Concawe

Environmental Science for European Refining

Report

Report no. 7/20

PetroTox version 4.0 User Manual

ISBN 978-2-87567-118-9

PetroTox version 4.0 User Manual

E. Vaiopoulou (Concawe Science Executive)

This report was prepared by:

Borja Serrano, Yves Verhaegen, Klaas den Haan, Louise Camenzuli, and David Saunders, A. Martin Aparicio with support of **Penman Consulting Ltd.**

Under the supervision of:

E. Vaiopoulou (Concawe Science Executive)

At the request of:

Concawe's Environmental Management Group (EMG)

Thanks for their contribution to: Aaron Redman

Reproduction permitted with due acknowledgement

© Concawe
Brussels
April 2020

ABSTRACT

PetroTox is a modelling tool implementing the Target Lipid Model (TLM) for the calculation of the aquatic toxicity of complex hydrocarbon mixtures. The model allows for the calculation of acute and chronic toxicity values for several trophic levels, as well as Predicted No-Effect Concentrations (PNEC). The previous implementation in an excel sheet has been migrated to the open source data mining platform KNIME, providing several advantages in terms of performance, transparency and potential for connecting to external tools, while offering the same options and results that the previous version. The herein text serves also as a manual for the use of the PetroTox too.

KEYWORDS

Critical Target Lethal Body Burden (CTLBB), Hydrocarbon Block (HB), (KNIME), Predicted No Effect Concentration (PNEC), Target Lipid Model (TLM), Target Lipid-Water Partition Coefficient (KTLW).

INTERNET

This report is available as an Adobe pdf file on the Concawe website (www.concawe.eu).

CONTENTS

		Page
1.	INTRODUCTION TO PETROTOX V4.0	1
2.	CHANGES FROM SPREADSHEET PETROTOX V3.06	2
3.	QUICK START	3
4.	CONCEPTUAL DESIGN	9
4.1.	COMPOSITION AND MAPPING	9
4.2.	THREE PHASE DISTRIBUTION (SOLUBILITY) MODEL	10
4.3.	TARGET LIPID MODEL	11
4.3.1.	Bioavailability Correction	13
4.3.2.	Predicted No Effect Concentration (PNEC)	13
5.	MODEL DESCRIPTION	15
5.1.	OVERVIEW	15
5.2.	INPUT MASSES	15
5.3.	DATA USED TO BUILD THE MODEL	16
5.3.1.	Species list (CTLBB)	16
5.3.2.	Concawe Library	16
5.4.	SAVING RESULTS	19
6.	RUN TIME	20
7.	GLOSSARY	21
8.	REFERENCES	22
APPENDIX: ABOUT KNIME		23

List of Figures

Figure 1. The KNIME PetroTox v4.0 workflow. 4
Figure 2. Screenshot of the empty 'Input masses' worksheet of the PetroTox v4.0 High Resolution input Excel workbook. 5
Figure 3. The 'Select calculation, bioavailability and endpoint mode' node and the 'Select calculation mode' dialog window. 6
Figure 4. The 'Input masses' node and the 'List files (Browse to the)' dialog window. 6
Figure 5. The 'Input loadings' node and the 'loading settings' dialog window. 7
Figure 6. The 'Input test conditions' node and the 'test condition settings' dialog window. 7
Figure 7. The 'Select a species' node. 8
Figure 8. The KNIME toolbar with the "Execute all" button highlighted. 8
Figure 9. The different meta-node states. Ports on the left of the node icon are input ports, the amber, red and green dots graphically represent the output ports. 8
Figure 10. The 'Output Summary' node. 8
Figure 11. The 'Concawe library' table creator node. 16
Figure 12. A detail of the Concawe (constituent) library table creator settings. 17
Figure 13. The 'Chemical classes' QuickForm node and detail of the 'Chemical classes' table. 17
Figure 14. The 'Substances selector' QuickForm node and detail of the 'Substances selector' table. .. 18
Figure 15. The 'PNEC calculation variables' metanode. 18
Figure 16. The 'log(KMW) cut off and slope' metanode. 18
Figure 17. The 'Export KNIME workflow(s)' dialog box, with the check box to reset the workflow(s) before export highlighted. 19

1. INTRODUCTION TO PETROTOX V4.0

PetroTox v4.0 is an implementation of the PetroTox model in the KNIME analytics platform. It implements the model from the PetroTox v3.06 Excel workbook and provides equivalent results. KNIME provides a more visual, traceable, and extendable environment for the implementation of the PetroTox model. The data processed by KNIME is visible at every step of the calculations, which allows for:

- analysis of intermediary data,
- flexible support of experimental datasets (solubility and toxicity),
- increased transparency for communicating results to stakeholders, and
- easier debugging.

PetroTox is designed to calculate the toxicity of petroleum products to aquatic organisms. Petroleum hydrocarbons are complex mixtures that exert a narcotic mode of toxic action, which is assumed to be additive. This model calculates the solubility of a petroleum product and then uses the TLM [1] and toxic unit (TU) theory of additivity to calculate the toxicity or environmental risk limits of these mixtures.

The model uses a three-phase (petroleum product, water, air) oil solubility calculation [2] that is coupled with a database of physical and chemical properties of typical petroleum hydrocarbons to calculate the distribution of petroleum hydrocarbons among the exposure water, headspace and free product phases. The properties database was developed by Concawe and contains physical/chemical properties such as boiling point (BP), solubility and octanol-water partition coefficients (K_{ow}) for 1512 hydrocarbon structures that are possibly found in petroleum products (some are hypothetical structures). This version of the PetroTox model performs the solubility and toxicity calculations with physiochemical properties derived from the database for each hydrocarbon block.

The computed dissolved hydrocarbon concentrations are then used by the TLM to calculate aquatic toxicity. A modification to the TLM is the use of membrane-water partition coefficients (K_{mw}) rather than octanol/water partition coefficient K_{ow} to describe the partitioning between the water phase and the organism [3]. This modification is critical for petroleum products that have very hydrophobic compounds ($\log K_{ow} > 6.0$).

2. CHANGES FROM SPREADSHEET PETROTOX V3.06

Most changes to the original Excel PetroTox v3.06 model are in the form of interaction with the system. The results produced by both models should be near identical. The changes implemented are:

- Two additional approaches to reassign missing compositional mass: (1) distribute the remaining mass to the last blocks of all classes [appropriate for 'heavy' unknown or variable composition, complex reaction products of biological materials (UVCBs)], and (2) distribute the remaining mass to the first blocks of all classes (appropriate for 'light' UVCBs) (see Section 5.2).
- The possibility to select just some classes for the calculation as explained the previous section (see Section 5.3.2).
- The possibility to select the way to calculate the PNEC [Hazardous Concentration that affects 5% of species (HC5 calculation)] (see Section 5.3.2).

3. QUICK START

1. **Open KNIME and import the workflow**
   1. After launching KNIME for the first time you will be asked to create a workspace. Please create it on the local drive and not on a network folder if possible, otherwise you could have problems loading the workflow. Then **load the preferences** from the file "PetroTox_4.0_Preferences.epf" via: **"File → Import preferences"** which will change the level of log displayed.
   2. Open the PetroTox workflow via **"File-> Import KNIME Workflow…"**. Click on the first **"Browse…"** option at the right of "Select file…" input box and select the file with the workflow that you want to use (low or high resolution mode) which will be "PetroTox_4.0_High_resolution.knwf" or "PetroTox_4.0_Low_resolution.knwf". These steps won't be needed in further executions of the model as the workflow will appear directly in KNIME.

2. **Prepare input mass file**
   3. Open the **Excel template file** "PetroTox_4.0_High_resolution_Input_template.xlsx" or "PetroTox_4.0_Low_resolution_Input_template.xlsx" depending on the model you want to use (low or high resolution mode).
   4. Insert **the substance name** in cell B3 and select the **"Mass imbalance"** mode in cell C3 in case the masses do not sum 100% ("Scale to 100%", "Add to non-toxic n-Paraffin/aliphatic", "Distribute across highest blocks" and "Distribute across lowest blocks").
   5. **Input** the corresponding **masses** as a percentage of the total in cells E6:T31 (high resolution) or E6:F31 (low resolution).
   6. In cell C33 you will see the **sum of the masses**. In general, they must sum **100%**. In case the masses sum less or more than 100% you will see a red message under the table saying e.g. "Masses do not sum 100%. Base on the Mass imbalance value (C3) the calculations will: Scale to 100%".
   7. **Save the Excel document** with a name associated to your substance (for instance "My_substance_INPUT.xlsx") on a local folder of your machine (for instance "C:\PetroTox Inputs")
   8. You can **repeat the process** saving several files **in the same folder** in case you want to launch the model for different substances or under different "Mass imbalance" settings. This would be necessary for a batch mode calculation.

3. **Select calculation, bioavailability and endpoint mode**
   Double-click on the node "Select calculation mode". If you want to change the value of any of them just click on the **"Change"** checkbox, select the new value and click "OK". If you select **"Enter particulate organic carbon (POC) data"** or **"Enter lipid concentration"** under **"Bioavailability mode"** you will have to set the values in step 3 (Input Product Loading / POC data).

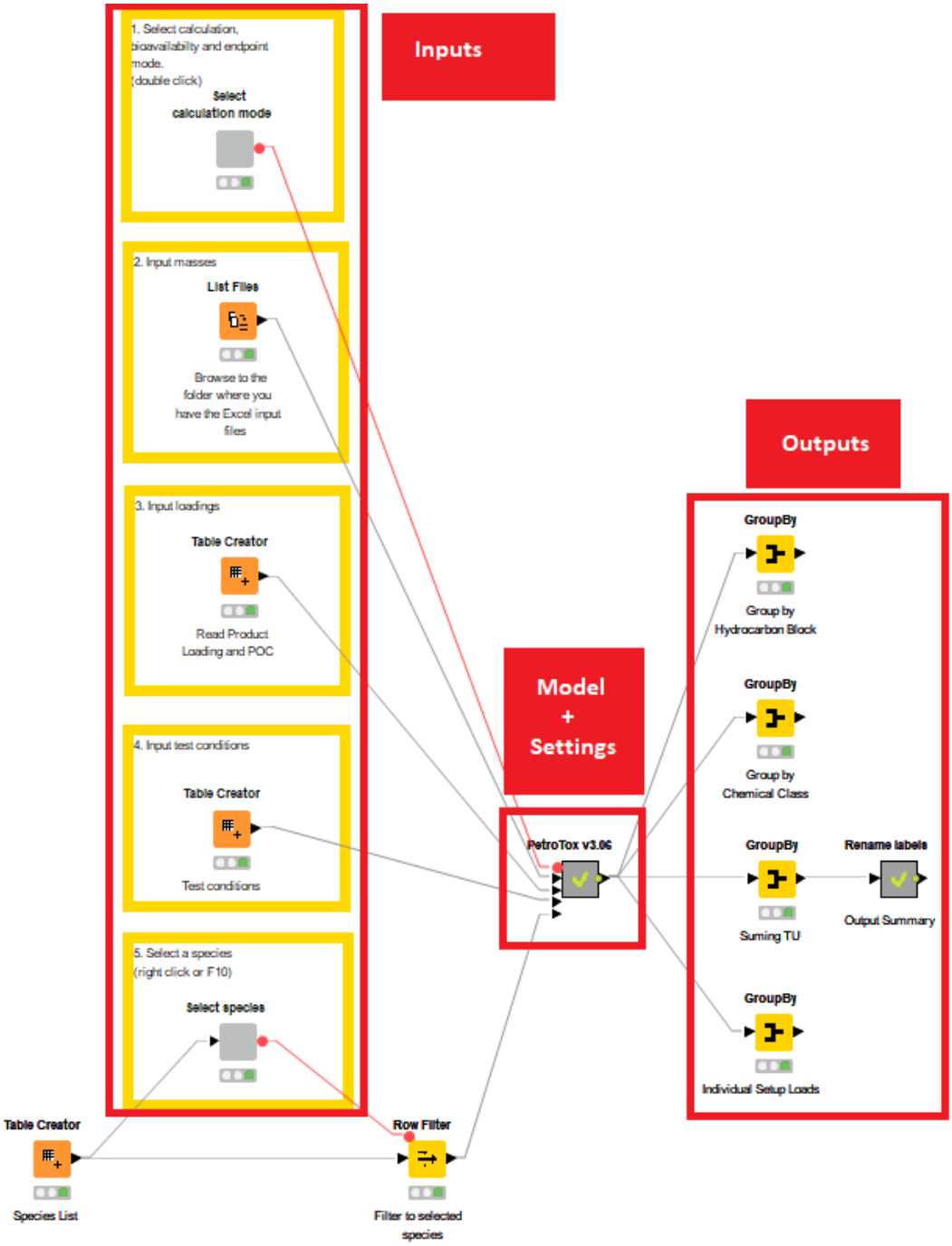

Figure 1. The KNIME PetroTox v4.0 workflow.

Substance name	Mass Imbalance
SUBSTANCE_NAME	Scale to 100%

Hydrocarbon Block	Starting Carbon Number	Ending Carbon Number	n-P	i-P	n-CC5	n-CC6	i-N	Di-N	n-Olefins	i-Olefins	Poly-N	AlS	MoAr	NMAr	DiAr	NDiAr	PolyAr	ArS
1	5	6																
2	6	7																
3	7	8																
4	8	9																
5	9	10																
6	10	11																
7	11	12																
8	12	13																
9	13	14																
10	14	15																
11	15	16																
12	16	17																
13	17	18																
14	18	19																
15	19	20																
16	20	21																
17	21	22																
18	22	23																
19	23	24																
20	24	25																
21	25	26																
22	26	27																
23	27	28																
24	28	29																
25	29	30																
26	30	31																

Sum of masses	0	Masses do not sum 100%. Based on the Mass imbalance value (C3) the calculations will: Scale to 100%

Input masses

Figure 2. Screenshot of the empty 'Input masses' worksheet of the PetroTox v4.0 High Resolution input Excel workbook.

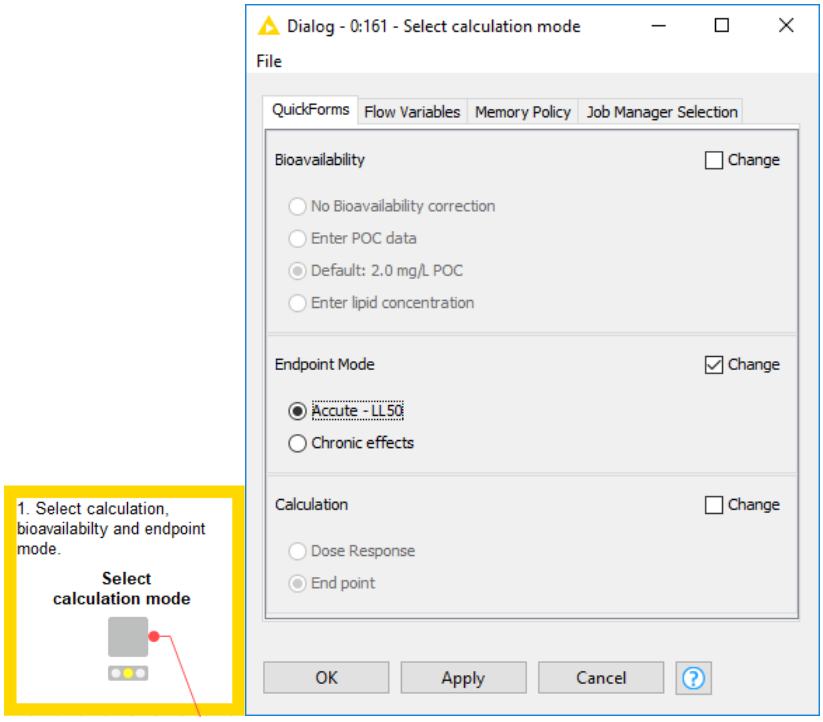

Figure 3. The 'Select calculation, bioavailability and endpoint mode' node and the 'Select calculation mode' dialog window.

4. **Input masses**

   Select the folder where you have saved the "Input masses" spreadsheet(s). **Double-click** on the node "List files". Click on the "**Browse**" button to select the folder where you have saved your input files.

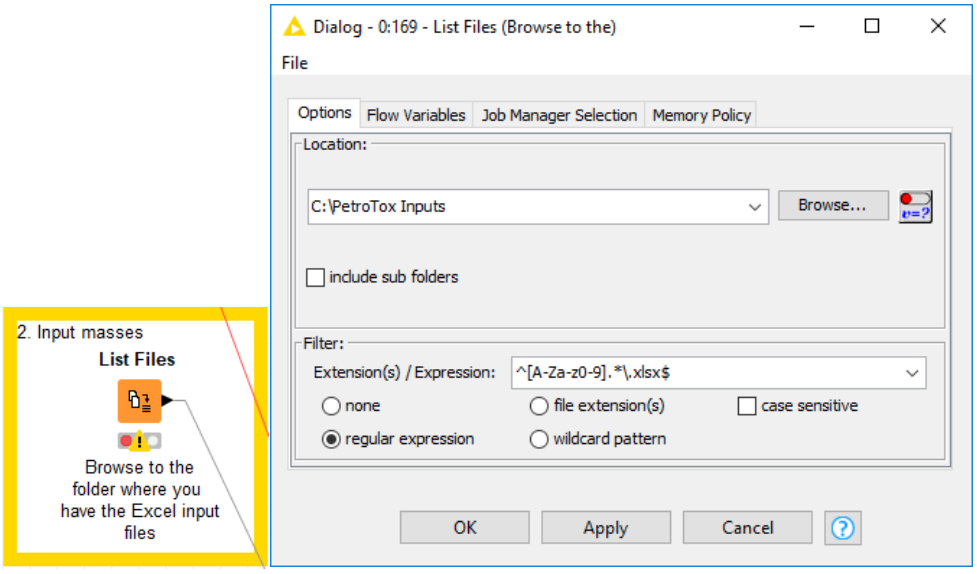

Figure 4. The 'Input masses' node and the 'List files (Browse to the)' dialog window.

5. **Input loadings**
   Double click on the node "**Table creator**" to input the product loading and the POC/data. Enter the desired loading settings. Please note that the "**product loading**" will only apply on **Dose response calculations**. Please note as well that the **POC or Lipid input will be set to 2 if the default Bioavailability** mode is selected on the first step of the KNIME model. More rows can be added to perform multiple runs with different settings for dose response calculations. To add more rows simply start typing the values into *Row1*.

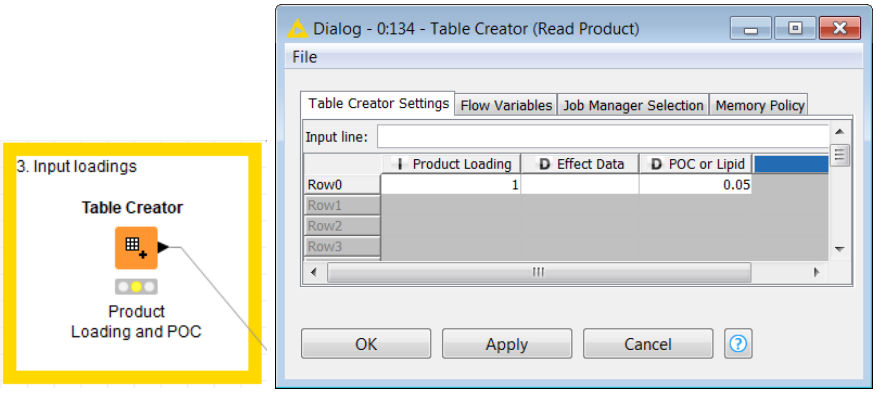

Figure 5. The 'Input loadings' node and the 'loading settings' dialog window.

6. **Input test conditions**
   Double click on the "**Table Creator**" node to input the test conditions. Input the volume of water and air. The default values of 0.9L of exposure water with a 0.1L headspace result in a total system volume of 1.0L, which is equivalent to a system with a 10% headspace. The minimum headspace volume can be 0.0L. The headspace can be an important phase for particularly volatile petroleum products like gasoline products where much of the product load volatilizes to the headspace.

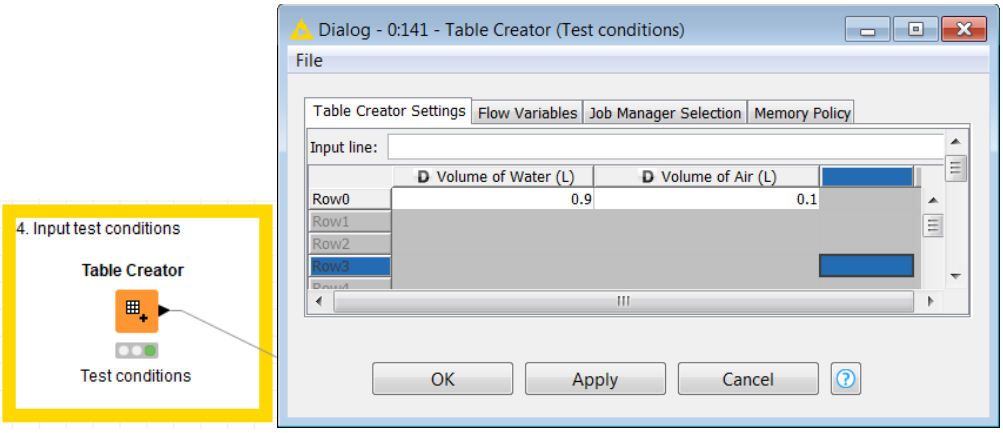

Figure 6. The 'Input test conditions' node and the 'test condition settings' dialog window.

7. **Select a species**
   Double click on the "**Select Species**" node to select the species-specific CTLBB. If the CTLBB for a desired species is not available, a user-defined CTLBB can be entered. If you want to change the value just click on the "**Change**" checkbox, select the new value and click "OK". (Optional) If a custom CTLBB is required the data can be manually entered by double clicking

on the **"Species List"** node and editing the **"USER-DEFINED"** row as needed. *Note:* the **"USER-DEFINED"** option will also have to be selected from the **"Select Species"** node.

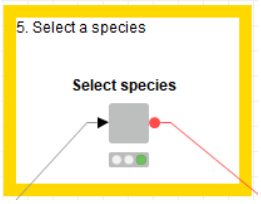

Figure 7. The 'Select a species' node.

8. **Run PetroTox**
Press the **"Execute all"** button, or press **"Shift+F7"**. You will see the **workflow icons** changing from a clock (paused) to a "fast forward" icon (in execution) when the "execute all" button is clicked. When the model run is finished, a check icon (finished task) will appear on all nodes previously showing "in execution". If the node changes to "paused" after the "execute all is clicked, then an error has occurred, and the model needs to be restarted. If there is a problem during the calculation or you want to **restart** the flow because you want to change some input values you can reset the workflow **selecting all the nodes** (Ctrl+A or with the mouse) and pressing F8 or clicking on the **"Reset"** button ( ) on the toolbar.

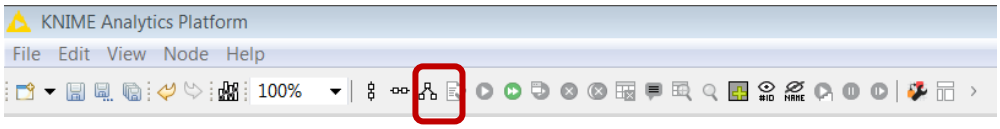

Figure 8. The KNIME toolbar with the "Execute all" button highlighted.

Figure 9. The different meta-node states. Ports on the left of the node icon are input ports, the amber, red and green dots graphically represent the output ports.

9. **View output**
Right click on the "Output Summary" node and select **"Connected to: Output Table"** option (at the bottom).

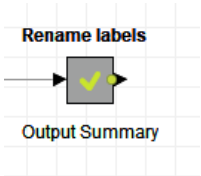

Figure 10. The 'Output Summary' node.

4. CONCEPTUAL DESIGN

4.1. COMPOSITION AND MAPPING

The model can accommodate two types of inputs, low resolution or high resolution, which depend on the information known about the mass distribution of the petroleum product. In the low-resolution approach, the mass distribution of two general chemical classes (aliphatic and aromatic) is entered over user-defined BP intervals. This format is patterned after the information given from simpler hydrocarbon block analyses. In the high-resolution approach, the mass distribution for up to 16 chemical classes can be entered: n-paraffins, iso-paraffins, n-substituted cyclohexanes, n-substituted cyclopentanes, other mono-naphthenics, di-naphthenics, poly-naphthenics, n-olefins, iso-olefins, sulphur-bearing aliphatics, mono-aromatics, naphthenic mono-aromatics, di-aromatics, naphthenic di-aromatics, poly-aromatics and sulphur-bearing aromatics. This format is patterned after the information derived from highly detailed 2-dimensional Gas Chromatography (GCxGC) that provides mass distribution information for the 16 chemical classes over discrete carbon number (C#) intervals. These hydrocarbon blocks (C# and chemical class) are used for environmental fate and effects estimates for complex petroleum substance. The Concawe database contains structures for each of these 16 chemical classes over a range of physicochemical properties.

The library structures are assigned to a hydrocarbon block and the mass fraction given to that block is evenly distributed among all of the assigned structures. For example, a hydrocarbon block in the iso-paraffin class is assigned a mass fraction of 5.0% with an initial carbon number (CNO) interval of C9-C10. The Concawe library is sorted for structures within the iso-paraffin class that fall within the defined CNO interval. In this case there are eight iso-paraffins in the database within this CNO interval (ID: 676, 677, 678, 679, 681, 682, 683, 684). The mass fraction of this hydrocarbon block (5.0%) is evenly distributed among the candidate structures so that each has 0.625% of the total mass. The mass fraction that is assigned to each structure is later used to determine the mole fraction of a given hydrocarbon, which impacts the aqueous solubility of that structure.

If a hydrocarbon block has a chemical class and/or CNO interval configuration that does not identify any candidate structures in the database, the mass fraction of this so-called ‘orphan’ block is reassigned to a neighbouring block. The unassigned mass in these ‘orphan’ blocks is added to a neighbouring block generally within the same chemical class where there are structures available for assignment depending on the BP configuration of the block. If the CNO interval for the ‘orphan’ block is above C14 in high resolution mode, or the BP interval above 275.2°C in low resolution mode, the unassigned mass is added to the first available hydrocarbon block at higher BP intervals. This is done to account completely for all of the mass at any given loading. ‘Orphan’ blocks below this BP value are added to the first available hydrocarbon block at a lower BP interval. This is based on the distribution of hydrocarbon structures in the database. In a few cases where there are no hydrocarbon structures within the same chemical class available for the mass reassignment, the unassigned mass is added to the first available chemical class to the left (unassigned PolyAr mass added to structures in NDiAr block, etc. See Figure 2) within the same BP interval. The mass in orphan blocks from the n-Paraffin class is reassigned to the available block to the right (unassigned n-Paraffin mass added to iso-Paraffin mass).

4.2. THREE PHASE DISTRIBUTION (SOLUBILITY) MODEL

The following equations describe a multi-phase (petroleum product, water, air) dissolution model that considers the changes in product volume and composition as a result of dissolution. Several models for computing equilibrium concentrations in multiphase systems have been presented elsewhere [4] and only a brief outline of the equations are presented here. The mass balance for component i in the system is:

$$m_{T,i} = m_{G,i} + m_{W,i} + m_{A,i} \quad (1)$$

where $m_{T,i}$ is the total mass of component i in the system and $m_{G,i}$, $m_{W,i}$, and $m_{A,i}$ refer to the mass of the component i in the product, water, and air phases, respectively. The mass balance can be expressed in terms of volume and concentration of component i:

$$m_{T,i} = V_G C_{G,i} + V_W C_{W,i} + V_A C_{A,i} \quad (2)$$

Where

$C_{G,i}$= petroleum product phase concentration of i, mol/L
$C_{W,i}$= water phase concentration of i, mol/L
$C_{A,i}$= air phase concentration of i, mol/L
V_G= volume of petroleum product, L
V_W= volume of water, L
V_A= volume of air, L

Assuming petroleum products are ideal mixtures [5], the equilibrium between the product-phase concentration and the total water-phase concentration ($C_{W,i}$) is described by Raoult's Law:

$$C_{W,i} = x_i S_i \quad (3)$$

where x_i is the mole fraction of component i in the product phase and S_i is the aqueous solubility of component i. The concentration of monomeric species is calculated in a similar manner:

$$C_{D,i} = x_i A_i \quad (4)$$

where x_i is the mole fraction of component i in the product phase and A_i is the aqueous activity (e.g., monomer solubility) of component i. For many compounds with $\log(K_{OW})$ less than approximately 5.5 the activity is equal to the sub-cooled solubility indicating that these compounds do not form polymeric species. for some compounds with $\log(K_{OW})$ greater than 5.5 the activity can be a very small fraction of the sub-cooled solubility (< 0.001) indicating that the monomer species is a small fraction of the total dissolved chemical.

The Henry's Law describes the equilibrium between the air phase concentration and the water phase concentration of the monomer species:

$$H_i = C_{A,i}/C_{D,i} \quad (5)$$

where H_i is the unitless Henry's Law constant for component i.

The relationship between the product phase concentration on a mole fraction basis and a mol/L basis is via the petroleum product density, ρ_G (mol/L):

$$C_{G,i} = x_i \rho_G \quad (6)$$

Substituting equations 3, 4, 5, and 6 into equation 2 and solving for the mole fraction yields:

$$x_i = \frac{m_{T,i}}{(V_G \rho_G) + V_W S_i + V_A H_i A_i} \quad (7)$$

There is one additional equation needed to describe the system. The mole fractions of the components in the petroleum product sum to 1:

$$\sum_i x_i = 1 \quad (8)$$

There are a total of $n + 1$ equations and $n + 1$ unknowns (n x_i, and the product of $V_G \rho_G$). This system of equations is solved numerically as described below. The inputs are the total mass ($m_{T,i}$) the solubility (S_i), Henry's Law constant (H_i) and density (ρ_G) of each component i, as well as the volumes of air and water in the system (V_A and V_W).
The strategy is to minimize the error in Equation 9. The procedure is coded into a KNIME workflow. The starting value for $V_G \rho_G$ (the mass of petroleum product) is obtained by assuming all the mass in the product phase:

$$(V_G \rho_G) = \sum_i m_{T,i} \quad (9)$$

An iterative method is used to vary the V_G until the overall sum of the working mole factions is 1.0.

4.3. TARGET LIPID MODEL

For the risk assessment of petroleum-derived products, it is accepted that the mode of action of petroleum related hydrocarbon components is via narcosis [6]. Chemicals that act via narcosis are referred to as narcotics. It has been demonstrated experimentally that the effects of narcotics are strictly additive. This is important for petroleum products since they consist of many hydrocarbons. The toxic unit approach is used to express the toxicity of mixtures with components that exert a similar mode of action. The TLM [1] and the TU concept have been adopted for use in computing environmental risk limits for mineral oil [7].

The TLM predicts the aquatic toxicity of narcotic chemicals to a variety of aquatic species, including fish, algae and invertebrates. It is based on the inverse relationship observed between the log(LC50) (lethal concentration to 50% of test organisms) and $\log(K_{OW})$ and the observation by McCarty et al. [8] that this relationship can be understood as the result of a constant body burden of narcotic chemical that causes adverse effects. In the TLM a single universal slope for the $\log(\mathrm{LC50}) - \log(K_{OW})$ relationship has been demonstrated, independent of species. The TLM equation for computing the LC50 (mmol/L) for a particular chemical for a specific species is:

$$\log \mathrm{LC50} = -0.936 \log(K_{OW}) + \Delta c + \log C_L^* - \log \mathrm{ACR} \quad (10)$$

where Δc_ℓ is the chemical class correction factor for chemicals that were found to be slightly more potent than chemicals acting via a baseline toxicity mode of action

[2] [e.g.: polyclic aromatic hydrocarbon (PAHs)], -0.936 is the universal slope, and C_L^* is the CTLBB for narcosis effects (μmol/g octanol) and ACR is the acute-to-chronic ratio. The ACR is used to convert acute effects (e.g., LC50, ACR of 1) to chronic effects levels [e.g., no observed effect concentration (NOEC), LC10, ACR of 3.83]. The TLM has been recalibrated with current $\log(K_{OW})$ values extracted from SPARC v4.2 in May 2008 [9]. The revised TLM parameters are provided in the KNIME node labelled 'CTLBB' for 42 organisms including daphnids, fish, algae and other species. The parameters include the CTLBB, universal narcosis slope, chemical class corrections and their associated standard errors. Chemical classes that require a correction include the mono-, di- and polyaromatic hydrocarbon classes. The acute-to-chronic ratio can be applied here to calculate chronic effect concentrations.

For application to petroleum products, the term LL50 (lethal loading concentration) is used instead of the LC50 due to the insoluble nature of most petroleum products. Additionally, for application to petroleum products, the K_{TLW} (target lipid-water partition ratio) is used in place of K_{OW} as follows:

$$\log K_{TLW} = \begin{cases} 0.947 \log K_{OW} + \Delta c & ; for\ \log K_{OW} < 6 \\ 6 + \Delta c & ; for\ \log K_{OW} \geq 6 \end{cases} \quad (11)$$

The overall result is that the maximum $\log K_{TLW}$ is 6.0 even for structures with very high $\log K_{OW}$ values. This adjustment dramatically limits the impact that extremely hydrophobic hydrocarbons have on the calculated toxicity. Without the adjustment, the predicted LL50s were orders of magnitude lower than observed LL50s indicating that the model overestimated the toxicity. The reader is referred to Verbruggen [7] and Redman et al. [3] for additional details on the derivation of the K_{TLW} cutoff. There is an option available to the user where the cutoff and slope of the cutoff can be customized to meet specific the needs of a specific application.

The toxicity of the petroleum product (a combination of multiple hydrocarbons) can be expressed using the concept of Tus [10]. A TU is defined as the ratio of the exposure concentration of the freely dissolved chemical (equation 4) to the effect concentration for a specific medium (e.g. water). The monomer species is assumed to be the bioavailable form of the total dissolved hydrocarbons. A TU posed by each specific hydrocarbon block *i* in a mixture is computed as:

$$TU_i = \frac{C_{W,i}}{LC50_i} \quad (12)$$

where $C_{W,i}$ is the aqueous freely dissolved concentration of the hydrocarbon block (mmol/L) and $LC50_i$ is the aqueous effect concentration (mmol/L) for that block for a specific species as derived from equation 12 or 13. To compute the toxicity of the mixture, the TUs for all hydrocarbon components in the mixture are summed. When the sum of the TUs for the mixture equals one, the mixture is expected to be toxic to 50% of the organisms.

Loading data (aqueous loading, the ratio of oil/water) are not required for the model to calculate median acute or chronic effect endpoint (e.g., LC50). The model starts with an initial guess at the loading and calculates solubility and the subsequent toxic units, iteratively changing the load with each pass until the final toxic units equal 1.0 according to the following definitions:

$$\text{Acute Toxic Units} = 1 = \sum_n TU_n \quad (13)$$

$$\text{Load}\ (i+1) = \text{Load}(i) \frac{1}{TU_i} \quad (14)$$

4.3.1. Bioavailability Correction

Under certain exposure conditions the partitioning of hydrocarbons to the test organisms can reduce the bioavailability of dissolved hydrocarbons. For example, in algal exposures there can be substantial particulate organic carbon (POC) concentrations (1-2 mg/L) from the algal cells. The particulate algal carbon acts as a sorbent for dissolved hydrocarbons. The organic carbon partition coefficient (K_{OC}) and the fraction of dissolved hydrocarbon are calculated following equilibrium partitioning approach (EqP) [11]:

$$\log(K_{OC}) = 0.00028 + 0.983 * \log(K_{OW}) \quad (15)$$

$$\text{Fraction dissolved} = 1/(1 + \text{POC} + K_{OC}) \quad (16)$$

If lipid concentrations are entered with the corresponding biodilution option selected, then the $\log(K_{MW})$ value is used to calculate the fraction of dissolved hydrocarbons in a similar manner:

$$\text{Fraction dissolved} = 1/(1 + \text{Lipid Concentration} + K_{MW}) \quad (17)$$

There are 4 bioavailability correction options in the model. The first option, Option 1, is to proceed with the calculations and assume that no correction occurs. Option 2 allows the user to enter POC concentrations in units of mg POC / L water. Under typical test conditions algal cell density can reach 10^5 cells/ml, which corresponds to a POC concentration near 2 mg/L assuming a cellular C-content of 20 pg/cell [12]. The model will calculate the fraction of dissolved hydrocarbons that is bioavailable according to the POC partitioning model described by equations 16 and 17. Option 3 will perform the same calculation with the default POC concentration of 2 mg/L, representing a typical upper bound. Options 2 and 3 are usually enabled during calculations for algal endpoints but the calculation can be performed for any species in the TLM database, including the user-defined option. If Option 2 is selected but no POC concentrations are entered, the model will proceed with POC concentrations of zero. Option 4 allows the user to enter lipid concentrations in place of POC concentrations. A convenient estimate of the lipid concentration by calculating an organism loading (g wet weight / L water) from the exposure condition and assuming at lipid content of 5% or with other estimation techniques or measurements. The fraction of bioavailable hydrocarbons is calculated with membrane-water partition coefficients as described in equation 19.

4.3.2. Predicted No Effect Concentration (PNEC)

Aquatic PNECs were calculated for each hydrocarbon structure according to the HC5 equation as described in McGrath et al. [2]. The equation was modified to calculate HC5 with target lipid-water partition coefficients (K_{TLW}) instead of K_{OW} as described in equation 11:

$$\log HC_5 = E(m)\log K_{OW} + E(\log C_L^*) + \Delta c_l - E(\log ACR) - k_Z\sqrt{V(m)\log K_{TLW}{}^2 + V(\log C_L^*) + V(\log ACR)} \quad (18)$$

PetroTox v4.0 has the possibility to select an alternative formula to calculate HC5:

$$\log HC_5 = E(m)\log K_{TLW} + E(\log C_L^*) + \Delta c_l - E(\log ACR) - k_Z\sqrt{V(m)\log K_{TLW}{}^2 + V(\log C_L^*) + V(\log ACR) + 2\log K_{OW}C_{OV}} \quad (19)$$

m = universal target lipid slope (-0.936)
HC_5 = hazardous concentration to 5% of test species, mmol/L
C_L^* = mean critical target lipid body burden, µmoles/g octanol
ACR = mean Acute to Chronic Ratio
Δc_l= chemical class correction
k_Z = 95% confidence extrapolation factor
Cov = covariance

where the $E\{\log(C_L^*)\}$ is the logarithmic value of the geometric mean of all critical target lipid body burdens (µmoles/g octanol) in the TLM database (Di Toro et al. [1] with Sept. 2005 recalibration) and has a value of 2.13. By default, the mean ACR for all species in the toxicity database is 3.83 for hydrocarbons [13]. The term 'k_Z' is the 95% confidence sample size-dependent extrapolation factor and has a value of 2.21. The variance in the universal target lipid slope (V(m)) is 2.25E-4 and the variance in the C_L^* ($V\{\log(C_L^*)\}$) is 0.015. The variance in the ACR ($V\{\log(ACR)\}$) is 0.126. In addition, a chemical class-correction of -0.352 for di- and poly-aromatic hydrocarbons and a correction of -0.109 for monoaromatic hydrocarbons are included in these calculations since the compounds exhibit slightly higher toxicity that other hydrocarbon classes that conform to "baseline" ecotoxicity. Additional validation of PetroTox are provided in Redman et al. [3]. All these values can be changed in the workflow in case the variable values need an update.

5. MODEL DESCRIPTION

5.1. OVERVIEW

There are two different workflows available for use depending on the resolution of the available data. The high resolution workflow allows the entry of a detailed product composition data to be used for in the calculations. The resolution is set at 16 chemical classes. The low resolution workflow allows the entry of the product based only on the aliphatic and aromatics classes.

Both workflows have an identical structure (Figure 1):

Input nodes - All required inputs are located on the left side of the workflow and are all required.

Output nodes – The important outputs can be found in the output of node "Output Summary".

Processing node – The processing node contains all the "logic" behind the model and is contained in the "PetroTox v4.0" node in the middle of the workflow.

Data nodes – The workflow contains several predefined datasets which are used in the calculations described previously.

Settings node – The settings for the model are available within the "PetroTox v4.0" node, at the bottom of the workflow. All settings are provided with default values but can be adjusted as needed.

5.2. INPUT MASSES

The 'input masses' Excel worksheet is used to enter product-specific data on mass fraction distribution by chemical class and BP range or carbon number interval (Figure 2). There are two modes for inputting compositional data that correspond to GCxGC (high resolution) and total petroleum hydrocarbons (TPH) (low resolution) analytical approaches. Compositional data for High resolution is organized by discrete chemical class (n=16) and carbon number. Compositional data for Low resolution is organized by total aromatic and total aliphatic general chemical classes and BP. If needed, the user can filter the chemical classes and the constituents used during the calculation (see Section 5.3.2). The orange-shaded section indicates an input range for carbon number or BP intervals. The light blue (cyan) colour indicates a section for entering mass fraction information. The purple shaded section is for entering dose response information: loadings, effect data, or particulate organic carbon concentrations. The pink cells are for entering the dimensions of the exposure chamber: water and headspace volumes. The user may also input date and title information specific to the petroleum product being analysed in the top two cells in column C.

The values entered into each composition input cell, or hydrocarbon block, are entered as weight percent of the total. The model allows for some variability in the overall reported composition of +/- 0.5% out of 100%. This means that the sum of total input mass can be between 100.5 and 99.5% and the model will proceed with the calculations. However, if the total entered mass is >100.5% then the model will stop and report the error in the console. If the mass is less than 99.5%, the weights will be recalculated depending on the user's choice (dropdown cell C3 of the "Product composition" sheet): the model can (1) evenly scale all individual block concentrations to a total input mass of 100%, (2) add the remaining mass to the last

block of the n-Paraffin subclass, (3) distribute the remaining mass to the last blocks of all classes (appropriate for 'heavy' UVCBs), or (4) distribute the remaining masses to the first blocks of all classes (appropriate for 'light' UVCBs). For options (3) and (4), the remaining mass is distributed across the classes based on the total measured concentrations of each class, e.g.:

$$DiAr_{C>30} = DiAr_{C>30} + (100\% - All_{Total}) \times {}^{DiAr_{Total}}/_{All_{Total}}$$

5.3. DATA USED TO BUILD THE MODEL

5.3.1. Species list (CTLBB)

The workflow lists the CTLBB for 42 species that are in the TLM da1tabase in the node "**Species List**". This table can be expanded with new species just adding new lines. The important fields that need to be filled always are "Name" and "Intercept".

5.3.2. Concawe Library

Physical and chemical properties data computed with SPARC for 1512 library structures are stored in this database. The TLM is calibrated with SPARC $\log(K_{OW})$ values and to maintain internal consistency with the workflow model, physicochemical properties extracted from SPARC are used in all solubility and toxicity calculations. All physicochemical parameters correspond to conditions of standard temperature ($25°C$) and pressure. There are several columns of property data: Molecular Weight, BP, Sub-Cooled Solubility, Activity, $\log(K_{OW})$, Henry's Law Constant, and a few other data fields necessary to run the model calculations. Physical/chemical properties in this data collection were calculated with (Environment Protections Agency's (EPA) properties calculator SPARC at the standard temperature and pressure. PNEC values are computed with the HC5 equation.

If **new substances** need to be added to the library, the node which contains the library can be edited. To find it go to the node "PetroTox 4.0" → "Assign masses to library chemicals" → "Concawe library". Then add new rows with the values of the different fields of the table and it will be used in any further calculation.

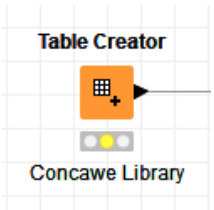

Figure 11. The 'Concawe library' table creator node.

Input line: 123

	CAS	Name	SMILES	Carbon ...	Class	SubClass	General ...	Molar W...	Boiling P...	D
Row 1495	?	Dimethylhept...	C(C(C(CC1)C...	29	PentaN	9	1	398.7140000...	378.9	5.208
Row 1496	?	Isohexyl hyd...	CCC(C)CCC6...	30	HexaN	9	1	410.7250000...	400.279	1.041
Row 1497	?	BENZO(P)NA...	c1cc6cccc7c6...	28	HexaAr	15	2	352.4344000...	706.425	2.406
Row 1498	?	Hydro-BENZ...	C1CC6CCCC...	28	HexaN	9	1	380.6556000...	419.59	4.875
Row 1499	?	DIBENZO(B,K...	c5c6ccccc6cc...	26	HexaAr	15	2	328.4124000...	669.407	2.446
Row 1500	?	Hydro-DIBEN...	C5C6CCCCC...	26	HexaN	9	1	354.6178000...	378.228	2.987
Row 1501	?	n-Butyl dimet...	CCCCC2CC1...	26	PentaN	9	1	356.6336000...	362.066	3.385
Row 1502	?	2,2,4,6,6-pe...	C(C)(C)(C)C...	12	Alkane, bran...	2	1	170.3374000...	208.362	4.317
Row 1503	?	2,2,3-trimeth...	CCCCCCCC(...	13	Alkane, bran...	2	1	184.3642000...	234.644	1.450
Row 1504	?	3,3,6,6-Tetr...	CCC(C)(C)CC...	12	Alkane, bran...	2	1	170.3374000...	228.451	2.038
Row 1505	?	2,2,4,4,6,8,...	CC(C)(C)CC(...	16	Alkane, bran...	2	1	226.4446000...	287.525	2.772
Row 1506	?	6-n-Butyl-2,3...	Cc2cc1ccc(C...	16	Naphthalene	13	2	212.3340000...	349.611	4.564
Row 1507	?	2,3 Dimethy-...	Cc2cc1c(CCC...	18	Naphthalene	13	2	240.3876000...	354.416	6.529
Row 1508	?	2-(4,8,14,18...	C(C)CCCC(C)...	33	Naphthalene	13	2	450.7896000...	456.459	4.600
Row 1509	?	n-phenylcycl...	c(cccc1)(c1)C...	12	Naphthenic M...	12	1	160.2584000...	232.81	1.829
Row 1510	?	cyclopropylb...	c1ccccc1C2CC2	9	Naphthenic M...	12	2	118.1780000...	166.112	1.451
Row 1511	?	2-methyloctyl...	c12ccc3c(c2C...	31	Naphthenic M...	12	2	420.7202000...	433.324	1.752
Row 1512	99-99-9999	My new subs...		12	HexaN	9	1	123	150	
Row 1513										

Figure 12. A detail of the Concawe (constituent) library table creator settings.

You can **filter** the structure classes used during the calculation using the node "PetroTox 4.0" → "Assign masses to library chemicals" → "Chemical classes to be used in the calculation".

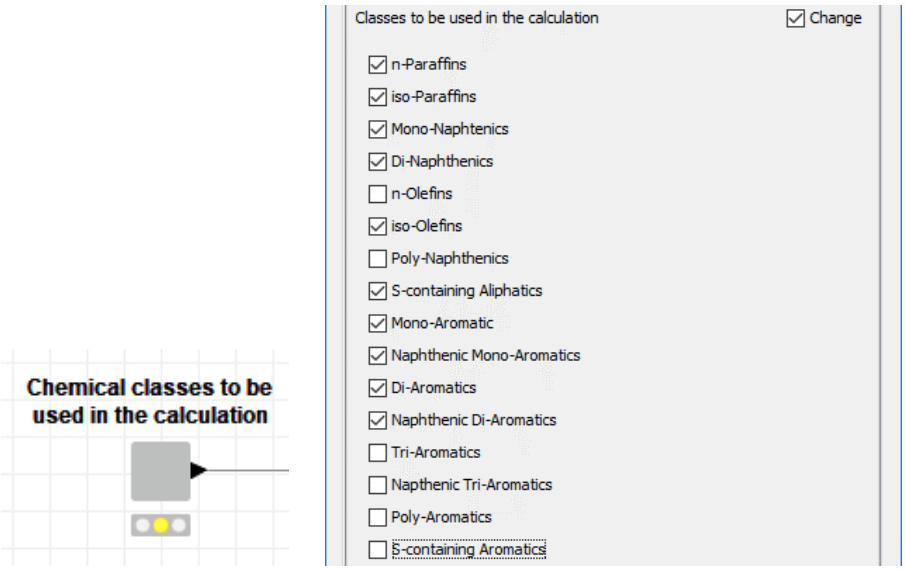

Figure 13. The 'Chemical classes' QuickForm node and detail of the 'Chemical classes' table.

Additionally, if the user would like to run the model on a limited number of structures, or would like to exclude structures, a filter by constituent name can be applied. The user needs to access the node: "PetroTox 4.0" → "Assign masses to library chemicals" → "Substances selector"

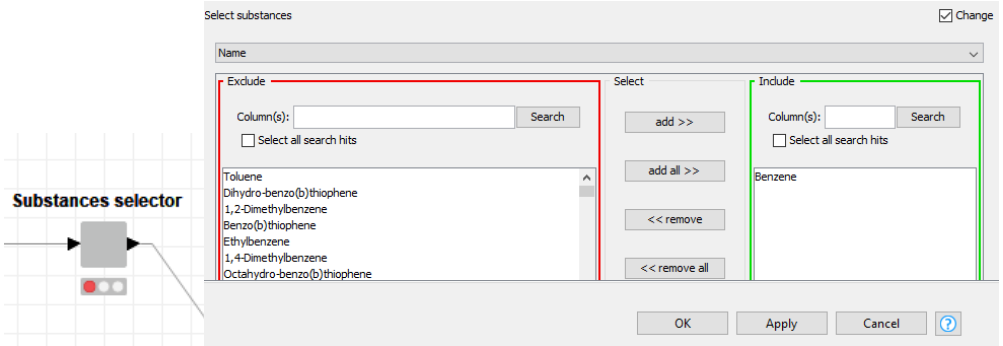

Figure 14. The 'Substances selector' QuickForm node and detail of the 'Substances selector' table.

If you wish to permanently save the exclusion of hydrocarbon classes and/or constituents in the existing or a new workflow for future use, please do not forget to save the workflow (Ctrl + S).

You can also change the way to calculate the PNEC as there are two different **HC5 formula** (section 4.4.2). You can select the appropriate formula in the node "PetroTox 4.0" → "Assign masses to library chemicals" → "PNEC calculation variables" and update the values of the different variables (e.g. Universal Narcosis Slope, geomean of CCLTB, kz). The default in PetroTox is formula 17 (Fraction dissolved=1/(1+Lipid Concentration+K_MW).

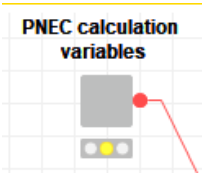

Figure 15. The 'PNEC calculation variables' metanode.

Finally, you can change the **cut-off** limit to calculate membrane water partitioning coefficient (KMW). As explained for equation 11, the calculation of that value change from a specific limit (by default 6). You can adapt the values for that calculation in the node "PetroTox 4.0" → "Assign masses to library chemicals" → "log(KMW) cut off and slope".

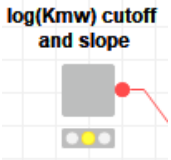

Figure 16. The 'log(KMW) cut off and slope' metanode.

5.4. SAVING RESULTS

The user can save the status of the workflow after finishing the calculations by exporting the workflow via

- **File**
- **Export KNIME Workflow…".**

The workflow being used should be selected, uncheck the option to "reset the workflow" and save it with an appropriate name that includes the date and the name of the analysed substance

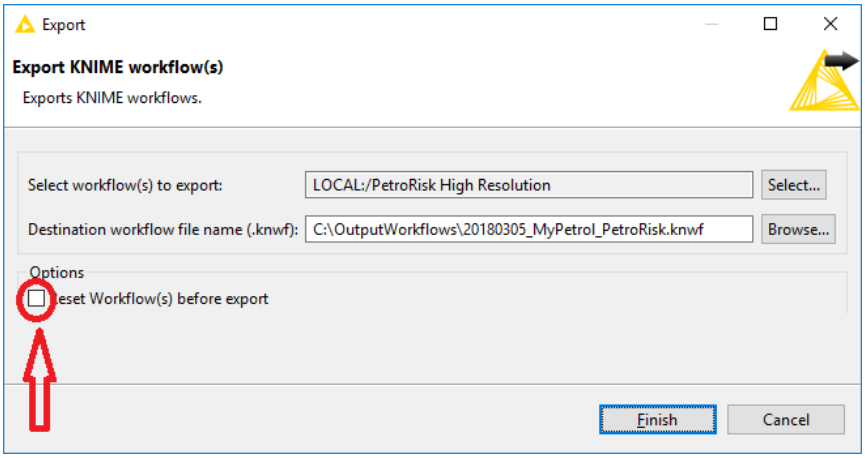

Figure 17. The 'Export KNIME workflow(s)' dialog box, with the check box to reset the workflow(s) before export highlighted.

6. RUN TIME

Depending on the data provided and number of loading setups, the run time of the model can vary significantly. Here are some options to increase the model execution speed.

1. **Increase the memory used by KNIME.** You can increase the amount of memory that KNIME can use. In the KNIME installation directory there is a file called knime.ini. Open the file, find the entry -Xmx1024m, and change it for example to –Xmx12288m if your computer has 16Gb Random Access Memory (RAM), save the .ini file, and (re)start KNIME. If KNIME fails to restart even though your memory capacity is sufficient, re-install it using the most recent installer (.exe file) which allows you to allocate the memory during the installation process.
2. **Use a computer with more processing cores and/or memory.** The KNIME framework scales well with the computer power, so a faster machine will produce the results faster. A desktop will often have more processing power than a laptop, so would be recommended for a large number of runs.
3. **Advanced option: reduce the constituent library size.** The processing considers the whole chemical library provided to PetroTox KNIME. If a subset of the library is used the calculations will be performed faster. You can select the substances used for the calculation as explained in section.

7. GLOSSARY

ACR	Acute to Chronic Ratio of toxicities
BP	Boiling Point
CNO	Carbon Number
CTLBB	Critical Target Lipid Body Burden
EPA	Environment Protection Agency
EqP	Equilibrium Partitioning Approach
GCxGC	2-dimensional Gas Chromatography
HB	Hydrocarbon Block
HC5	Hazardous Concentration that affects 5% of species
KMW	Membrane Water partitioning coefficient
KNIME	Konstanz Information Miner
KTLW	Target Lipid Water partitioning coefficient
LC50	Lethal Concentration (at 50% mortality)
LL50	Lethal Loading rate (at 50% mortality)
NOEC	No Observed Effect Concentration
PAH	Polycyclic Aromatic Hydrocarbon
PNEC	Predicted No-Effect Concentration
POC	Particulate Organic Carbon
RAM	Random Access Memory
SAS	Statistical Analysis System
SPARC	SPARC Performs Automated Reasoning in Chemistry (http://ibmlc2.chem.uga.edu/sparc/)
TLM	Target Lipid Model
TPH	Total Petroleum Hydrocarbons
TU	Toxic Unit
UVCB	Unknown or Variable Composition, Complex reaction products of Biological materials

8. REFERENCES

1. Di Toro, D. M.; McGrath, J. A.; Hansen, D. J. (2000) Technical Basis for Narcotic Chemicals and Polycyclic Aromatic Hydrocarbon Criteria. I. Water and Tissue. *Environ. Toxicol. Chem. 19* (8), 1951–1970. https://doi.org/10.1002/etc.5620190803.

2. McGrath, J. A.; Parkerton, T. F.; Toro, D. M. D. (2004) Application of the Narcosis Target Lipid Model to Algal Toxicity and Deriving Predicted-no-effect Concentrations. *Environ. Toxicol. Chem. 23* (10), 2503–2517. https://doi.org/10.1897/03-538.

3. Redman Aaron D.; Parkerton Thomas F.; Paumen Miriam Leon; McGrath Joy A.; den Haan Klaas; Di Toro Dominic M. (2014) Extension and Validation of the Target Lipid Model for Deriving Predicted No-effect Concentrations for Soils and Sediments. *Environ. Toxicol. Chem. 33* (12), 2679–2687. https://doi.org/10.1002/etc.2737.

4. McGrath, J. A.; Parkerton, T. F.; Hellweger, F. L.; Toro, D. M. D. (2005) Validation of the Narcosis Target Lipid Model for Petroleum Products: Gasoline as a Case Study. *Environ. Toxicol. Chem. 24* (9), 2382–2394. https://doi.org/10.1897/04-387R.1.

5. Di Toro, D. M.; McGrath, J. A.; Stubblefield, W. A. (2007) Predicting the Toxicity of Neat and Weathered Crude Oil: Toxic Potential and the Toxicity of Saturated Mixtures. *Environ. Toxicol. Chem. 26* (1), 24–36. https://doi.org/10.1897/06174R.1.

6. Concawe (1996) *Acute Aquatic Toxicity of Gasolines - Report on CONCAWE Test Programme*. Report No. 96/57. Brussels: CONCAWE

7. Verbruggen, E. (2004) Environmental Risk Limits for Mineral Oil (Total Petroleum Hydrocarbons).

8. McCarty, L. S.; Mackay, D.; Smith, A. D.; Ozburn, G. W.; Dixon, D. G. (1991) Interpreting Aquatic Toxicity QSARs: The Significance of Toxicant Body Residues at the Pharmacologic Endpoint. *Sci. Total Environ. 109–110*, 515–525. https://doi.org/10.1016/0048-9697(91)90205-S.

9. Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. (2004) Prediction of the Solubility, Activity Coefficient and Liquid/Liquid Partition Coefficient of Organic Compounds. *QSAR Comb. Sci. 23 (*9), 709–720. https://doi.org/10.1002/qsar.200430866.

10. Sprague, J. B.; Ramsay, B. A. (1965) Lethal Levels of Mixed Copper–Zinc Solutions for Juvenile Salmon. *J. Fish. Res. Board Can.* 22 (2), 425–432. https://doi.org/10.1139/f65-042.

11. Di Toro, D. M.; Zarba, C. S.; Hansen, D. J.; Berry, W. J.; Swartz, R. C.; Cowan, C. E.; Pavlou, S. P.; Allen, H. E.; Thomas, N. A.; Paquin, P. R. (1991) Technical Basis for Establishing Sediment Quality Criteria for Nonionic Organic Chemicals Using Equilibrium Partitioning. *Environ. Toxicol. Chem.*10 (12), 1541–1583. https://doi.org/10.1002/etc.5620101203.

12. Mayer, P.; Halling-Sørensen, B.; Sijm, D. T. H. M.; Nyholm, N. (1998) Toxic Cell Concentrations of Three Polychlorinated Biphenyl Congeners in the Green Alga Selenastrum Capricornutum. *Environ. Toxicol. Chem.*17 (9), 1848–1851. https://doi.org/10.1002/etc.5620170927.

13 . McGrath, J. A.; Toro, D. M. D. (2009) Validation of the Target Lipid Model for Toxicity Assessment of Residual Petroleum Constituents: Monocyclic and Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.28* (6), 1130–1148. https://doi.org/10.1897/08-271.1.

APPENDIX: ABOUT KNIME

KNIME, the Konstanz Information Miner, is an open source data analytics, reporting and integration platform based on Eclipse and written in Java. An intuitive drag and drop style graphical user interface allows visual assembly, execution and analysis of complex data flows (termed workflows) without the extensive need for coding. To some extent, KNIME can be considered as a statistical analysis system (SAS) alternative.

NODES AND WORKFLOWS, META-NODES AND SUB-WORKFLOWS

A node is the smallest processing unit in KNIME. Each node has been dedicated to perform a specific task with minimal programming. For example, training a decision tree, filtering data rows, calculating a pivoting matrix, applying a model. Nodes in the editor appear with their icon, their name, their background (indicating the node type), their ports and a traffic-light status display. A KNIME user can choose from over 2 000 nodes from the nodes repository, including nodes which integrate external tools (see 2.4) into a workflow. After being created, a node needs to be configured to execute the task. After configuration, a node needs to be executed to actually carry out the assigned task.

The available node types are: source (any kind of data source, green), sink (any kind of persisting nodes, blue), manipulator (data transforming nodes, brown), miner (orange), predictor (red), view (nodes that display data, dark-yellow), meta (nodes that can contain other nodes, violet) and other (grey-blue).

Ports on the left of the node icon are input ports, the ones on the right are output ports. Ports with a black triangle are data ports. Data model ports are represented by small blue squares.

Each node has a state, and a traffic light depicted under each node represents the node's state. The red light means the node has been created, but it is not ready to do anything, it needs to be configured (it is either not fully connected, some settings may be missing or incorrect, or the state of the predecessor node is also red). The amber light means that the node has been configured and it is ready to run its task. The green light indicates that the node has executed its task successfully. The crossed red light indicates that the node has executed its task unsuccessfully.

A sequence of nodes makes a workflow. A workflow is the graphic equivalent to a script as a sequence of instructions. Nodes can be connected to each other through their input and output ports to form a workflow.

Meta nodes are nodes that contain sub-workflows, i.e. in the workflow they look like a single node, although they can contain many nodes and even more meta nodes. They are created with the help of the meta node wizard, which contains customisable pre-defined meta-nodes. Meta nodes look different to normal nodes: the icon is not rounded and has a dark grey background. There is no status light and no progress. The state of a meta node is displayed by an icon in the meta node (seen from outside). The states of a meta node are the same as the states of a workflow. A meta node can be: (1) idle/configured if there is at least one node inside the meta node that is neither executed nor executing (an orange clock icon); (2) executing if at least one node is executing (a green 'fast forward' icon); (3) executed if all contained nodes are executed (a green 'check' icon) (Figure 9).

INTERFACE

The KNIME Workbench is made up of the following components:

- **KNIME Explorer**: Overview of the available workflows and workflow groups in the active KNIME workspaces, i.e. your local workspace. Once imported, you will have PetroTox there.
- **Workflow Coach**: Lists node recommendations based on the workflows built by the wide community of KNIME users. It is inactive if you don't allow KNIME to collect your usage statistics and it won't be useful to run PetroTox.
- **Node Repository**: All nodes available in core KNIME Analytics Platform and in the extensions you have installed are listed here. The nodes are organized by categories but you can also use the search box on the top of the node repository to find nodes. If you don't create new workflows or change the current one you don't need to use this component either.
- **Workflow Editor**: Canvas for editing and running the currently active workflow.
- **Description**: Description of the currently active workflow, or a selected node (in the Workflow Editor or Node Repository).
- **Outline**: Overview of the currently active workflow.
- **Console**: Shows execution messages indicating what is going on under the hood.

You can select what components you want to see from the "View" menu. To use PetroTox you just need the KNIME Explorer, the workflow editor and the Console. A more detailed explanation can be found online at: https://docs.knime.com/2019-06/analytics_platform_workbench_guide/index.html

VISIBILITY & TRACEABILITY

All data processed by KNIME is visible at every step of the calculations. This allows for: (1) analysis of intermediary data, (2) validation of the model, (3) explanation of model details to regulators and auditors, and (4) debugging.

EXTENSIBILITY

KNIME has many different integrations such as database connectors, molecular viewers, export formats, and more. This makes the PetroTox workflow a tool that can easily be modified to fit into other more complex processes (e.g. screening of substances). It also allows the workflow to be easily adjusted to add more visualisations, charts, or reports depending on the user's needs.

Concawe

Environmental Science for European Refining

Concawe
Boulevard du Souverain 165
B-1160 Brussels
Belgium

Tel: +32-2-566 91 60
Fax: +32-2-566 91 81
e-mail: info@concawe.eu
http://www.concawe.eu

ISBN 978-2-87567-118-9