILES
Halogenated	Contains F, Cl, Br, I.
Heavy Metal	Contains a heavy metal (metallic element with a density greater than 5)

All the listed parameters can be employed in various ways to query the available information.

Taxa Descriptions
Information shown in

 has been collated for species present in the database. Current taxonomic status has been harmonized as of 2017. It is recognized that some designations can be somewhat arbitrary. For example, functionally photosynthetic/non-photosynthetic protists may be categorized as algae (photosynthetic microbes) when they may have been tested in a state absent chloroplast in some situation(s). Common authoritative taxonomic websites including https://www.ncbi.nlm.nih.gov, http://www.algaebase.org/, http://www.marinespecies.org/about.php, and http://fishbase.org/home.htm were consulted to derive final classifications. A few ecotoxicologically important species are among these and ecotoxicologists should recognize their transitions to new names: *Pseudokirchneriella subcapitata* (formerly *Selenastrum capricornutum*), *Desmodesmus subspicatus* (formerly *Scenedesmus subspicatus*), *Danio rerio* (formerly *Brachydanio rerio*), *Oncorhynchus mykiss* (formerly *Salmo gairdneri*) and *Americamysis bahia* (formerly *Mysidopsis bahia*) have all undergone taxonomic revision in recent years. All species identified in the database were also assigned to a freshwater or saltwater habitat. For estuarine or facultatively freshwater to saltwater species, the primary habitat in which they are known or tested in was used to assign habitat. A species file is maintained in the database and will be updated as needed.

Table 4. Information included in the standalone (non-interactive) species file for taxa found in the EnviroTox database

Latin Name	**Linnaean Genus and species name**
Trophic level	Algae, Invertebrate, Fish, Amphibian, Macrophyte, Fungi
Taxonomic Kingdom	Consensus based designation
Taxonomic Phylum or Division	Phylum (animal) or Division (plant)
Taxonomic Sub-phylum	Not always available
Taxonomic Superclass	Not always available
Taxonomic Class	Taxonomic Class
Taxonomic Order	Taxonomic Order
Taxonomic Family	Taxonomic Family

It should be noted here, that ecotoxicological test performed on mixed communities of organisms (more than one taxon, well described at the genus level) are not included in the database. Tests performed on organisms designated above the genus level (i.e., Family, or higher) are not included. Note that the tests are identified in the database but are excluded from consideration due to the structure of queries and the SIFT process (Beasley et al. 2015, also see below).

Toxicity endpoints
Central to decision-making in the EnviroTox database were decisions on endpoints to include and how to ascertain if the study was an acute test or chronic test. Endpoints for ecotoxicity studies were evaluated for their utility in regulatory evaluations of ecotoxicity data (Moermond et al., 2017;Rudén et al., 2017). Further, the endpoint was then associated with appropriate statistical evaluations to arrive at a conclusion of "acute" or "chronic" toxicity or unassignable. As an illustrative example, a study on the ecotoxicity of a chemical to *Daphnia magna* (an accepted cladoceran) was performed over 17 days and response to a biomarker was measured and positioned as a No-observed effect-concentration (NOEC).

While the species, duration and statistic may be appropriate for a "chronic" interpretation, the biomarker is not presently used in any regulatory framework for environmental risk assessment so it would not be further used. Decision logics were established for studies on all taxonomic groups so that transparency for assignment as acute or chronic determinations based on endpoint, species, and statistic that were operationally defined. The logic used to classify non-photosynthetic microinvertebrates, macroinvertebrates, amphibians, fish, and macrophytes as acute or chronic are included in Belanger et al., 2021 as Supplemental Information.

Mode of Action Information

- In order to allow grouping of chemicals for eventual ecoTTC or other analyses, four mode of action classification assignment schemes (Verhaar, TEST, OASIS, and ASTER) were applied to each chemical. Details of this work can be found in Kienzler et al., 2017 and 2019. The specific MOA assignments obtained from each of the schemes are included in the database. However, to facilitate simpler groupings for ecoTTC and other applications, each chemical was assigned to one of three 'general' groupings based on the degree of consensus between the evaluated schemes and the concordance between schemes shown in

Metals

The toxicity of some metal-containing compounds can be driven by the presence of the freely dissolved metal ion. Consistent with the U.S. derivation of water quality criteria for aquatic life and international screening values (Barron and Wharton, 2005), specific divalent metal compounds were grouped by metal ion. Inorganic compounds were assigned to a 'dummy metal ion CAS' (e.g., Metalgrp.Ag') 'if 1) the metal ion could dissociate from the compound (e.g., acetate, lactate), 2) the toxicity of the compound would be driven by the metal ion. A compound was not assigned to a dummy metal CAS if it was caustic or highly reactive, if the metal was associated with ammonia or hydroxides, or if more than one metal was present in the compound. A total of 140 compounds in the database were assigned one of 24 different 'dummy metal ion CAS'. The original CAS for the compound and dummy metal ion CAS are both provided in the database.

Salts

Chemical compounds were excluded from the database if the desalted canonicalized SMILES resulted in the individual hydroxide (OH-), chloride (Cl-), ammonium (NH4+) or amino cation (NH2+). The corresponding effects data for excluded compounds were removed from the database because of uncertainty regarding the moiety that would produce the toxicological effect.

: narcotic (N), specifically acting (S) or Unclassified (U).

Each MOA scheme assignment was collapsed into one of the three bins as assigned in Table 5 below. A four-letter code, corresponding to the TEST, ASTER, OASIS, and Verhaar bins, respectively, was assigned to each chemical, a consensus MOA was determined, and a confidence score was assigned.

- All four in agreement (e.g., NNNN, SSSS, UUUU): Confidence score of 3
- Three schemes in agreement (e.g., NNNS, SSNS): Confidence score of 2
- Two schemes in agreement with the other two as "U" (e.g., NNUU, SUSU): Confidence score of 1 and assignment made on the non- "U" assignment
- All other combinations: assigned a consensus MOA of "U" and a confidence score of 0

Metals

The toxicity of some metal-containing compounds can be driven by the presence of the freely dissolved metal ion. Consistent with the U.S. derivation of water quality criteria for aquatic life and international screening values (Barron and Wharton, 2005), specific divalent metal compounds were grouped by metal ion. Inorganic compounds were assigned to a 'dummy metal ion CAS' (e.g., Metalgrp.Ag') 'if 1) the metal ion could dissociate from the compound (e.g., acetate, lactate), 2) the toxicity of the compound would be driven by the metal ion. A compound was not assigned to a dummy metal CAS if it was caustic or highly reactive, if the metal was associated with ammonia or hydroxides, or if more than one metal was present in the compound. A total of 140 compounds in the database were assigned one of 24 different 'dummy metal ion CAS'. The original CAS for the compound and dummy metal ion CAS are both provided in the database.

Salts

Chemical compounds were excluded from the database if the desalted canonicalized SMILES resulted in the individual hydroxide (OH-), chloride (Cl-), ammonium (NH4+) or amino cation (NH2+). The corresponding effects data for excluded compounds were removed from the database because of uncertainty regarding the moiety that would produce the toxicological effect.

Table 5. MOA classification table and concordance "Bins"

VERHAAR		ASTER		TEST		OASIS	
Classification	Bin	Classification	Bin	Classification	Bin	Classification	Bin
• Class 1 (narcosis or baseline toxicity) • Class 2 (less inert compounds)	N	• Non-polar narcosis • Polar narcosis • Ester narcosis	N	Narcosis	N	• Basesurface narcotics • Narcotic amines • Phenols and anilines • Alpha, beta-unsaturated alcohols • Esters	N
• Class 3 (unspecific reactivity) • Class 4 (compounds and groups of compounds acting by a specific mechanism)	S	• Diester toxicity • Reactive • Chloro-diester-based reactivity • Carbonyl (C=0)-based reactivity • Carbonyl reactivity • Alkylation/arylation-based reactivity • Acylation-based reactivity • Sulfhydryl (-S-H)-based reactivity • Reactive dinitroaromatic group • Nitroso-based reactivity • Quinoline reactivity • Acetamidophenol reactivity • Reactive diketones • Acrylate toxicity • N-halogenated acetophenone inhibition • Hydrazine-based reactivity • Isocyanate (-N=C=O)-based reactivity • Pyridinium compounds • Neurotoxicant: DDT-type • Neurotoxicant: pyrethroid • Neurotoxicant: cyclodiene-type • Neurotoxicant: strychnine • Neurotoxicant: nicotine • Organophosphate-mediated AChE inhibition • Carbamate-mediated AChE inhibition • Uncoupler of oxidative phosphorylation • Respiratory blocker: azides and cyanides	S	• Reactivity • Neurotoxicity • AChE inhibition • Electron transport inhibition • Iono/osmoregulatory / circulatory impairment	S	• Reactive unspecified • Aldehyde	S
Class 5	U	Unknown mode of action	U	Unknown	U	Unknown	U

EnviroTox Platform: Database searching & Tools

Access

The EnviroTox platform, which includes the database, search interface, and tools, can be accessed at www.envirotoxdatabase.org.   Upon agreement with HESI, users agree to appropriately utilize the tools and data. Primary and derivative works where the ecoTTC database and/or its calculation functions have been used for a scientific purpose should be cited when used. Information on how to cite use of the database and tools is available under the "About" tab on the website.

Database searching

On the top of the homepage the five clickable tabs of the ecoTTC web applications are found: Search, Analysis, Setup, Documentation and About. Two options are available to search the ecoTTC database: a "General Search" (Figure 3) and an "Advanced Search" (Figure 4). Both options results in the user being able to export their desired data as an Excel file, which will then be uploaded into the analysis tools. Additional search terms are available using the Advanced Tab.

General Search Fields	Advanced Search Fields (include those of general search PLUS)
Substance properties • CAS • Chemical Name • Desalted Canonical SMILES • Log K_{ow} • Water Solubility (mg/L) • MW (g/mol) • ECOSAR classification • US-EPA New Chemical Categories • Consensus MOA **Taxonomy Properties** • Latin name • Trophic Level **Test Properties** • Test type • Test statistic • Duration (A = Acute; C= Chronic)	**Substance Properties** • Canonical SMILES • Heavy Metals • Halogenated • Desalted Canonical SMILES • Log K_{ow} • TEST • ASTER • OASIS • Actual Verhaar Category **Test Properties** • Test type • Test statistic • Duration (days) • Duration (hours) • Effect is 5x above water solubility

General Search Fields:
Begin either type of search by clicking on the "Select a Field (optional)" dropdown menu and selecting a field to filter the dataset. Next, click the "contains" dropdown menu and select the qualifier you want to use for your chosen field (e.g., contains, =, <). Finally, begin typing in the "Search for…" how you would like to filter the data based on the field chosen (this could be letters or numbers, depending on the field chosen).

If you would like to filter the data by more than one field, then click the "+" box immediately to the right of the "Search for…" box to add another filter Field. All additional filter fields require

you to specify the desired Boolean operator (AND or OR) for how you wish to link the search filters together.
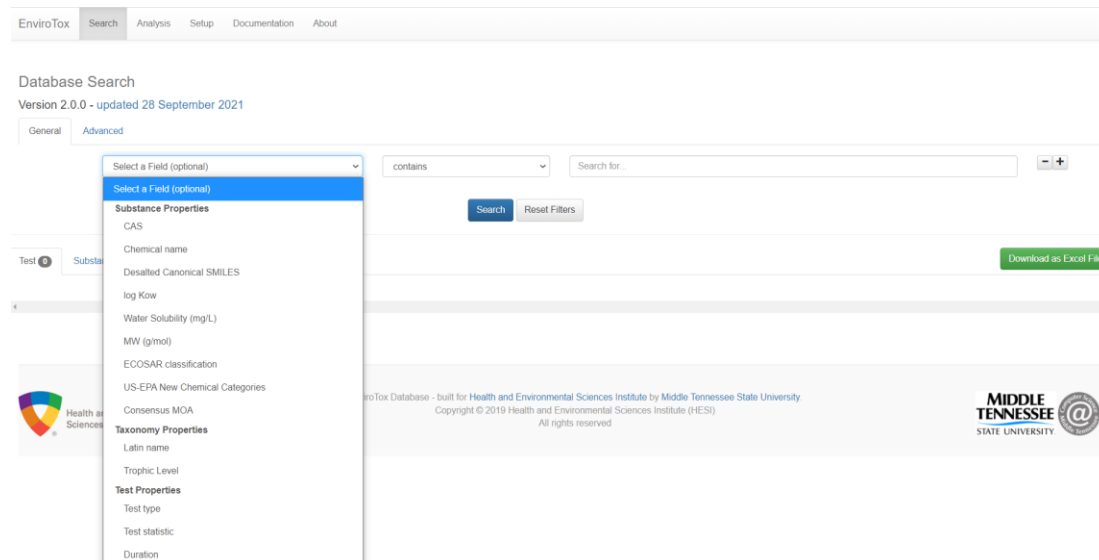
Figure 3. General search
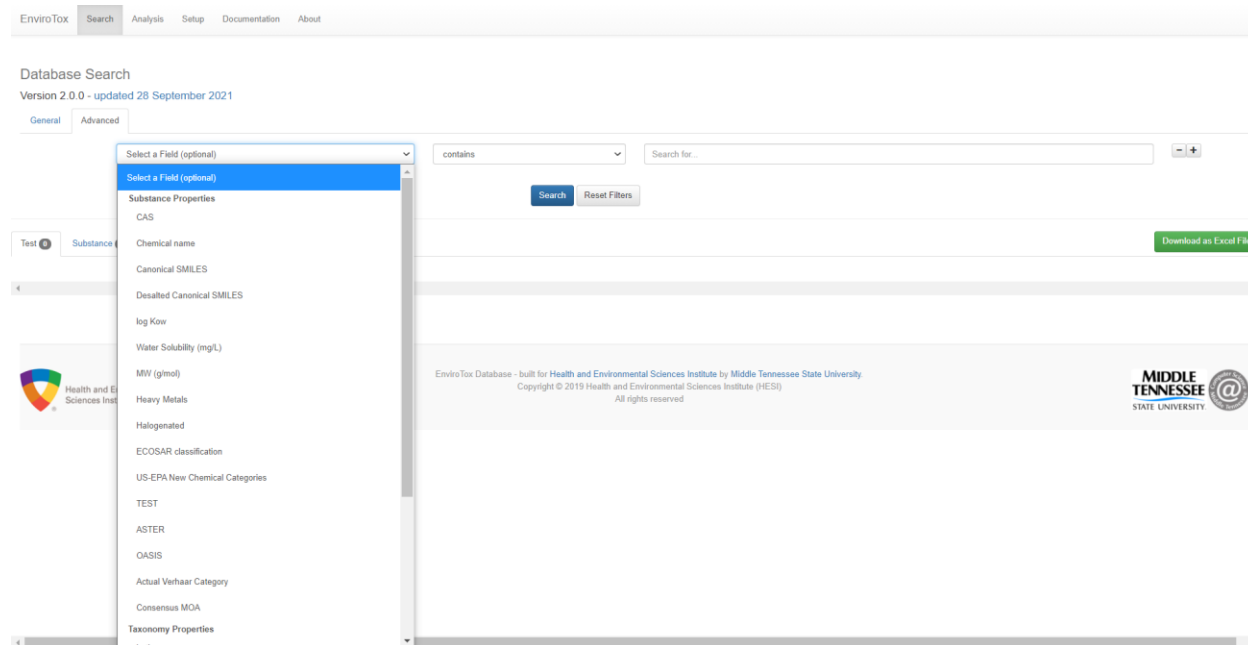
Figure 4. Advanced search

Once all filter Fields have been completed, click the blue "Search" button and a snapshot of the resulting data will appear with three tabs running across the top of the data that indicate the number of studies (Test), chemicals (Substances), and Species (Taxonomy) in the filtered dataset (Figure 5). If the search performed has returned the desired information, then click the

green button on the "Download as Excel File" right side of the "Search for…" box to export the data file. You will want to save this to your local drive in a place easy to locate, as this file will serve as the input for the analysis tools. If the search performed has not returned the desired dataset, then you can either modify your previous search or click the "Reset Filters" button and start the advanced search over.

Figure 5. Summary information provided after a search

IMPORTANT NOTE: The web tool saves no searches, once filters have been reset; there is no way to retrieve the search parameters used.

The database query system is sensitive to characters entered in the "Search for…" field (i.e., it employs smart search technology). In this case, the user is beginning to search for ecotoxicity data associated with non-polar narcosis that have been identified using ASTER. The more characters added, the narrower the search becomes. The user can scroll into the dropdown box and highlight the group desired at any point. Refinement of the search can continue by adding additional search terms with the + sign or remove search terms with the – sign at the right of the "Search for…" field. For example, if the user desired to narrow the data to be analyzed to those non-polar compounds that have log K_{ow} values between 3 and 5, these could appear as additional search terms.

At times, it may be desirable to perform a search that includes multiple identifiers for a selected search category. For example, a search for multiple CASNOs or multiple ECOSAR categories. A refined search functionality is provided to perform this action using the "=" sign within the dropdown field containing Boolean descriptors. After selecting the "equals sign" the user can enter multiple terms for the category into the "Search for…" field. Terms should be separated

by a space backwards slash. This function is particularly useful in the search for multiple CASNOs, as an example.

Please note that direct entry by the user into the "Search for…" box is fully enabled for CASNOs. Other Search categories, such as taxonomic names, mode of action assignments, and so forth need to be entered by first developing the search string in a text editor (Word, WordPad, Notepad, etc.) and pasting the string into the "Search for…" field. The "=" term requires an exact match to the entry in the field chosen. As an example, if "the ECOSAR category chosen by the user is "Phenols", the term "Phenol" will not return a hit.

CAUTION: If direct entry into the box is utilized for multiple terms the user cannot use the highlighting function to select the terms. Each term must be typed in or copied into the box a string from a text editor. Otherwise, only the most recent entry will appear in the search, not the entire string. Capitalization and correct spelling are essential.

Exporting a search to Excel

After clicking on the green "Download as Excel file" button, your database search will be saved to your "downloads" folder. A screenshot of the data file is provided below in **Error! Reference source not found.**. The file structure (column placement, content) is essential to being compatible with the analysis tools. A summary of the information included in the downloadable file is provided in **Error! Reference source not found.**.

Figure 6. Excel file export

If a user desired to include additional data, not in the EnviroTox Database, the additional data can be appended to the exported data file as described above. The file format for the output must be identical to those required by R analytical tools. If the file structure is maintained, additional analysis should be possible to conduct with no issue. The user is cautioned however

that the analytics are on a publicly accessible server, with limited control. Analyses are not saved nor tracked by MTSU or HESI.

NOTE: The file output nomenclature does not contain the search terms used to develop the EXCEL output file. It is advisable to rename the output file to indicate the type of search and date for future reference by the user.

Table 6. Output columns of a database search including the ecotoxicological information for each chemical.

CAS*	Harmonized CAS number as found in the Physical-chemical descriptor file
Chemical name*	Common name of the chemical
Latin name*	Genus and species name of the test organism
Trophic level*	Designation as algae, invertebrate, fish, amphibian, plant, or fungi
Effect*	Type of response, such as mortality, immobility, population growth rate
Effect value	Concentration at which the response was observed
Unit	Units associated with the Effect (universally mg/L)
Test type*	Acute or Chronic
Test statistic*	Varies depending on the study, e.g., LC50, NOEC, MATC. Consult logic diagrams for acceptable ranges
Duration*	Duration of the test (varies, may be hours, days, months, etc.)
Duration (days)*	Duration of the test in days
Duration (hours)*	Duration of the test in hours
Effect is 5X above water solubility*	Flagged with a "1" it is above; a "0" if not
Source	Citation of primary data source
Version	Version of EnviroTox
Reported chemical name	Original reported name from the primary source
Original CAS	Original CAS (important for metals groupings)

Terms that can be searched within the tool

PNEC Calculator Tool

The user selects the geographic Region for which the PNECs will be determined as seen below. Each region or federal regulatory authority has its own suite of considerations for PNEC determination. The combination of breadth of test species data with whether the data is of acute or chronic duration determines the Application Factors assigned to derive regional PNECs (see also Belanger et al. 2021).

A data file is loaded in the usual manner by placing the cursor into the "Browse" box and navigating to the location of the Excel file containing the data to be analyzed. As stated above, the file needs to conform to that structure initially distributed by the query to the user as this is the only file structure that is read by the system.

Once the file is uploaded a quick summarization of the information is displayed that includes an enumeration of the number of toxicity data available (= Rows), the number of unique chemicals (= Chemicals), the breadth of taxa tested (= Species), the amount of acute data (= Acute) and the amount of chronic data (= Chronic) that is in the uploaded file.

The user can move to the Full PNEC Table tab which will display the chemical-by-chemical output. The information can be viewed by adjusting the vertical and horizontal scroll bars. The user can also indicate how many entries they would like to see at a time from 10, 25, 50 or 100 entries. This tab also provides the ability to perform two different exports of information associated with the chemical specific PNECS in Excel format:
- The full PNEC Table (see **Error! Reference source not found.** below for a description of the PNEC Table contents)
- Geometric mean toxicity data by species for each chemical in the output

PNECs are regional in nature and reflect the level of conservatism applied to a local regulatory authority. **Error! Reference source not found.** provides a summary of representative Application Factors applied to the most sensitive available aquatic toxicity data at various levels of data availability.

Table 7. A summary of various aquatic PNEC assessment factors

Data	Canada[a]	Japan	OECD[b]	US EPA[c]	EU TGD[d]
QSAR			1000	1000	
Acute Data (one or two species)	1000	100 × ACR[e]	1000	1000	
Acute Data (3 taxa)	100	10 × ACR[e]	100	100	1000
Chronic Data[f] (1 taxa)		100		10	100
Chronic Data[f] (2 taxa)		50		10	50
Chronic Data (3 taxa)[f]	10	10	10	10	10
Chronic Probabilistic					1 – 5
Microcosm/Mesocosm Data	Case-by-case		Case-by-case	1	Case by case; 1-10

a) Environment Canada (1997). Maximum factors; however, new PNEC derivation approach is under development where AF are calculated based on a variety of criteria and not predefined.
b) OECD (1992)
c) Zeeman and Gilford (1993), Nabholz (1991)
d) EU TGD refers to short- and long-term toxicity instead of acute and chronic toxicity.
e) see: http://www.meti.go.jp/policy/chemical_management/english/cscl/about.html, Japan Chemical Substance Control Law (accessed 28 March 2017), ACR applied to algae is 20, for Daphnia ACR for amine and non-amine compounds are 100 and 10, respectively; ACR for fish = 100
f) Application factors generally assigned to the most sensitive data point available

The actual PNEC determination process is somewhat more complex than the above simplified table as a variety of data combinations may be available. For the purposes of the ecoTTC utilization of PNECs, the following logic diagrams were developed to allow the interpretation of EXCEL files and toxicity data by chemical into consistently applied Application Factors. Future regional PNEC determination logics will likely be added over time.

Table 8 provides an overview of the different data combinations that are employed in the R-logic. Users should not interpret any of these as endorsements of the logic for regulatory application as actual decisions for chemical approval have many other factors associated with the decision-making process. Further, the PNEC derivations utilize a single AF assignment logic for a region when it is known that multiple assessment types may be employed. For example, in the US, industrial chemicals evaluated under TSCA (now The Lautenberg Chemical Safety Act of 2016) are evaluated differently from pesticides (evaluated and regulated under the Federal Insecticide, Fungicide, and Rodenticide Act). Below is a table describing the assignment of Applications Factors (AF) in the PNEC derivation logic.

Table 8. PNEC codes associated with various combinations of acute and chronic data and the application factors associated with each.

Region	PNEC Code	Data combination	AF Assigned
Unspecified, not part of a regulatory implementation; not currently implemented into ecoTTC			
	PNEC1	ecoTTC already available expressed as 5th percentile of PNECs in a group	1
	PNEC2	QSAR output for a local type QSAR (e.g., one for a specific group of homologous compounds), applied to most sensitive taxon	10,000
	PNEC3	QSAR output for a generalized QSAR (e.g., ECOSAR class) applied to most sensitive taxon	10,000
United States	PNEC4	1 trophic level acute	1000
	PNEC5	2 trophic levels acute; use most sensitive taxon	1000
	PNEC6	3 trophic levels acute; use most sensitive taxon	100
	PNEC7	3 trophic level acutes; 1 chronic on less sensitive acute taxon	100
	PNEC8	3 trophic level acutes; 1 chronic on most sensitive acute taxon	10
	PNEC9	3 trophic level acutes; 2 chronics including most sensitive acute taxon	10
	PNEC10	3 trophic level acutes; 3 trophic level chronics including most sensitive acute taxon	10
	PNEC11	≥10 species chronic toxicity data; perform Species Sensitivity Distribution (SSD)	1-5[b]
	PNEC12	≥ 10 species chronic toxicity data; Mesocosm or microcosm	1-5[b]
Europe	PNEC13	1 trophic level acute	10,000[a]
	PNEC14	2 trophic levels acute; use most sensitive taxon	5000[a]
	PNEC15	3 trophic levels acute; use most sensitive taxon	1000
	PNEC16	3 trophic levels acute; 1 chronic available (fish or invertebrate) but not on most sensitive acute	1000
	PNEC17	3 trophic levels acute; 1 chronic available (fish or invertebrate) which is also most sensitive acute	100
	PNEC18	3 trophic levels acute; 2 chronics available including most sensitive acute taxon	50
	PNEC19	3 trophic levels acute; 3 trophic levels chronic including most sensitive taxon	10
	PNEC20	≥10 species chronic toxicity data; perform SSD	1-5[b]

PNEC21	≥ 10 species chronic toxicity data; Mesocosm or microcosm	1-5[b]

Figure 7 provides a screen shot of the Full PNEC Table tab. The table was constructed for a query employing the US PNEC algorithm. PNEC groupings may be especially useful to evaluate in greater detail the consequences of having more and less data for a given chemical data set for example.

Figure 7. PNEC tool table output

Table 9. Contents of the Full PNEC Table with explanations

Chemical name	Common chemical name
CAS	CAS number
Acute Algae (ug/L)	Geometric mean acute algae result in µg/L
Acute Invertebrate (ug/L)	Geometric mean acute invertebrate result in µg/L
Acute Fish (ug/L)	Geometric mean acute fish result in µg/L
Chronic Algae (ug/L)	Geometric mean chronic algae result in µg/L
Chronic Invertebrate (ug/L)	Geometric mean chronic invertebrate result in µg/L
Chronic Fish (ug/L)	Geometric mean chronic fish result in µg/L
Number of Acute Levels	Number of trophic levels with acute data
Number of Chronic Levels	Number of trophic levels with chronic data
PNEC Group	PNEC group defined by the particular combination of acute and chronic data for a chemical
Group Driving PNEC	Most sensitive taxonomic group
Application Factor	Application Factor assigned for the particular combination of acute and chronic data for a chemical
Final PNEC (ug/L)	Final PNEC value (lowest relevant toxicity data/AF)

The PNEC Group is defined by the particular combination of acute and chronic toxicity data available and follows the regional PNEC determination logic.

Initial data visualization tools are provided to the user which may be useful and are visible by scrolling to the bottom of the PNEC table. This summary information depicts a heat map (Figure 8) of the available acute and chronic data and the associated acute or chronic categorizations for the associated PNEC Table file is displayed along with an enumeration of the different PNEC groupings.

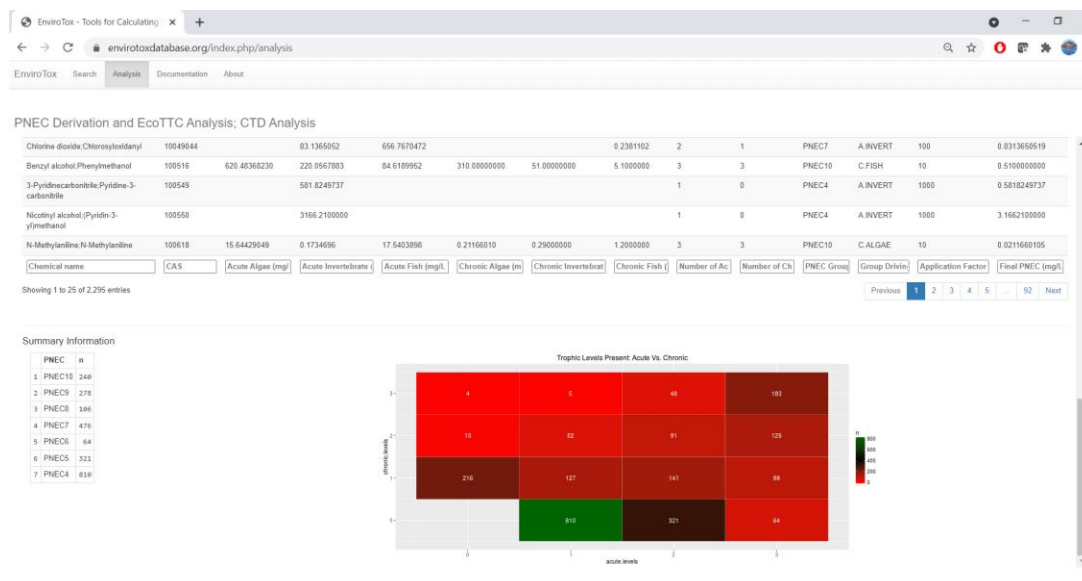*Figure 8. Heat-map of acute and chronic data within the PNEC tool*

A table with geometric means by species (Figure 9) can also be viewed and downloaded by selecting the appropriate tab on the analysis screen. This table can also be downloaded and may be useful for a variety of other analytical purposes such as Species Sensitivity Distribution analysis.

Figure 9. Table with geometric means by species

ecoTTC Tool

The tab labeled "TTC Analysis" is dedicated to the calculation of an ecoTTC for the chosen/queried data set which was uploaded for PNEC determination and the selected region. The user will see several items in this display which deserve attention (Figure 10)

Figure 10. ecoTTC tool page

1. <u>PNEC Groups to Include:</u> PNECs can vary in their "quality" in that the underlying ecotoxicological data may be composed of only acute data on one species to numerous acute and chronic data on a range of species. The actual combination of available data dictates the Application Factor applied to usually the most sensitive acute or chronic information available. The default ecoTTC is calculated using all available PNECs but under some conditions, the user can choose which PNECs to include. Very large data sets may have sufficient information to utilize only chemical data sets where the underlying ecotoxicological data is fully complete as an example. The definition of PNEC Groups is provided in the Reference tab in this screen.
2. <u>Chemical:</u> This region displays the compounds included in the initial upload and for which chosen PNEC groups are to be used. The user chooses to view 10, 25, 50, or 100 entries using the drop down to the left and above the Chemical name column.
3. <u>InterestVar:</u> This is the final PNEC value for the chemical.
4. <u>PNEC Group:</u> The PNEC grouping to which the chemical belongs based on the completeness of the underlying ecotoxicological data.
5. <u>Run Analysis:</u> This button initiates the computation of the ecoTTC.
6. <u>Reference:</u> This tab provides a convenient table of the generic Application Factors assigned to the available ecotoxicological data for the chemicals. Depending on whether US or

Europe was chosen for the initial PNEC determination, a different table of Application Factors will be displayed.

ecoTTC analysis

Once the analysis is initiated (Run Analysis button), progress towards completion is tracked in the upper right of the screen (progress tracker). Upon completion, a screen similar to the one shown in Figure 11 is displayed. Depending on the user's display settings, horizontal and vertical scroll bars are often displayed which can provide a view of additional information.
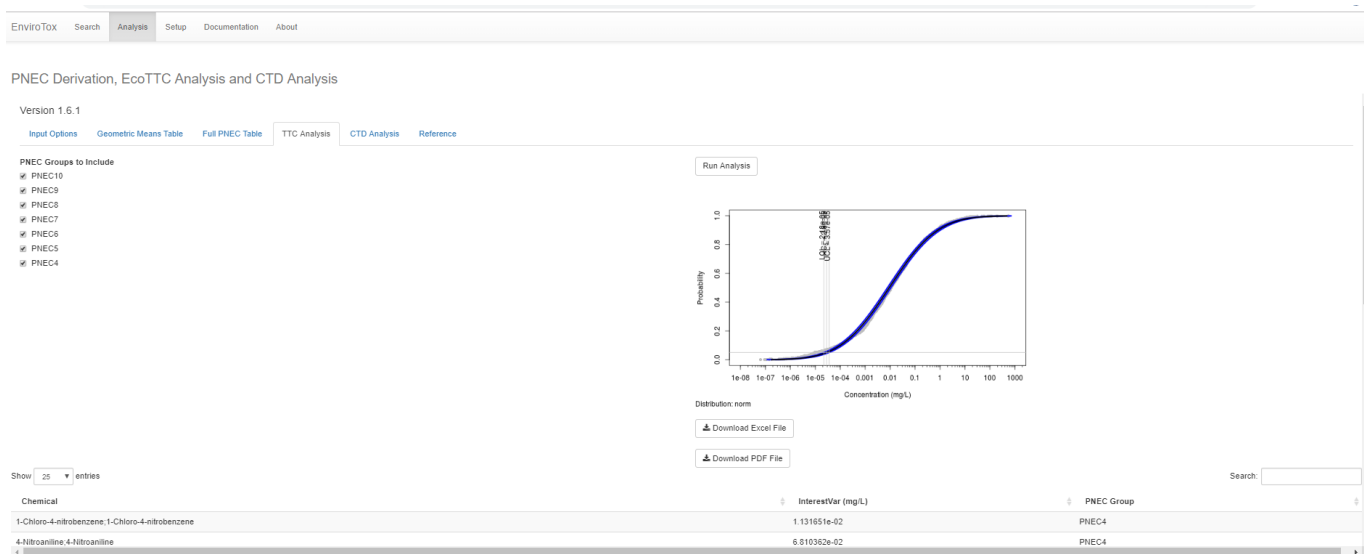
Figure 11. ecoTTC analysis

A thumbnail representation of the log normal cumulative frequency distribution of PNECs is displayed for convenience (note that the log-logistic distribution will also be found in the pdf output). The fifth percentile PNEC (=ecoTTC) and associated upper and lower confidence intervals are also given. Very importantly, a high- resolution pdf of all graphical outputs can be downloaded as well as a file of all statistical characterizations available for the calculated distributions. Note that the thumbnail will only be for the assumed log-normal distribution but both a log-normal and log-logistic analysis are available in the downloaded files.

The ecoTTC tool fits two distributions to each set of data provided: normal and logistic. Technically, all calculations are performed by fitting normal and logistic distributions to the log10-transformed concentration values, and fifth percentile estimates from these distributions are reported by back-transforming to the measured concentration scale. Note that this is equivalent to fitting log-normal/log-logistic distributions to the measured concentration scale of the data.

Both the normal and logistic distributions are symmetric about their centers and are generally difficult to differentiate in data analyses unless there is a large amount of data provided. Generally speaking, the logistic distribution has heavier tails, or a slightly wider spread than

does the normal distribution. Both the logistic and normal distributions are defined by two parameters: a location parameter, relating to the mean of the data, and a scale parameter, relating to the variability or spread of the data.

For each distribution, calculations are performed to estimate these parameters, and subsequently derive a joint confidence set for the two parameters of the distribution, from which confidence intervals on percentiles of the distribution can be calculated. Concrete guidance to the user cannot be advised solely on the statistical outcomes as with all PNEC implementations, these can also reflect matters of environmental policy (e.g., the level of conservativeness employed). In general, the developers advise to use the distribution which provides the best empirical fit to the data, regardless of the 5th percentile calculated outcome. Individual data points at the tails, which may profoundly influence the distribution fits can, be further inspected as to their inherent study qualities. Censoring of data can be done but should be fully documented and justified.

The analysis output file in Excel is shown in Figure 12, and for each distribution, the HC5 estimate (the 5th percentile of the PNEC distribution), a confidence interval on the HC5, the location and scale estimates from the log-scale fit, and an Anderson-Darling (AD) goodness-of-fit test p-value are provided. A significant AD p-value generally indicates that distribution is not a good fit to the data, however, it could also reflect the presence of outliers, which should be considered. For large sets of data, it may prove more difficult to identify a distribution that fits the data well. Note that because the logistic distribution has heavier tails than the log-normal distribution, the log-logistic is less sensitive to the presence of outliers.

Distribution	X	HC5 (mg/L)	LowerCL (mg/L)	UpperCL (mg/L)	location	scale	AD
Logistic	0.05	3.04903E-05	2.29702E-05	3.99463E-05	-1.98674535	0.858938899	0.046574287
Normal	0.05	2.80338E-05	2.17601E-05	3.57215E-05	-2.041244864	1.526623634	0.003780657

Figure 12. Excel download of ecoTTC analysis

The pdf plot output consists of scale location diagrams, cumulative probability plots and distribution density plots for both normal and logistic distributions. It is up to the user to ascertain the most appropriate model and plot for their application.

- *Scale-location diagram*: This plot (Figure 13) is primarily included as a quality check. The points represent combinations of the two parameters (after log10 transformation, location on the x-axis and scale on the y-axis) that determine the shape of the distribution. Points inside of the light-blue line represent parameter combinations that would be considered reasonable for the data. The colored region should be approximately elliptical. If the shape of this region is not elliptical, it may be an indication that the distribution is not a good fit, or other problems with the data. Each point inside of the ellipse corresponds to a unique distribution, with its own HC5 estimate. The range of these HC5 estimates from within the confidence region defines the HC5 confidence interval. Other percentile confidence intervals (blue lines in the following figure) are calculated similarly. The plot below is for a representative data set conforming to the desire for an elliptical shape and visual confidence boundaries.
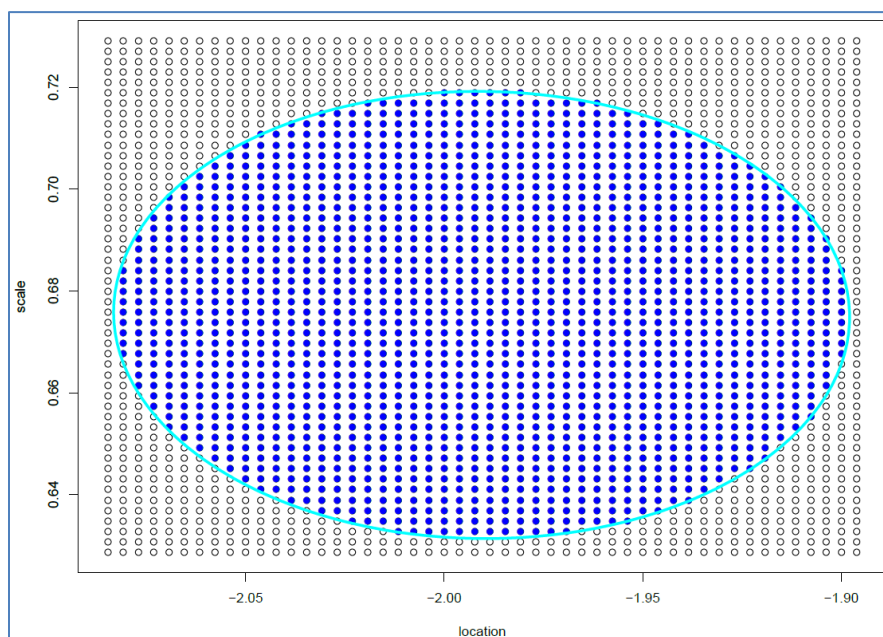
Figure 13. Scale-location diagram

- *Cumulative Probability*: The black line in Figure 14 represents the best reasonable fit for each distribution and the blue lines show the 95% confidence interval at each concentration. Systematic departures of the raw data points from these lines, especially those falling outside of the confidence interval may indicate that the distribution type (logistic or normal) is not a good fit for the data. It could also indicate that there are outliers in the data that are influencing the estimation of distribution parameters. Data at the tails of the distribution (very high and/or very low concentrations) can heavily influence estimation of distribution parameters and cause a poor fit. The ecoTTC point estimate (HC5 label in the graph as a vertical line) is equivalent to the 5th percentile PNEC of the distribution. Individual points (gray) represent a distinct chemical. LCL and UCL are the Lower and Upper 95% confidence limit estimates around each concentration.
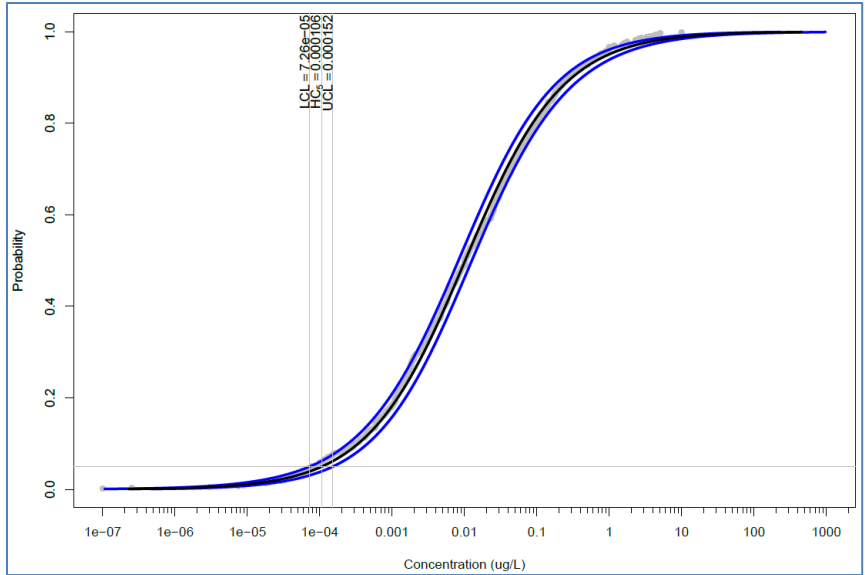
Figure 14. Cumulative probability distribution

- *Distribution Density*: This is an alternative visualization of the fit and confidence interval (Figure 15). The best-fitting distribution to the data is shown in black, while the best-fitting distributions that are restricted to have 5th percentiles equal to the confidence interval limits are shown in color. The plot is intended to show the range of reasonable distributions that could fit the data. The raw data density is represented just above the x-axis with the gray ticks. The higher the sample size, the closer these distributions will look to each other. HC_5, LCL and UCL have the same meanings as in the cumulative distribution.
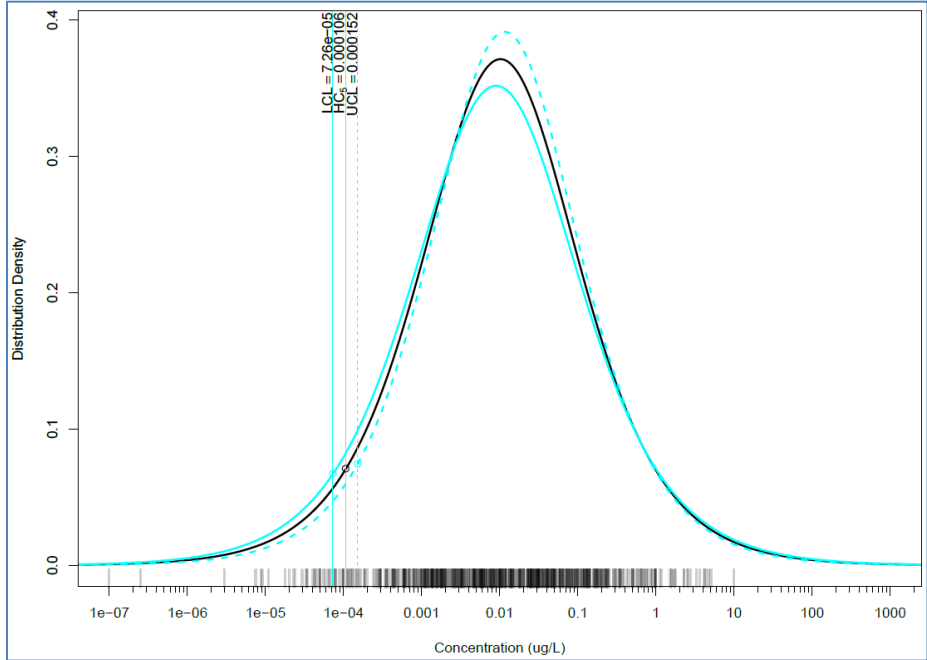
Figure 15. Distribution density

In cases where both the normal and logistic clearly do not fit the data, the expertise of a trained statistician may be required. While alternative distributional choices can be explored, the ecoTTC analysis may contain sufficient data (N>=80) that nonparametric estimation of the HC5 is possible (see for example, Hahn and Meeker 1991), avoiding the need to choose a distribution, and being robust to outliers, particularly on the high end of the distribution. Future versions of this tool may expand the distributions available for analysis.

Most outputs from the ecoTTC database search, PNEC algorithms, and ecoTTC calculations can easily be re-formatted and applied to other statistical and/or graphics display programs as they are essentially EXCEL flat files.

Chemical Toxicity Distributions (CTD)

The ecoTTC approach relies on the development of chemical and region specific PNEC values using regulatory assigned application factors. These PNEC values are then statistically modeled to derive a PNEC HC5. There may be situations where a researcher or assessor is interested in the toxicity distributions, without the added conservatism introduced by the assessment factors or the regional overtones of their application. A chemical toxicity distribution (CTD) can be used to perform this type of analysis.

ecoTTCs and CTDs rely on the same underlying theory and statistical methods. These approaches differ solely based on the type of input data: ecoTTCs contain distributions of PNECs and CTDs contain distributions of hazard values.

Traditionally, CTDs have been used to probabilistically model hazard values from a single species and test type (e.g., acute *Daphnia* toxicity tests; Williams et al 2011). Slight modifications of this approach have been included in the CTD Analysis tool to allow for CTDs to be performed at single species level, trophic level, or incorporating all trophic levels. Additionally, the tool allows for CTDs to be constructed with just acute or just chronic data, or chronic data supplemented with acute values. Users need to carefully examine the data being loaded into the CTD tool and critically think about how the CTD results may be influenced by the relative contribution of data from different species, trophic levels, and experimental durations.

CTD analysis can be conducted with the same exported Excel data file used in an ecoTTC analysis. To initiate data analysis, the user should click on the "Analysis Tab".

A data file is loaded in the usual manner by placing the cursor into the "Browse" box and navigating to the location of the Excel file containing the data to be analyzed. A "PNEC region" does not need to be selected, as no assessment factors will be used and no PNECs will need to be derived.

Once the file is uploaded a quick summarization of the information is displayed that includes an enumeration of the number of toxicity data available (=Rows), the number of unique chemicals (= Chemicals), the breadth of taxa tested (=Species), the amount of acute data (=Acute) and the amount of chronic data (=Chronic) that is in the uploaded file.

After the data file has been loaded, the user should click on the "CTD Analysis" tab. Here, the user will decide what type of data will be included within the CTD analysis (Figure 16).

Figure 16. CTD tool page

1. <u>Method</u>: Here the user chooses which type of data will be included within the CTD:
    a. "Acute Only".
    b. "Chronic Only".
    c. "Chronic Supplemented with Acute"

    The "Acute Only" and "Chronic Only" data are self-explanatory. This selects only data that is of a specific test type.

    The "Chronic Supplemented with Acute" data is generated in several steps. First, all chronic data is collected. If a chemical has acute toxicity data, but not chronic toxicity data, an Acute-to-Chronic ratio (ACR) is applied to convert the acute data point into a "chronic" value. This value is supplemented to the "Chronic Only" dataset.

    The applied ACR is chosen based on the trophic level of the test organism. Acute fish and invertebrate data are divided by an ACR of 10 to generate a chronic toxicity value. Acute algae data is divided by an ACR of 3 to generate a chronic toxicity value.

2. <u>Split Analysis by</u>: This drop-down menu lets the user decide if the CTD should be performed with:
    a. "Species". An individual species, like a traditional CTD.
    b. "Trophic level". A specific trophic level (e.g., Fish, Invert, Algae).
    c. "All". This selects all data.

3. <u>Levels of by Var to Analyze (n>5)</u>: The user manually types in the name of the species or trophic level they wish to analyze. Only options that have a minimum of 5 unique CAS will be accepted. This prevents the CTD tool from generating non-sensical distributions. This window will not be present if you have selected to run "Split Analysis by: All", as no further data parsing is required to run this option.

It is worth noting that a 5-data point CTD (or ecoTTC) is not considered to be robust and should be interpreted with extreme caution. A minimum of 8 data points is suggested, though the predictive power of the model will increase with 10+ data points.

After you have made your selection, a completed summary table of the data will appear (Figure 17). In the example below, the following selections have been made: "Chronic Supplemented with Acute". "Trophic Level", "ALGAE".

Table 10 provides an overview of the data used in the CTD analysis.

Figure 17. CTD analysis

Table 10. Contents of the CTD summary table with explanations

CAS	CAS number
Chemical name	Common chemical name
Trophic level	Trophic level of the species contained within the CTD. This column reflects the choices made in the "Split Analysis By" drop-down menu.
Acute (mg/L)	A geometric mean of all acute toxicity values contained within this database subset for a specific CAS.
Chronic (mg/L)	A geometric mean of all chronic toxicity values contained within this database subset for a specific CAS.
ACR Adjustment	The Acute to Chronic Ratio (ACR) that was applied to derive the final value. If no chronic toxicity value is present, and ACR of 3 (algae) or 10 (fish, inverts) would be applied to acute toxicity data. If chronic toxicity data is present, no ACR is needed. An ACR of 1 is written to show that the chronic toxicity value was used.
InterestVar	The final toxicity value (mg/L) that will be loaded into the CTD calculation.

After the analysis methods have been assigned, a geometric mean is calculated for each CAS. If data is summarized on a "Trophic Level", a CAS-level geometric mean would be calculated for all species within that trophic level. If the data is summarized using "All" data, a geometric mean would be calculated using all data. This geometric mean is *not* weighed based on species representation or by trophic level representation.

We recommend carefully examining the data file you want to analyze before making decisions on how to best summarize the data in a CTD. Trophic level CTDs or CTDs using all data may be heavily influenced by the relative make-up of the data. For example, if a chemical had 4 studies with Daphnia magna and 30 studies by various Algae species, all studies would be treated equally. The geometric mean for this CAS would be heavily influenced by the Algal data.

After the data analysis decisions have been made the button "Run CTD Analysis" will appear. Clicking this button will initiate the CTD analysis to occur. Two outputs from this analysis will be generated: "Numeric Results (XLS)" and "Plot Results (PDF)". These results are analogous to the ecoTTC outputs.

References

Barron, M.G., Lilavois, C.R., and Martin, T.M. (2015). MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquat Toxicol* 161**,** 102-107.

Barron, M.G., Otter, R.R., Connors, K.A., Kienzler, A., and Embry, M.R. (2021). Ecological Thresholds of Toxicological Concern: A Review. *Frontiers in Toxicology* 3.

Beasley, A., Belanger, S., and Otter, R. (2015). Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way: Stepwise Information-Filtering Tool (SIFT). *Environmental toxicology and chemistry / SETAC* 34.

Belanger, S.E., Beasley, A., Brill, J.L., Krailler, J., Connors, K.A., Carr, G.J., Embry, M., Barron, M.G., Otter, R., and Kienzler, A. (2021). Comparisons of PNEC derivation logic flows under example regulatory schemes and implications for ecoTTC. *Regulatory Toxicology and Pharmacology* 123**,** 104933.

Belanger, S.E., Sanderson, H., Embry, M.R., Coady, K., Dezwart, D., Farr, B.A., Gutsell, S., Halder, M., Sternberg, R., and Wilson, P. (2015). It is time to develop ecological thresholds of toxicological concern to assist environmental hazard assessment. *Environ Toxicol Chem* 34**,** 2864-2869.

Brill, J.L., Belanger, S.E., Barron, M.G., Beasley, A., Connors, K.A., Embry, M., and Carr, G.J. (2021). Derivation of algal acute to chronic ratios for use in chemical toxicity extrapolations. *Chemosphere* 263**,** 127804.

Connors, K.A., Beasley, A., Barron, M.G., Belanger, S.E., Bonnell, M., Brill, J.L., De Zwart, D., Kienzler, A., Krailler, J., Otter, R., Phillips, J.L., and Embry, M.R. (2019). Creation of a Curated Aquatic Toxicology Database: EnviroTox. *Environmental Toxicology and Chemistry* 38**,** 1062-1073.

De Wolf, W., Siebel-Sauer, A., Lecloux, A., Koch, V., Holt, M., Feijtel, T., Comber, M., and Boeije, G. (2005). Mode of action and aquatic exposure thresholds of no concern. *Environmental Toxicology and Chemistry* 24**,** 479-485.

De Zwart, D. (2002). *Observed regularities in species sensitivity distributions for aquatic species.*

EC (2007). "Corrigendum to Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Offic J Eur Union L136:3–282. See also: http://eur-lex.europa.eu/".).

ECCC (2018). Canadian Domestic Substances List. Accessed 02 October 2018. Available at: https://www.canada.ca/en/environmentclimate-change/services/canadian-environmental-protection-act-registry/substanceslist/domestic.html.

Gross, M., Daginnus, K., Deviller, G., De Wolf, W., Dungey, S., Galli, C., Gourmelon, A., Jacobs, M., Matthiessen, P., Micheletti, C., Nestmann, E., Pavan, M., Paya-Perez, A., Ratte, H.T., Safford, B., Sokull-Klüttgen, B., Stock, F., Stolzenberg, H.C., Wheeler, J., Willuhn, M., Worth, A., Comenges, J.M., and Crane, M. (2010). Thresholds of toxicological concern for endocrine active substances in the aquatic environment. *Integr Environ Assess Manag* 6**,** 2-11.

Gutsell, S., Hodges, G., Marshall, S., and Roberts, J. (2015). Ecotoxicological thresholds-practical application to an industrial inventory. *Environ Toxicol Chem* 34**,** 935-942.

Hahn, T., Diamond, J., Dobson, S., Howe, P., Kielhorn, J., Koennecker, G., Lee-Steere, C., Mangelsdorf, I., Schneider, U., Sugaya, Y., Taylor, K., Dam, R.V., and Stauber, J.L.

(2014). Predicted no effect concentration derivation as a significant source of variability in environmental hazard assessments of chemicals in aquatic systems: an international analysis. *Integr Environ Assess Manag* 10**,** 30-36.

Hendriks, A.J., Awkerman, J.A., De Zwart, D., and Huijbregts, M.a.J. (2013). Sensitivity of species to chemicals: Dose–response characteristics for various test types (LC50, LR50 and LD50) and modes of action. *Ecotoxicology and Environmental Safety* 97**,** 10-16.

ICCA (2018). High Production Volume (HPV) Chemicals Challenge Program. Accessed 02 October 2018. Available at: http://www.icca-chem.org/.

Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., and Embry, M.R. (2017a). Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. *Environmental Science & Technology* 51**,** 10203-10211.

Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., and Embry, M.R. (2017b). Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. *Environ Sci Technol* 51**,** 10203-10211.

Kienzler, A., Bopp, S., Halder, M., Embry, M., and Worth, A. (2019a). Application of new statistical distribution approaches for environmental mixture risk assessment: A case study. *Science of The Total Environment* 693**,** 133510.

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., and Embry, M.R. (2019b). Mode of Action Classifications in the EnviroTox Database: Development and Implementation of a Consensus MOA Classification. *Environ Toxicol Chem* 38**,** 2294-2304.

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., Van Schothorst, F., Vos, J.G., and Würtzen, G. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem Toxicol* 42**,** 65-83.

Mayer, F., and R. Ellersieck, M. (1986). *Manual of Acute Toxicity:Interpretation and Data Base for 410 Chemicals and 66 Species of Freshwater Animals.*

Moermond, C., Beasley, A., Breton, R., Junghans, M., Laskowski, R., Solomon, K., and Zahner, H. (2017). Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches. *Integrated Environmental Assessment and Management* 13**,** 640-651.

Mons, M.N., Heringa, M.B., Van Genderen, J., Puijker, L.M., Brand, W., Van Leeuwen, C.J., Stoks, P., Van Der Hoek, J.P., and Van Der Kooij, D. (2013). Use of the Threshold of Toxicological Concern (TTC) approach for deriving target values for drinking water contaminants. *Water Research* 47**,** 1666-1678.

Rudén, C., Adams, J., Ågerstrand, M., Brock, T.C., Poulsen, V., Schlekat, C.E., Wheeler, J.R., and Henry, T.R. (2017). Assessing the relevance of ecotoxicological studies for regulatory decision making. *Integrated Environmental Assessment and Management* 13**,** 652-663.

Russom, C., Bradbury, S., Broderius, S., Hammermeister, D., and Drummond, R. (1997). *Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales Promelas).*

USFDA (1995). Food additives: threshold of regulation of substances used in food-contact articles; Final Rule. *Federal Register* 60**,** 36582-36596.

Vestel, J., Caldwell, D.J., Constantine, L., D'aco, V.J., Davidson, T., Dolan, D.G., Millard, S.P., Murray-Smith, R., Parke, N.J., Ryan, J.J., Straub, J.O., and Wilson, P. (2016). Use of acute and chronic ecotoxicity data in environmental risk assessment of pharmaceuticals. *Environ Toxicol Chem* 35**,** 1201-1212.

Williams, E.S., Berninger, J.P., and Brooks, B.W. (2011). Application of chemical toxicity distributions to ecotoxicology data requirements under REACH. *Environ Toxicol Chem* 30**,** 1943-1954.