

Illustrative examples with the OECD QSAR Toolbox workflow

Part 2b



Disclaimer

This document aims at assisting users in using the OECD QSAR Toolbox workflows in a systematic and reproducible manner. However, users are reminded that the text of the REACH Regulation is the only authentic legal reference and that the information in this document does not constitute legal advice. Usage of the information remains under the sole responsibility of the user. The European Chemicals Agency does not accept any liability with regard to the use that may be made of the information contained in this document.

Illustrative examples with the OECD QSAR Toolbox workflow. Part 2b

Reference: ECHA-15-R-23-EN

Cat. num: ED-01-15-872-EN-N

ISBN: 978-92-9247-682-3

d.o.i: 10.2823/279129

Publ. date: November 2015

Language: EN

© European Chemicals Agency, 2015

Cover page © European Chemicals Agency

Reproduction is authorised provided the source is fully acknowledged in the form "Source: European Chemicals Agency, <http://echa.europa.eu/>", and provided written notification is given to the ECHA Communication Unit (publications@echa.europa.eu).

ECHA invites interested parties to submit comments on this document which may be incorporated in future updates. These can be submitted using the Contact ECHA page at: http://echa.europa.eu/about/contact_en.asp

European Chemicals Agency

Mailing address: P.O. Box 400, FI-00121 Helsinki, Finland

Visiting address: Annankatu 18, Helsinki, Finland

Table of Contents

1. Long-term aquatic toxicity to fish	4
AQUATIC TOXICITY CLASSIFICATION BY ECOSAR	4
2. Example 1, 3, 5 - trichlorobenzene	5
INPUT	5
PROFILING	5
ENDPOINT	7
CATEGORY DEFINITION	8
DATA-GAP FILLING	8
REPORT AND CATEGORY SUMMARY	20
ADDITIONAL WEIGHT OF EVIDENCE CONSIDERATIONS AND CONCLUSIONS	22

1. Long-term aquatic toxicity to fish

The example illustrates the elements ECHA recommends to take into account when predicting long-term toxicity to fish, in a regulatory context, with the QSAR Toolbox. Please refer to the document “Illustrative examples with the OECD QSAR Toolbox workflow part 1”¹ for an introductory note on the QSAR Toolbox, which provides background information on the software including general considerations on the approach taken. The example does not focus on the reporting of the results. Further guidance on QSARs and how to report QSARs predictions is provided in “Chapter R.6: QSARs and grouping of chemicals” of REACH Guidance on information requirements and chemical safety assessment² and in “Practical Guide 5: How to report (Q)SARs”³. Endpoint specific guidance can be found in the “Guidance on Information Requirements and Chemical Safety Assessment Chapter R. 7b: Endpoint specific guidance.”⁴ In this example, we address the information requirement (regulatory endpoint) for fish long-term toxicity in Annex IX, 9.1.6 of REACH. The recommended test for this endpoint is the OECD TG 210. Short-term toxicity to fish in Annex VIII of REACH may also be covered with this information. Please note that the same reasoning illustrated in this example is applicable for the prediction of other REACH endpoints too.

The example illustrates the workflow for a quantitative hazard assessment of fish long-term toxicity for 1,3,5-trichlorobenzene (CAS number 108-70-3, SMILES: C1=C(C=C(C1Cl)Cl)Cl) using QSAR Toolbox version 3.3. The endpoint covered in the example is no observed effect concentration (NOEC) for growth.

The QSAR Toolbox v3.3 contains specific aquatic toxicity profilers, which are accessible in the module “Profiling”. The “profilers” are attributes that describe chemical or biological (including endpoint-specific) properties of the substances. The Toolbox calculates them starting from the chemical structure. Three profilers cover aquatic toxicity. They differ on the underlying data they were developed from and on the substructures they identify. Among the three aquatic toxicity profilers, “Acute aquatic toxicity classification by Verhaar” and “Acute aquatic toxicity MOA by OASIS” are specific profilers for aquatic acute toxicity. “Aquatic toxicity classification by ECOSAR” is developed for covering both acute and long-term toxicity on fish. It consists of structural alerts that were associated with trends in toxicity from an analysis of aquatic toxicological data. A short description of the profiler is accessible under Options\Modules\Profilers\Aquatic toxicity classification by ECOSAR:

AQUATIC TOXICITY CLASSIFICATION BY ECOSAR

“The Aquatic Toxicity Classification by ECOSAR profiler consists of molecular definitions developed by LMC and OECD to mimic the structural definitions of chemical classes within the US Environmental Protection Agency’s Ecological Structure-Activity Program (ECOSAR). ECOSAR contains a library of class-based SARs for predicting aquatic toxicity, overlaid with an expert decision tree based on expert rules for selecting the appropriate chemical class for evaluation of the compound. ECOSAR is currently programmed to identify 118 chemical classes. The profiler is introduced for chemical categorisation purposes using the class definitions from ECOSAR.

ECOSAR is a grouping method for identifying chemical classes. It is especially relevant for aquatic toxicity endpoints. Since this method is the most robust of the mechanistic grouping methods it is often the method of choice.”

For a more detailed description of the profilers related to acute fish toxicity, please refer to section 2.4 of the “Illustrative examples with the OECD QSAR Toolbox workflow part 2a” available on [ECHA’s webiste](https://echa.europa.eu/webiste)

¹ Previous illustrative examples with the OECD QSAR Toolbox workflow can be found on the ECHA website under http://echa.europa.eu/documents/10162/21655633/illustrative_example_qsar_part1_en.pdf (part 1). Examples on predictions for skin sensitisation and fish short-term toxicity arte described at: https://echa.europa.eu/documents/10162/21655633/illustrative_example_qsar_part2_en.pdf (part 2)

² https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

³ http://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

⁴ https://echa.europa.eu/documents/10162/13632/information_requirements_r7b_en.pdf

The QSAR Toolbox includes five databases (accessible in the module “Endpoint”) that contain experimental data on aquatic toxicity⁵:

- » Aquatic ECETOC (734 chemicals, 9 487 data points, 33 endpoints; adopted 06/2014)
- » Aquatic Japan Ministry of Environment (464 chemicals, 2 900 data points, four endpoints; adopted 08/2014)
- » Aquatic OASIS (2 390 chemicals, 4 826 data points, eight endpoints; adopted 07/2014)
- » ECHA CHEM (6 200 chemicals, 346 792 data points, 1 958 endpoints; adopted 08/2014)
- » ECOTOX (7 677 chemicals, 575 792 data points, 196 endpoints; version 12 June 2014, adopted 09/2014)

Keeping in mind that these databases are donated to the Toolbox, the user needs to verify the quality of data that are used for the prediction, as we will see in the next paragraphs.

2. Example 1, 3, 5 - trichlorobenzene

INPUT

In the “Input” module the user clicks CAS# under single chemical and enters the CAS number without dashes:

- “108703”

In this example the structure of the target chemical is introduced in the QSAR Toolbox in the “Input” module using the CAS number. Alternatively, the use of the drawing tool or the SMILES notation returns two chemicals with the same structure but with different CAS numbers. The user can then select the chemical structure according to the relevant CAS number. In general, a structure can also be selected from a database, inventory, or user list.

PROFILING

The user can select all the available profilers to have an overview of the alerts associated to the structure. In fact, this selection will only affect the computing time but not the following steps that will lead to the prediction.

The outcome of the endpoint specific profiler for aquatic toxicity is:

- » Neutral organics (Aquatic toxicity classification by ECOSAR)

The user can check in the Toolbox the explanation of the profilers and their outcomes. The boundaries for each possible outcome are also described. This information supports the user in checking the reliability of the profilers outcomes.

Please note that the QSAR Toolbox also contains profilers specific for acute aquatic toxicity: “Acute aquatic toxicity MOA by OASIS” and “Acute aquatic toxicity classification by Verhaar”, which can provide useful information.

According to the Aquatic toxicity classification by ECOSAR profiler, the chemical is a neutral organic substance. Most of the chemicals in this group exhibit only narcotic effects. Acute aquatic toxicity profilers

⁵ Note that the numbers given in the list refer to the total number of studies and chemicals collected in the databases. Data on long-term aquatic toxicity are a subset of those. Moreover, there is overlap between the different databases. Hence the same study may appear in several databases.

give indeed an indication of narcotic MoA at least in short-term exposure. A look at the protein binding profilers can also give indications on possible elevated toxicity (i.e. toxicity above the so called baseline toxicity of neutral organics). Investigating the toxicity of the analogues that will be identified later in the workflow may give further evidence on the presence/absence of specific long-term effects not covered by this profiler. In this case, the target structure does not trigger any protein-binding alert. A screenshot of the QSAR Toolbox with some profiling results is shown in Figure 1.

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', and 'Data Gap Filling'. The 'Profiling' section is active, showing a list of profiling methods on the left and a detailed view of the target chemical on the right. The target chemical is 1,3,5-trichlorobenzene, with its chemical structure shown. The results table on the right lists various identifiers and categories, including 'Class 1 (narcosis or baseline toxicity)'.

Filter endpoint tree...	1 [target]
Structure	
Substance Identity	
CAS Number	108-70-3
Chemical IDs	EINECS:2036086
Chemical Name	1,3,5-trichlorobenzene trichlorobenzene, 1,3,5- 1,3,5-tri-chlorobenzene benzene, 1,3,5-trichloro- C6H3Cl3
Molecular Formula	Clc1cc(Cl)cc(Cl)c1
Structural Formula	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	
Profile	
Predefined	
OECD HPV Chemical Categories	Not categorized
US-EPA New Chemical Categories	Neutral Organics
General Mechanistic	
Protein binding by OASIS v1.3	No alert found
Protein binding by OECD	No alert found
Endpoint Specific	
Acute aquatic toxicity classification by Verhaar (Modified)	Class 1 (narcosis or baseline toxicity)
Acute aquatic toxicity MOA by OASIS	Basesurface narcotics
Aquatic toxicity classification by ECOSAR	Neutral Organics

Figure 1. Screenshot of the profiling results for the target chemical.

The QSAR Toolbox provides information to explain profilers, which can easily be accessed by moving the mouse on the profiling result, clicking the right button and selecting 'Explain'. As an example, the results of the OECD HPV Chemical Categories and US-EPA New chemical categories profilers check whether the target chemical belongs to one of the chemical categories analysed in OECD or US-EPA programmes. The results for 1,3,5-trichlorobenzene are shown below:

OECD HPV Chemical Categories – Not categorized

“The target chemical does not fall under any of the category identified in the OECD list of High Production Volume chemicals.”

US-EPA New chemical categories – Neutral organics

“This broad category includes non-reactive non-ionizable organic chemicals such as alcohols, ketones, ethers, alkyl halides, aryl halides and aromatic hydrocarbons. Most neutral organics exhibit toxicity only due to their ability to produce simple narcosis, which is a disturbance of membrane integrity and functioning. A correlation of this type of toxicity with the octanol-water partition coefficient of the chemicals is reported in the scientific literature. Moreover, by applying subcategorization steps we can reduce the uncertainty due to other factors contributing to the long-term toxicity and base the prediction on the octanol-water partition coefficient only.”

ENDPOINT

In the “Endpoint” module, the user retrieves experimental data for the target chemical from the selected databases. In this case the selected databases are:

- all databases in the “Ecotoxicological Information” group

To retrieve the aquatic toxicity experimental data that may be in one or more of the QSAR Toolbox databases, click “Endpoint” in the top menu, select the relevant databases and click “Gather”. The retrieval can be further limited to endpoints (in this case for example to “aquatic toxicity” by selecting “choose...” in the pop-up window: “Read data?”). The next pop-up window is reporting repeated values originating from different sources, and it is recommended choosing the “Select one” option.

The five relevant databases selected are:

- » Aquatic ECETOC
- » Aquatic Japan MoE
- » Aquatic OASIS
- » ECHA CHEM
- » ECOTOX

For predicting long-term fish toxicity, it is sufficient to retrieve only data about aquatic toxicity. No further data selection is possible at this stage, but the next step will allow the selection of long-term fish toxicity data. There are 32 experimental values found for the target (Figure 2).

Upon assessment, no single point was found suitable to cover the endpoint and this is explained in the text that follows.

Endpoint	Value	Study
ChV		
EC-10	(1/1) M: >2 mg/L	
EC50	(1/10) M: 2.25 mg/L, 1.96 mg/L, 0.59 mg/L, 1 m...	
GC50		
LC50	(1/4) M: 12.3 mg/L, 3.3 mg/L, 3.3 mg/L, 3.2 mg/L	
LD50	(1/2) M: 5.53E3 mg/kg, 7.19E3 mg/kg	
LOEC	(1/3) M: 59.4 milligrams per kilogram dry food, ...	
NOEC		
Animalia		
Arthropoda (Invertebrates)	(1/2) M: 0.32 mg/L, 0.2 mg/L	
Chordata (Vertebrates)		
Actinopterygii (Fish)		
Danio rerio	(1/1) M: >0.41 mg/L	
Oryzias latipes	(1/1) M: 0.35 mg/L	
Plantae	(1/2) M: 0.59 mg/L, 0.71 mg/L	
NR-ZERO	(1/3) M: 58.2 milligrams per kilogram dry food, ...	
PT	(1/3) M: 12.8 mg/L, 14.1 mg/L, 14.1 mg/L	
Toxicity on Daphnia Magna		

Figure 2. Experimental aquatic toxicity values for the target chemical found in the QSAR Toolbox

Among others, two NOEC experimental values are available for fish. However, a look at the experimental details (accessible via the QSAR Toolbox by double clicking on the values) shows that one value (0.35 mg/L) is a result of a OECD TG 204 study, which is suitable for short-term fish toxicity but not for long-term. The other value (NOEC>0.41 mg/L) is from an unpublished report of the German Federal Health Office and thus it is not possible to assess the reliability of the study.

Therefore, additional evidence is needed to assess the chronic fish toxicity of the target substance. A read-across and/or a trend analysis using the QSAR Toolbox shall be considered for a weight of evidence approach.

CATEGORY DEFINITION

In the “Category Definition” module, the user chooses the first profiler for the analogue selection. In this case, the selected profiler is:

- “US-EPA New Chemical Categories” (strict)⁶

The starting group is formed in the module “Category Definition” by using the chemical profiler “US-EPA New Chemical Categories” (strict), which determines the molecule as “Neutral Organic” (see Figure 1). To do so, under “Grouping methods”, select “US-EPA New Chemicals Categories” and click on “Define”. Afterwards select “Strict” at the bottom of the pop-up window. It is recommended to start the selection of analogues with a “structural” profiler rather than an endpoint specific one. Indeed, the final selection of analogues has to be structurally similar to the target to obtain a reliable prediction that can be accepted. Once the final group of analogue is identified, this similarity assessment is done by applying expert judgement as shown later in this example.

After selecting the starting profiler, the user can now gather experimental data for all chemicals found in the selected databases that are part of the category. There are 1 224 chemicals gathered in the category (the target and 1 223 chemicals from the databases) and aquatic toxicity data is extracted/gathered for all category members. In total, 34 326 data points are gathered for 1036 chemicals (aquatic toxicity data). At this point, the category is still too broad and no endpoint has been defined, therefore further sub-categorisation is needed. The sub-categorisation will be performed in the data-gap filling module, where the chemicals removed in each step can be visually highlighted, helping the user to follow the process.

⁶ During the definition of the category, it is possible to select whether the analogues have to comply strictly with the functional groups/alerts of the target or if they can also contain additional functionalities/alerts. In the first instance, it is recommended to tick the “strict” option to select only close analogues of the target, excluding compounds with additional functionalities.

DATA-GAP FILLING

In the “Data Gap Filling” module the user chooses:

- » the endpoint to predict (NOEC for fish, will be more specifically identified later)
- » the type of approach for filling the data-gap (trend analysis)
- » the sub-categorisation strategy (described below)

In order to get an overview of the available data, the user can arrange the tree hierarchy so that the endpoint is displayed upfront. To do so, move the mouse over “Aquatic Toxicity” (this cell is marked with a small blue triangle to indicate that the tree hierarchy can be changed at this level) and click on the right mouse button. Select “Set tree hierarchy...” and move “Endpoint” in the Sub-nodes window on first place, then press “OK”. In order to display only fish data, there is a possibility to filter the data matrix. Write in the upper left cell of the data matrix where it says “Filter endpoint tree...” the term “fish”). The following information can then be retrieved:

- » EC10: 56 datapoints for 52 chemicals
- » LC10: 72 datapoints for 22 chemicals
- » NOEC: 2491 datapoints for 259 chemicals

Most often results for long-term fish toxicity are reported as NOEC and EC10 (LC10). Experimental test results expressed as NOECs are available in far higher numbers in the QSAR Toolbox databases. Thus, the choice of NOEC is preferable for this example (Figure 3).

TIP

Via the filter option, LC10, EC10 and NOEC can be grouped together for subsequent analysis. Instead of choosing for example “NOEC” in the sub-nodes as the starting point for the trend analysis, the cell in the row “Aquatic Toxicity” in the first column (target chemical) is selected. The data are filtered by writing in the upper left cell of the data matrix the term “fish” where it says “Filter endpoint tree...”. Then “Trend analysis” is selected, and pressing “Apply” will lead to a pop-up window allowing for the choice of the endpoints, for example ticking the check boxes only for LC10, EC10 and NOEC. With the right mouse button a pop-up window can be accessed giving the options of selecting all, unselecting all or inverting the selection within this pop-up window.

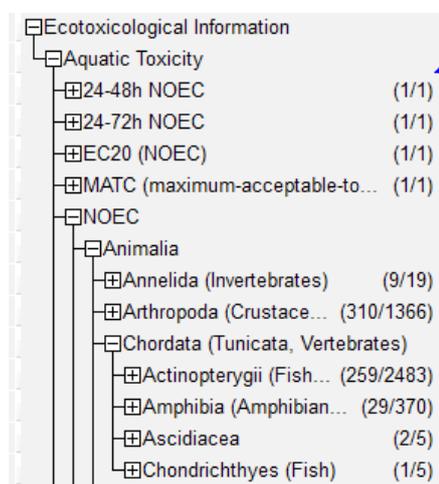
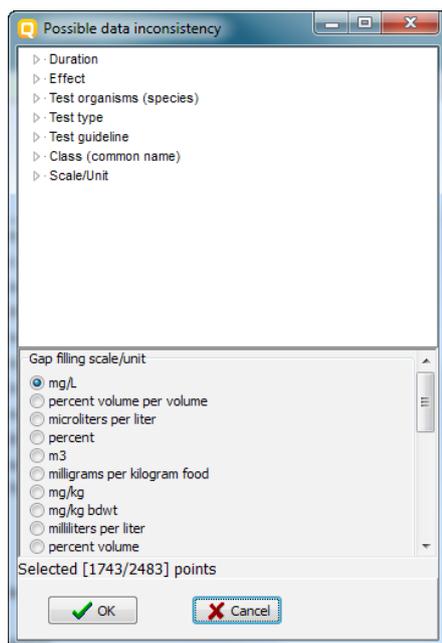


Figure 3. Selection of the endpoint in the “endpoint tree”. 2483 values for 259 chemicals are gathered in the branch NOEC > Actinopterygii(Fish). The tree hierarchy is arranged as following: Endpoint, Test organism (species), Duration, Effect; and the keyword “NOEC” is used to filter the tree and better visualize the selected endpoint.

The databases contain 2483 fish NOEC data points (Actinopterygii) for 259 chemicals for the category of neutral organics. For chemicals with multiple values, by default the QSAR Toolbox uses the average value for the calculation. Note that it is possible to use all the available values rather than the average and this can be a better choice in some cases as discussed later in the document. Before starting the data gap filling, it



is important to select the correct cell in the data matrix (for this example select the cell in the row “NOEC\Animalia\Chordata (Tunicata, Vertebrates)\Actinopterygii (Fish, Neoperygii)” in the first column (target chemical). Given that fish long-term toxicity is a quantitative endpoint, a good number of analogues is available and that a trend between octanol-water partition coefficient and aquatic toxicity for narcotics is often observed, the approach of first choice for the data-gap filling is “Trend analysis.” When clicking on “Apply” to start the prediction process, a pop-up window of the QSAR Toolbox appears checking for possible data inconsistency (Figure 4). The window offers the possibility to filter the data according to the experimental details and to choose the scale/unit for data gap filling. Given that filtering according to the experimental conditions is possible also later in the process, in this step we do not limit the selection. The filtering selects 2483 data points.

Figure 4. “Possible data inconsistency” pop-up window. The picture shows that by selecting “mg/L”, 1743 out of 2483 data points are expressed in this unit or can be converted to it. Only these points will be used for the analysis.

Following these steps, the first plot is generated (Figure 5). It shows the aggregated test data for fish NOEC on Y-axis, expressed as $\log 1/\text{NOEC}$ in mol/L, and the octanol-water partition coefficient ($\log K_{ow}$) calculated from EPIWIN (KOWWIN ver. 1.67) on the X-axis. Please note that the active descriptor on the X-axis can be changed at any time, as shown later in the document.

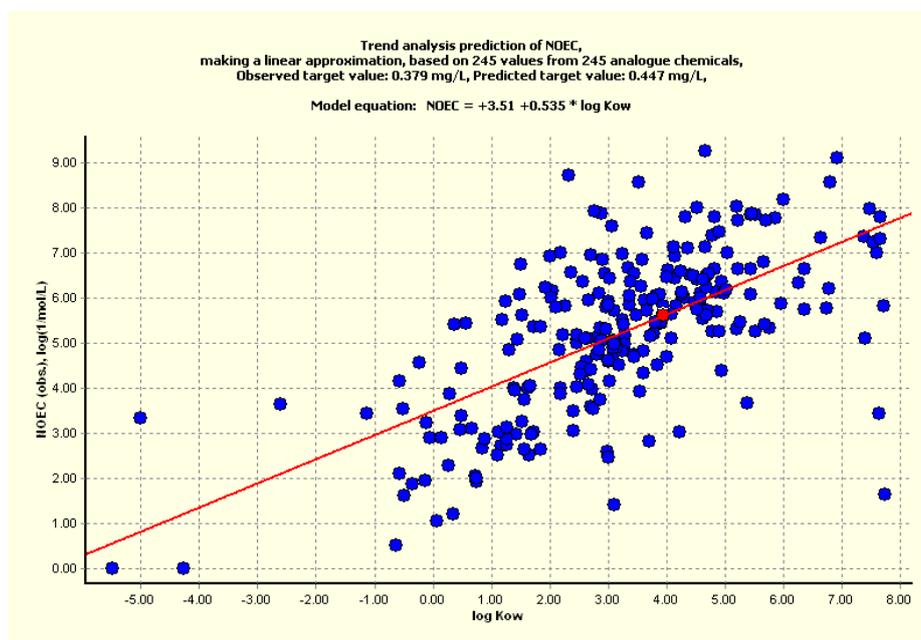


Figure 5. Automatically generated plot for the initial selection (all fish long-term toxicity NOEC values expressed in mg/L). The red dot indicates the target substance, the orange dot represents observed values for the target substance (will be visible in the following graphs), and the blue dots are the selected analogues (i.e. all neutral organics in the selected databases). The trend is represented by the red line. 245 chemicals contribute to define this initial trend.

The plot suggests the existence of a linear relationship between the NOEC values and the log Kow. This first selection is based on a broad chemical category definition and needs a further refinement.

In the “Select/filter data” option the user applies additional filters for analogue selection. As a first step it is advised to:

- remove chemicals with NOEC values above water solubility

For each chemical, it should be checked that the NOEC value does not exceed the water solubility value. Normally, the data points with NOEC values above the water solubility should be excluded from the analysis as these data points are associated with uncertainty. This operation can be performed automatically with the QSAR Toolbox in the section “Select/filter data” > “Mark chemical by WS” > “Water solubility (fragments)” (see Figure 6), and then using the “Remove marked chemicals/points” button. This operation excludes 49 data points, leaving a group of 220 analogues.

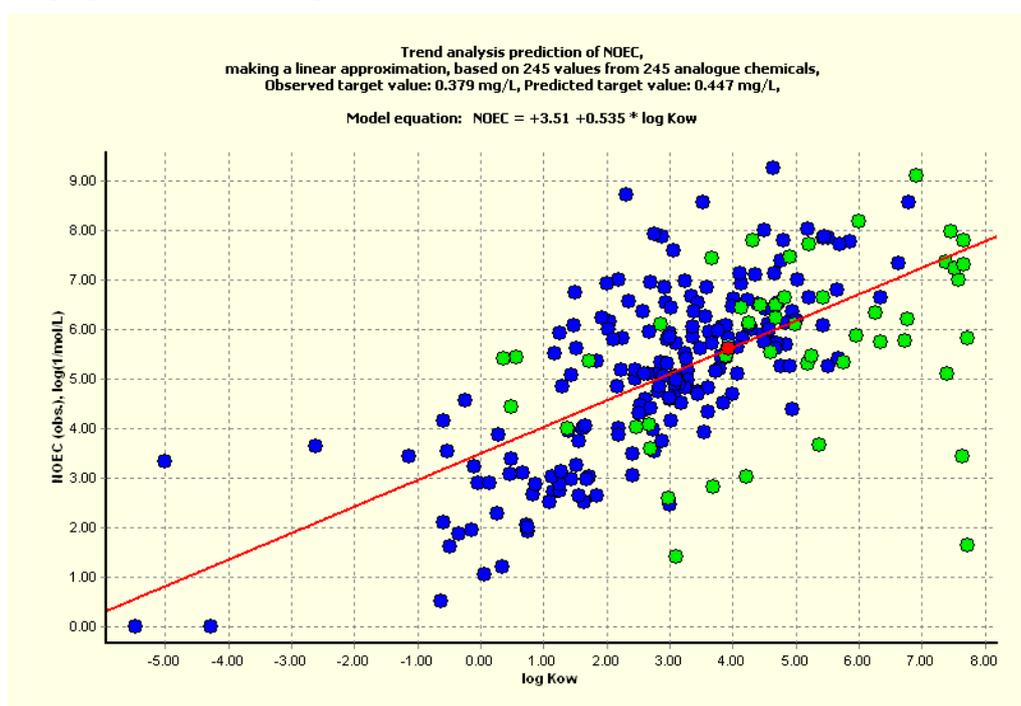


Figure 6. Result values above the calculated water solubility are shown as green dots. These values are recommended to be removed as they are associated with uncertainty.

The category is further sub-categorised utilising pre-defined, endpoint specific and empiric profilers in order to form a more robust category. To do so, select “Subcategorize” in the section “Select/filter data”. A pop-up window opens and various grouping methods can be selected to subcategorise the data set. With the “Remove” button, the analogue groups highlighted in the “Analogues” window can be removed from the data set.

The next sub-categorisation identifies structurally similar compounds from the neutral organics category and is performed according to the following steps:

- » Substance type: discrete chemicals (215) are kept in the category, while the dissociating chemicals (5) are removed (Figure 7a);
- » Organic functional groups (US EPA): chemicals profiled as the target, i.e. containing “Aromatic Carbon”, “Chlorine, aromatic attach”, “Chlorine, olefinic attach” and “Olefinic carbon” are kept (13); note that

depending on the structure of the target, different organic function groups profilers may perform better (see Figure 7b);

- » Aquatic toxicity classification by ECOSAR: Neutral organics (12) are selected (see Figure 7c).

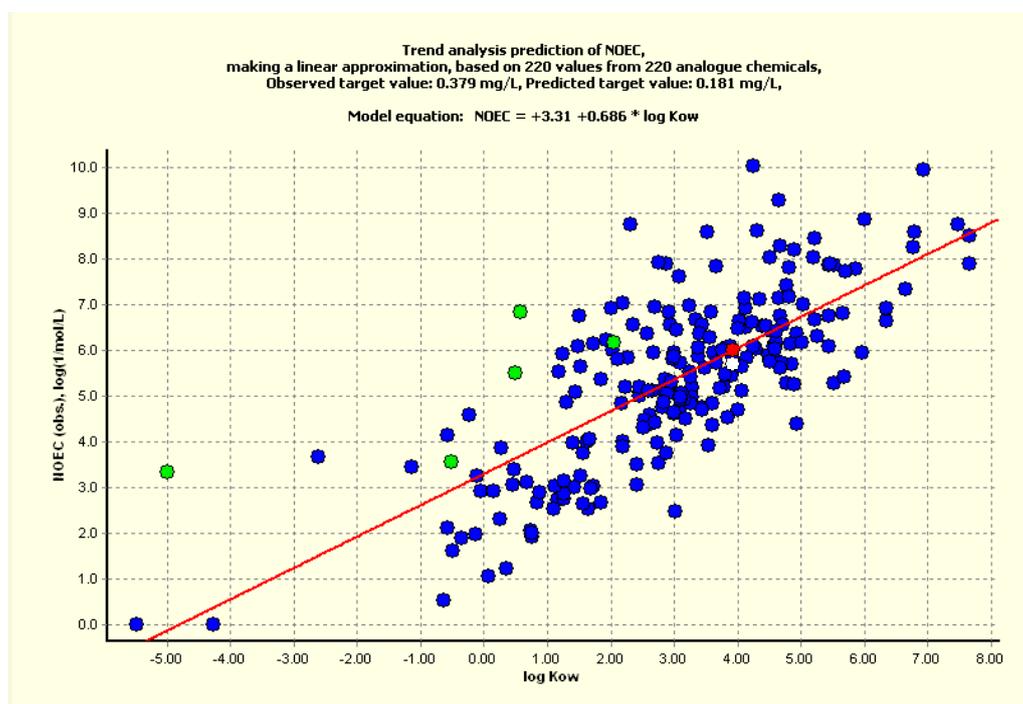


Figure 7a. Result values of dissociating chemicals (Substance type) are shown as green dots. The logKow of dissociating chemicals depends on the pH and therefore result values of dissociating chemicals can introduce higher uncertainty. It is in general recommended to remove result values of dissociating chemicals for trend analysis of discrete chemicals.

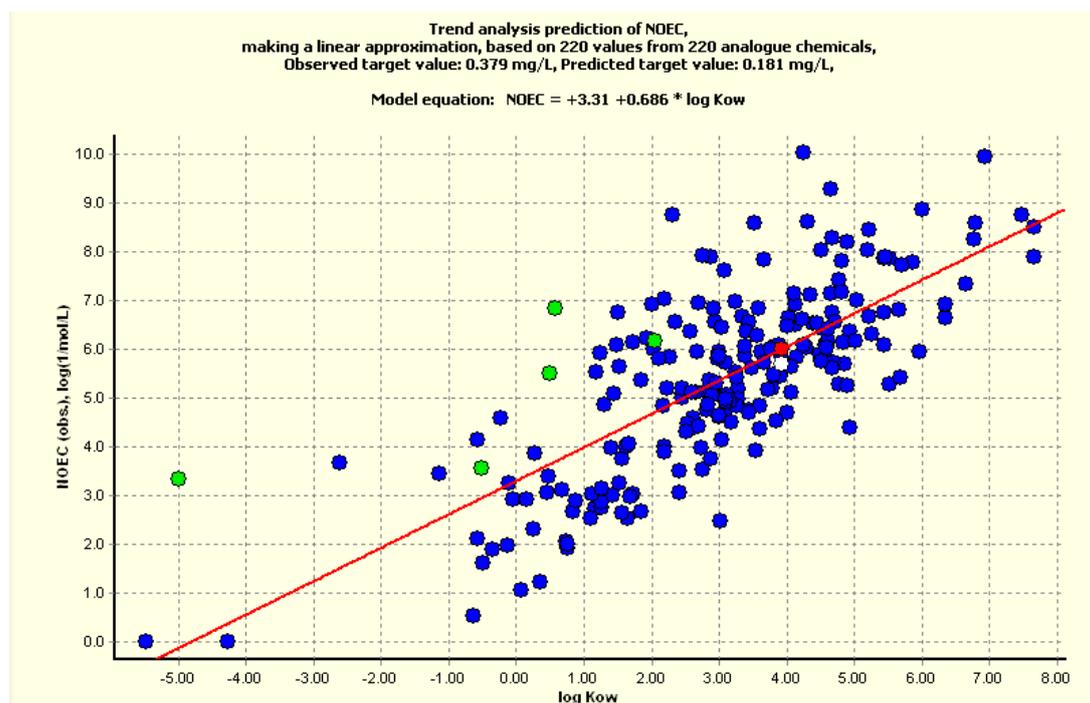


Figure 7b. Result values of Organic functional groups (US EPA) selection: chemicals profiled as the target, i.e. containing “Aromatic Carbon”, “Chlorine, aromatic attach”, “Chlorine, olefinic attach” and “Olefinic carbon” are kept (13).

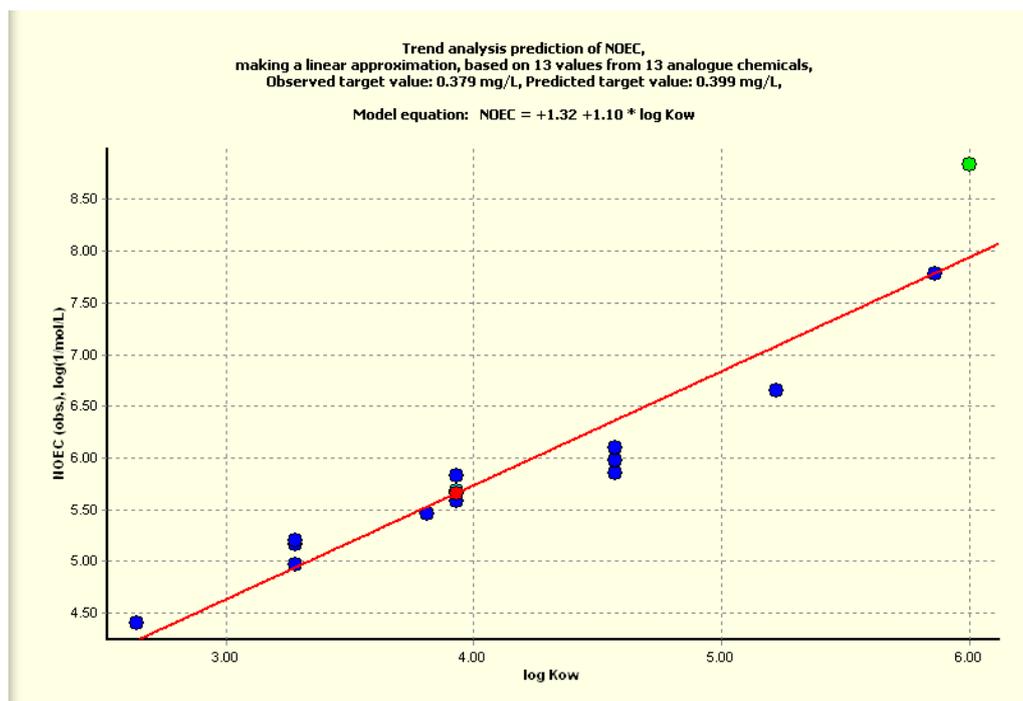


Figure 7c. Further subcategorization “Aquatic toxicity classification by ECOSAR profiler” excludes one additional data point (in green).

There are now 12 analogues in the category (Figure 8).

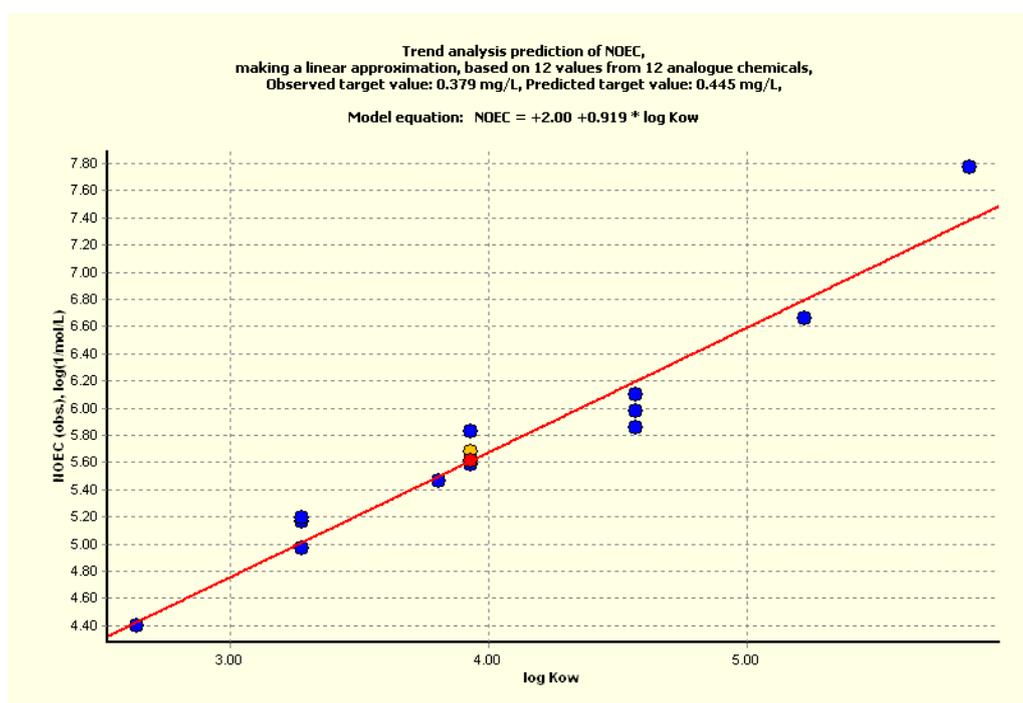


Figure 8. The category of 12 analogues remain after chemical (Substance type, Organic functional groups (US EPA)) and endpoint specific profilers (Aquatic toxicity classification by ECOSAR) have been applied to subcategories the data

The experimental endpoint (i.e. “test conditions” such as fish growth in Toolbox terms) has not been defined yet. To do so, it is possible to subcategorise further according to test conditions, but first it is useful to

display in the plot all data points available per each chemical rather than the average value, which is given by default. This option is available under “Calculation options” the item “Data usage”. In the new window “All” instead of “Average” is selected” and confirmed by “OK” (see Figure 9). With this step, the different endpoints and species as well as variability in one measured endpoint per chemical is differentiated and visualised (by looking at the endpoint (Y values) for the same chemical (X value)).

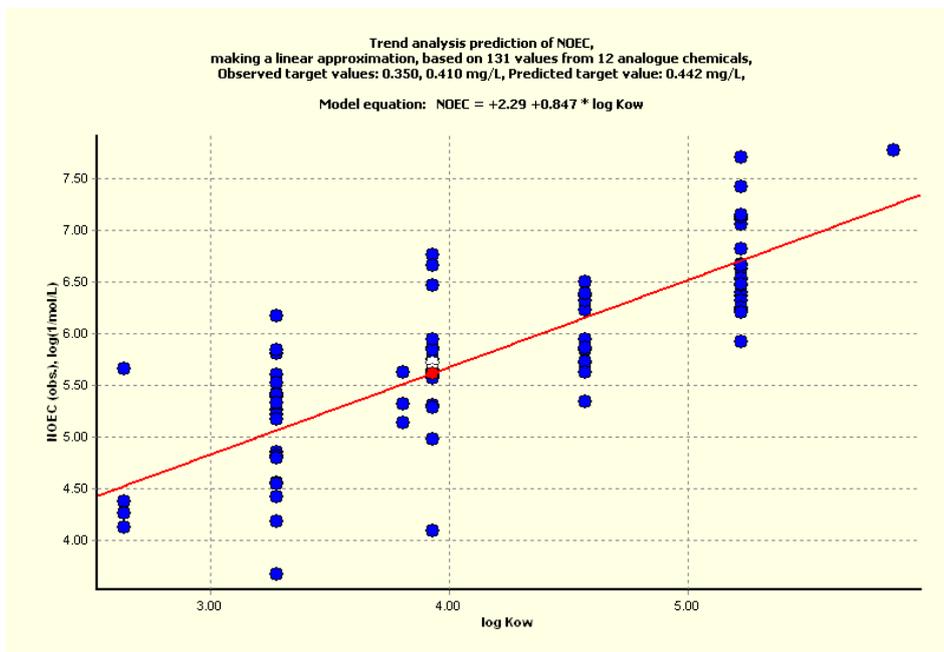


Figure 9. All data points per chemicals are displayed instead of an average value (default setting).

As indicated by the Toolbox above the graph, the category consists of 131 datapoints for 12 analogues. However, some isomers among the analogues have the same calculated logKow, and therefore cannot be easily distinguished in the plot. A way to resolve this overlap is to use experimental logKow values instead of the calculated ones. To do so, “Exp log P” is selected as “Descriptor X”, and under “Descriptors/data” in the right window, “Make active descriptor” is selected (see Figure 10).

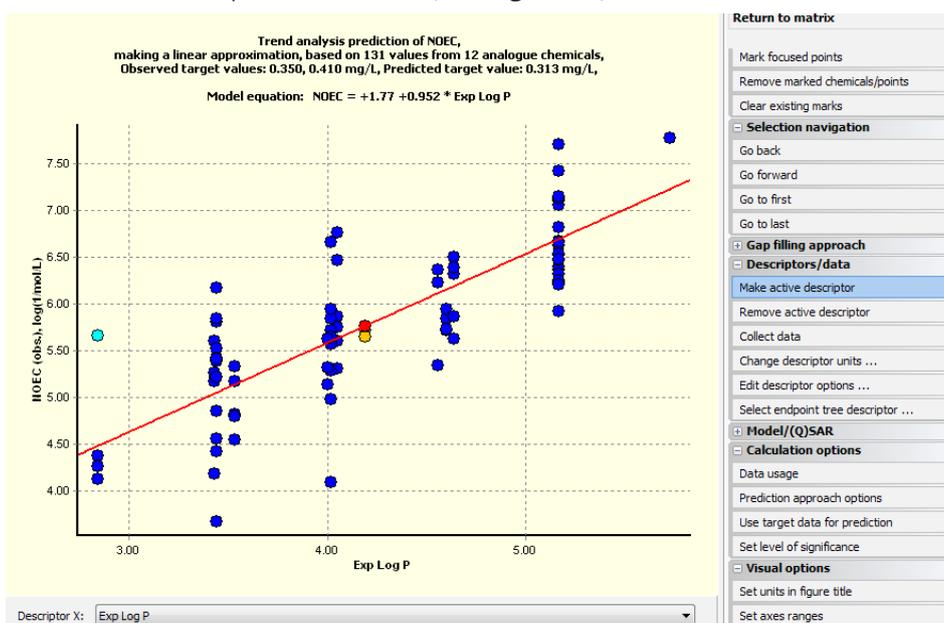


Figure 10. In order to distinguish isomers, the descriptor of the x-axis is changed to experimental logKow

values (“exp log P”) and made an active descriptor.

Experimental Kow values are taken from EPISUITE/KOWWIN database. In this example, the experimental Kow values are available for all the 12 analogues (Table 1). The general agreement observed between experimental and predicted log Kow (the difference between pcalculated and observed value is in the same order of magnitude as the experimental variability) adds confidence to the prediction. Consequently, the predicted NOEC for the target using experimental log Kow is 0.313 mg/L, very close to that obtained with the predicted log Kow (0.442 mg/L). Table 1 lists the experimental and predicted log Kow values. The user has the possibility to extract other physico-chemical properties by selecting other active descriptors and find additional trends that might support the prediction or the robustness of the category.

Table 1 Experimental and calculated log Kow value for the group members identified so far.

CAS number	Name	Experimental log Kow	Calculated log Kow
108-90-7	Chlorobenzene	2.84	2.64
90-50-1	1,2-dichlorobenzene	3.43	3.28
106-46-7	1,4-dichlorobenzene	3.44	3.28
541-73-1	1,3-dichlorobenzene	3.53	3.28
90-13-1	1-chloronaphtalene	4.00	3.81
120-82-1	1,2,4-trichlorobenzene	4.02	3.93
87-61-6	1,2,3-trichlorobenzene	4.05	3.93
634-90-2	1.2.3.5-tetrachlorobenzene	4.56	4.57
634-66-2	1,2,3,4-tetrachlorobenzene	4.60	4.57
95-94-3	1,2,4,5-tetrachlorobenze	4.64	4.57
608-93-5	1,2,3,4,5-pentachlorobenze	5.17	5.22
118-74-1	1,2,3,4,5-esachlorobenze	5.73	5.86

In the “Select/filter data” option the user can specify the endpoint to be predicted by choosing to filter points by test conditions:

- » Duration
- » Effects
- » Species

For this last step, select “Filter points by test conditions” in the section “Select/filter data”. A pop-up window opens and various parameters can be selected in the window “Use subcategorization” to subcategorise further the data set. With the “Remove” button, the analogue groups highlighted in the “Analogues” window are removed from the data set. In the “Analogues” window, items can be selected manually by pressing the

Ctrl-button on the keyboard and selecting the items by clicking on them with the left mouse button.

The sub-categorisation according to test conditions in this example is done as follows (other parameters, such as test guideline and GLP conditions shall be checked when available):

- » Duration: In this example, experimental data with at least 28 days duration were considered. It can be seen in Figure 11 that the selected data points (52 out of 131) with less than 28 days are not among the ones giving highest toxicity. Nine chemicals are left, while two chemicals for which there are no data measured for a duration of at least 28 days are removed. Note that the unit of the NOEC is $\log(1/\text{mol/L})$ and therefore the higher the data point is, the higher is the toxicity.
- » Effects: For consistency, it is preferable not to mix different effects. Among the observed effects, growth appears to be the most sensitive one with a good number of data points for this category (Figure 12). Data points referring to other effects (i.e. larval development, mortality, N/A, percentage of normal larvae hatching, reproduction) are removed in this example (46 points) in order to not mix different effects with lower sensitivities. Nine analogues are left.
- » Test organisms (species): only the species indicated in OECD TGs for fish long-term toxicity are kept (cyprinodon variegatus, oncorhynchus mykiss, pimpephales promelas). 4 data points and one chemical are removed, leaving 33 values for 8 chemicals for the prediction. The prediction is derived from experimental data from three different fish species. The interspecies variability and the existence of a more sensitive species shall be investigated. Ideally, the prediction could be based on data derived from only one fish species, but in this example there are not enough data points to further eliminate data. It is recommended as a good standard practice to always explore further the impact of mixing different species on the variability of the data. In this case the species *Gambusia affinis* appears to be more sensitive than other species (Figure 13). This species was excluded because this fish species is not recommended in OECD TGs and in any case the predicted NOEC value obtained without excluding this species is still comparable with the one obtained by excluding it, 0.210 mg/L vs 0.264 mg/L (Figure 16). Moreover, further exploration of the dataset reveals that there is a high variability even among one single species. The data for *Cyprinodon variegatus* can be traced back to a publication called "Interlaboratory Comparison of the Early Life-Stage Toxicity Test Using Sheepshead Minnows (*Cyprinodon variegatus*)". The paper investigates the inter-laboratory precision (reproducibility) of the test, reporting that the reproducibility has acceptable values, the ratio between the highest and lowest chronic values being equal to about 3.

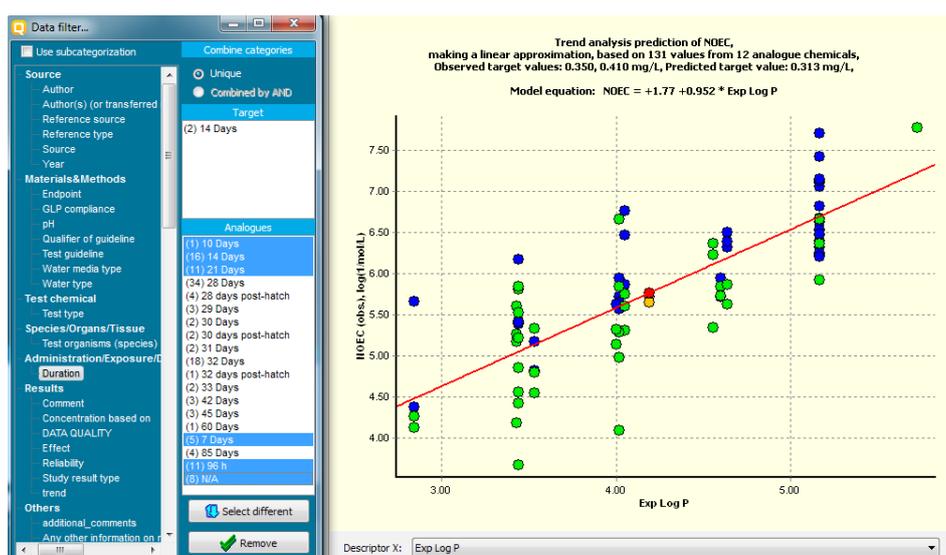


Figure 11. Data points identified with less than 28 days test duration are shown as green dots and are removed in this example.

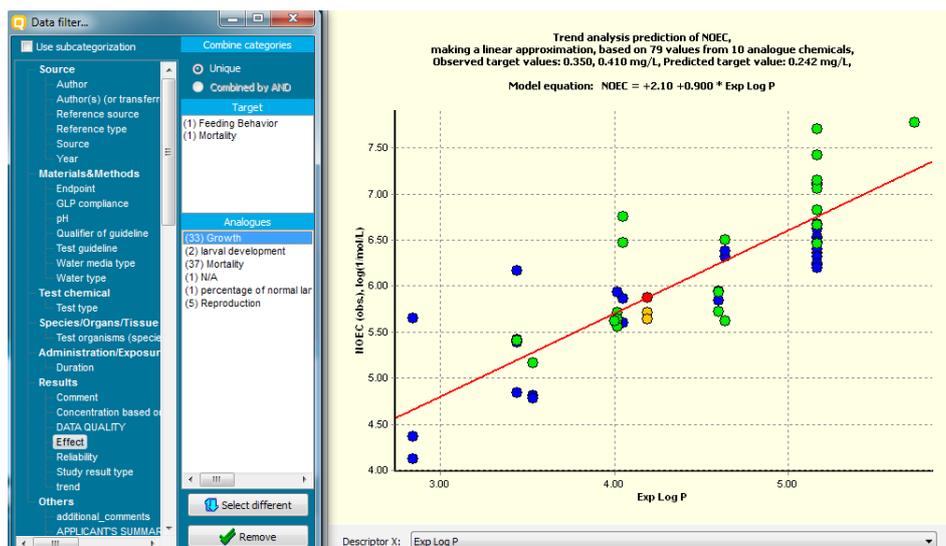


Figure 12. Among the observed effects for this category (listed in the white box on the bottom left of the figure), growth (highlighted as green dots) appears to be the most sensitive one and is used for the trend analysis.

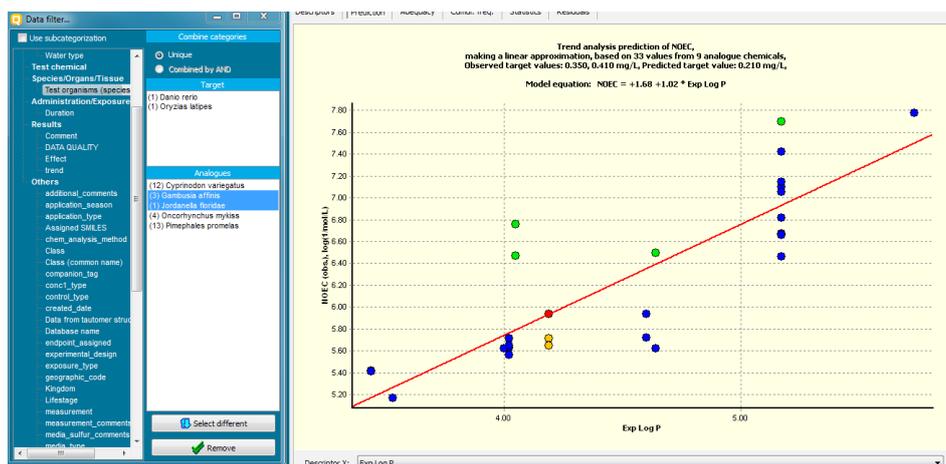


Figure 13. Species not recommended by OECD TGs for fish long-term toxicity (in blue in the white box on the left of this picture) are excluded. Data points derived from tests with *Gambusia affinis* and *Jordanella floridae* are highlighted as green dots. These species appear to be more sensitive than other species for this category but they are excluded because not indicated in OECD TGs.

The category now contains discrete, neutral organic chemicals with same functional groups, with test duration equal or above 28 days and growth as the most sensitive effect (Figure 14). Further details on the experimental conditions can be found in the Toolbox but are not described in this document.

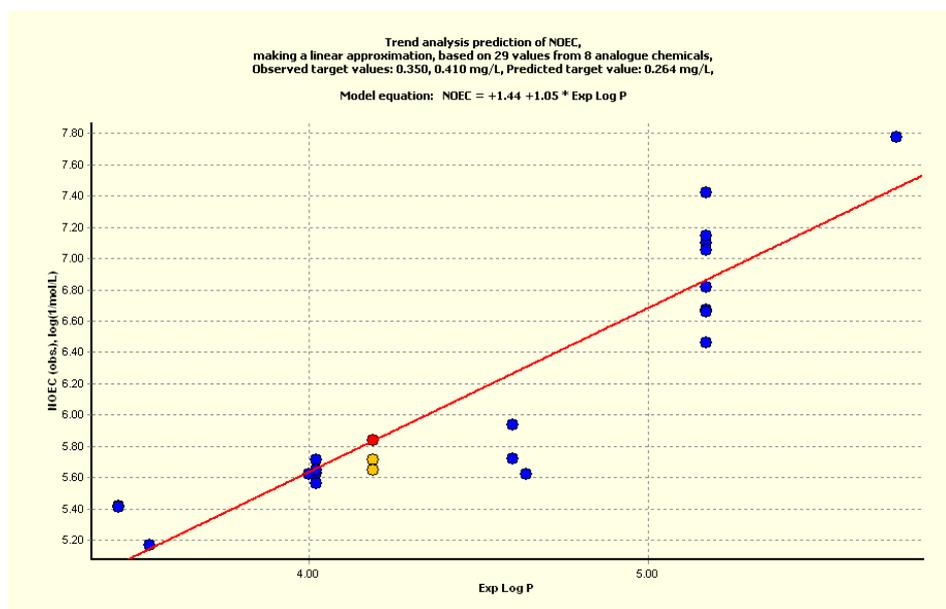


Figure 14. This category contains discrete, neutral organic chemicals with same functional groups, with test duration equal or above 28 days and growth as the most sensitive effect for fish species recommended by OECD TGs for fish long-term toxicity are kept (*Cyprinodon variegatus*, *Oncorhynchus mykiss*, *Pimephales promelas*).

In conclusion, the variability among data from one species with one effect (growth) for pentachloro benzene (logKow ca 5.2) differ already about 1.0 log units (see Figure 15), and further subcategorisation of the data will not lead to less. Hence, in this example the category is not further reduced to a single fish species.

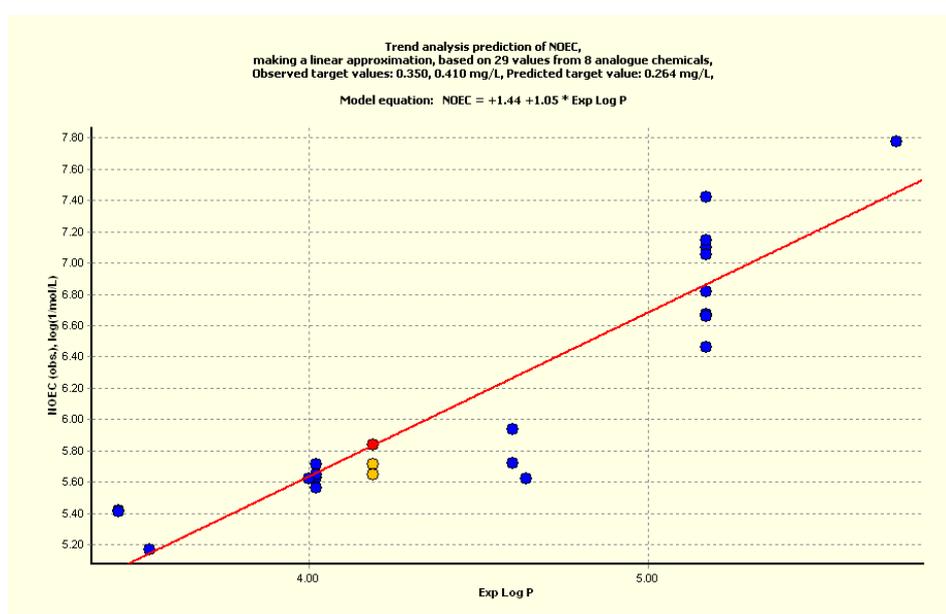


Figure 15. Data points derived from tests with *Cyprinodon variegatus* are highlighted as green dots and illustrate the variability within one species.

A visual inspection of the chemical structures of the analogues (Figure 16) shows that all members of the series are aromatic chlorides whose structures are very similar to that of the target, in fact most of them differ only in the number of Chlorine atoms and their position. There is also one naphthalenic structure, i.e

the chemical with CAS number 90-13-1, however it was not removed during the categorisation steps. It would be possible to remove this chemical on similarity basis, but given that its toxicity follows the trend and the profilers indicate the same (narcotic) mechanism of the target, the chemical 90-13-1 is kept in the analysis. Moreover, narcotic mechanism is not linked to specific structural groups and is therefore not much dependent on chemical similarity. The differences will influence the log Kow and thus the toxicity.

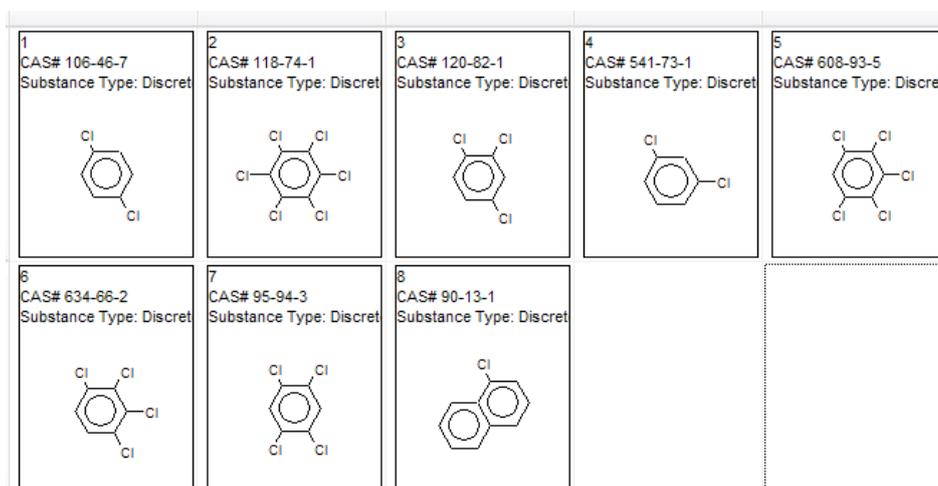


Figure 16. Analogues in the category after subcategorisation according to chemical and endpoint specific profilers and further removing data points according to test conditions as described in the text.

At this point, the analysis is concluded and can be entered in the data matrix by clicking on “Accept prediction”. One value per chemical should be used for the prediction to avoid that chemicals with a higher number of data points will have more weight in the final result. To follow a conservative approach, the data usage can be set to “Maximal” (i.e. for each chemical, maximum toxicity value is used for the trend analysis). The predicted value is 0.250 mg/L (Figure 17) and the QSAR Toolbox - as shown in Table 2 - automatically calculates the statistics related to the identified trend.

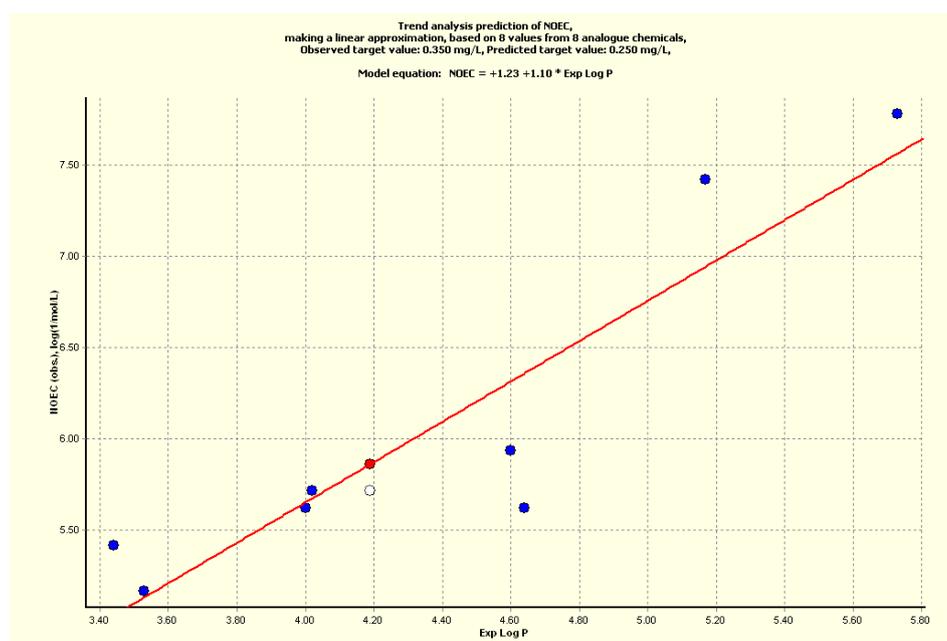


Figure 17. Trend derived from experimental log Kow values

Table 2 Statistic of the final trend used for the prediction

Statistical characteristics	TA model
Number of data points, (N)	8
Coefficient of determination, (R ²)	0.829
Adjusted coefficient of determination, (R ² _{adj})	0.801
Coefficient of determination - leave one out, (Q ²)	0.711
Coefficient of correlation for external set, (r ²)	-
Sum of squared residuals, (SSR)	1.11
Standard deviation of residuals, (sN)	-
Sample standard deviation of residuals, (s)	0.431
Fisher function, (F)	29.2
Fisher threshold for statistical significance, (F _a)	9.35
b0	
- model descriptor	Intercept
- coeff. value	1.23
- coeff. range	± 2.23
- significance	No
- max. covariation	0.994 (vs b1)
b1	
- model descriptor	Exp Log P
- coeff. value	1.10
- coeff. range	± 0.50
- significance	Yes
- max. covariation	0.994 (vs b0)

Table 3 summarises the values used to derive the trend.

Table 3 CAS number, name, experimental log K_{ow} and NOEC of the analogues used for the prediction

CAS number	Name	Experimental log K _{ow}	Experimental NOEC (28d, growth) [mg/L]
106-46-7	1,4-dichlorobenzene	3.44	0.565
541-73-1	1,3-dichlorobenzene	3.53	1.000
90-13-1	1-chloronaphtalene	4.00	0.390
120-82-1	1,2,4-trichlorobenzene	4.02	0.350
634-66-2	1,2,3,4-tetrachlorobenzene	4.60	0.250
95-94-3	1,2,4,5-tetrachlorobenze	4.64	0.520
608-93-5	1,2,3,4,5-pentachlorobenze	5.17	0.095
118-74-1	1,2,3,4,5-esachlorobenze	5.73	0.005

Data quality also needs to be addressed before accepting the predicted 28 days NOEC value (0.250 mg/L) by checking the metadata and underlying data of the original reference of each data point used for the prediction (not shown in the example). The QSAR Toolbox includes data as they are donated without further curation. This step has not been performed for the current example.

REPORT AND CATEGORY SUMMARY

At the end of the process, the user can decide if the result is satisfactory. Once the prediction is accepted for reporting, the “report” module of the QSAR Toolbox gives the possibility to semi-automatically generate

a standalone file to record the result with a pre-defined template, providing information on the modelling procedure (applicability domain, information about the group members, etc.). To generate a report for the prediction, the user has to open in the module "Report" the Predictions in the window "Available data to report" by double clicking, selecting the specific prediction, and then clicking the button "Create". Alternatively, the user can generate a report directly from the data matrix by hovering over the prediction and clicking the right mouse button. Selecting "Report" from the pop-up menu will generate a report, selecting IUCLID can be used to produce and attach a report directly to a IUCLID template file. The prediction can be reported as a "QSAR" study type.

The report also contains manually editable fields that the user should fill to justify the procedure that was followed (i.e., category definition, hypothesis, and description). Here, we summarise some of the considerations needed for the assessment of the category and the prediction. They are based on good practices and expert judgement and are not hard-coded in the QSAR Toolbox as such. To support and justify the category hypothesis, it is highly recommended to also include a data matrix (not shown in this example) reporting physico-chemical, fate and other (aquatic) toxicity properties of the category members.

a. Category definition

A category of 9 analogues (a target and 8 source chemicals) is defined for the purpose of data-gap filling of fish long-term toxicity (as 28 days NOEC for growth). The analogues are shown in Figure 17. In this example, considerations on impurities are omitted. For REACH substances, impurities need to be analysed and discussed.

b. Hypothesis for grouping chemicals

The working hypothesis is that the selected analogues can be grouped together for the prediction of fish long-term toxicity. The hypothesis is based on the profilers outcome which support the assumption that the analogues share the same mode of action for aquatic toxicity – the non-polar narcotic effect.

c. Category description

- The category members are phenyl chlorides and a naphtyl chloride with variable number and position of the chlorine atoms. The category members do not present any additional functional groups. For each analogue, a NOEC value for fish growth in long-term exposure is available. The NOEC values are expressed in mg/L.
- The covered hydrophobicity range in terms of log Kow is between 3.44 and 5.73 (based on log Kow experimental values available in the QSAR Toolbox). The log Kow of the target chemical is 4.19 (experimental) or 3.93 (calculated). The two values are similar to each other and both fit in the range defined by the eight analogues. Solubility covered by the analogues is from 0.344 to 104 mg/L. (based on calculated solubility values available in the QSAR Toolbox). We acknowledge the range is broad but the trend looks consistent over the whole range. The solubility for the target is 26.1 mg/L (calculated). The value fits in the range defined by the analogues. Chemicals tested above their water solubility were removed too. The molecular weight of the target is 181 Da and fits into the range defined by the analogues (from 147 to 285 Da).

d. Strategy used (data-gap filling method)

Predicted Value (NOEC 28d)= 0.250 mg/L

Quantitative trend analysis. The regression equation is:

$NOEC = 1.23 + 1.10 * \text{Exp log Kow}$.

The NOEC is presented in $\log(1/\text{NOEC mol/L})$.

Number of data points (N) = 8

Statistical characteristics of the model (exhaustive list in Table 2)

Coefficient of determination (R^2) = 0.829

Coefficient of determination, leave-one-out (Q^2) = 0.711

Sum of square residuals (SSR) = 1.11

Sample standard deviation of residuals (s) = 0.431

Mechanistic explanation and applicability domain: provided at points b and c, respectively.

ADDITIONAL WEIGHT OF EVIDENCE CONSIDERATIONS AND CONCLUSIONS

Read-across can provide additional confidence to the prediction within a weight of evidence context. Looking at the analogues with relevant data selected for the trend analysis, the structurally closest structure to the target with relevant long-term toxicity studies is the 1,2,4-Trichlorobenzene (CAS 120-82-1), which has the same amount of chlorine atoms and only differs in the position of the Chlorine atoms (structural isomer). This chemical has a NOEC of 0.350 mg/l and can be used as a source in a one-to-one read-across in a weight of evidence context (not shown in this example).

There is an OECD 204 (Fish, Prolonged Toxicity Test) study available for the target, conducted by the Ministry of the Environment in Japan (1997), with a result of 0.35 mg/L (*Oryzias Latipes* NOEC mortality at 14 days). Although sometimes referred to as a long-term study, it does not cover the REACH standard information requirement on its own. However, it can be used in a weight of evidence approach. The other available value (NOEC > 0.41 mg/L) was taken from an unpublished report of the German Federal Health Office (Danio Rerio feeding behavior at 14 days) and thus it is not possible to assess the reliability of the study.

The trend analysis using 8 structurally similar chemicals resulted in a 28 days growth NOEC of 0.250 mg/L.

The lowest NOEC is the one predicted by the trend analysis (0.250 mg/L) and, following a conservative approach, this is the value selected. In addition, the result is supported by the experimental NOEC value of the closest analogue 1,3,5-Trichlorobenzene (0.350 mg/L) and by the experimental NOEC value at 14 days (0.35 mg/L).

EUROPEAN CHEMICALS AGENCY
ANNANKATU 18, P.O. BOX 400,
FI-00121 HELSINKI, FINLAND
ECHA.EUROPA.EU

ECHA-14-R-19-EN - ED-02-14-81-EN-N - ISBN: 978-92-9244-752-6 - DOI: 10.2823/52268