Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSARTM

de Haas, E.M.; Eikelboom, T.; Bouwman, T.

published in
SAR and QSAR in Environmental Research
2011

DOI (link to publisher)
10.1080/1062936X.2011.569949

document version
Publisher's PDF, also known as Version of record

SAR and QSAR in Environmental Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsar20

Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR™

E.M. de Haas [a], T. Eikelboom [a] & T. Bouwman [a]

[a] Chemical Risk Analysis, TNO Triskelion BV, Zeist, The Netherlands

Available online: 07 Jul 2011

To cite this article: E.M. de Haas, T. Eikelboom & T. Bouwman (2011): Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR™, SAR and QSAR in Environmental Research, 22:5-6, 545-559

To link to this article: http://dx.doi.org/10.1080/1062936X.2011.569949

SAR and QSAR in Environmental Research
Vol. 22, Nos. 5–6, July–September 2011, 545–559

Taylor & Francis
Taylor & Francis Group

Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR™

E.M. de Haas*, T. Eikelboom and T. Bouwman

Chemical Risk Analysis, TNO Triskelion BV, Zeist, The Netherlands

(*Received 25 January 2011; in final form 7 February 2011*)

This study concentrates on the external validation of an existing Quantitative Structure–Activity Relationship (QSAR) model widely used for long-term aquatic toxicity to fish. In the context of the REACH legislation, QSARs are used as an alternative for experimental data to achieve a complete environmental assessment without the need for animal testing. The predictivity of the model was evaluated in order to increase the reliability of the model. We assessed whether the model met all of the OECD principles. The model was adapted to become more robust, and predictions were made with an external validation set collected from several databases. For the internal validation of the QSAR, the r^2, Q^2_{LOO} and Q^2_{LMO} were used as validation criteria, and for the external validation r^2, Q^2_{ext}, h and the validation ratio were used. A few substances were classified as outliers and therefore the applicability domain of the QSAR had to be adjusted. The QSAR passed all validation criteria and met all the OECD principles for QSAR validation, and the long-term toxicity QSAR for fish can be applied with high certainty of a correct prediction within the limits of the inherent uncertainty of the model in cases where the substance falls within the applicability domain.

Keywords: ECOSAR; external validation; long-term aquatic toxicity to fish; QSAR for neutral organics; REACH

1. Introduction

(Quantitative) Structure–Activity Relationships ((Q)SARs) are theoretical models used to predict activities from the chemical structure of a substance. These models are used, for instance, in aquatic risk assessment to predict the potential effects in ecotoxicology. In addition to filling the experimental data gap, (Q)SAR predictions may also be used as a supporting tool for the evaluation of the accuracy of available experimental data, and they can assist in the decision as to whether further testing is needed. Also, the application of (Q)SARs will reduce animal testing and costs, and speed up the number of risk assessments for hazardous chemicals.

According to recent policy developments in the European Union [1], it is expected that the use of (Q)SARs for REACH (Registration, Evaluation and Authorization of Chemicals) legislation will increase if regulators and industries rely sufficiently on these (Q)SAR predictions. The regulatory use of (Q)SARs requires validation to ensure that they have acceptable predictive power. However, most of the currently available (Q)SARs

*Corresponding author. Email: elske.dehaas@tno.triskelion.nl

ISSN 1062–936X print/ISSN 1029–046X online

DOI: 10.1080/1062936X.2011.569949
http://www.informaworld.com

lack details about their training set and the statistical validation, and can therefore not be used adequately [2].

The validation of a (Q)SAR is the process by which the performance and mechanistic interpretation of the model are assessed for a particular purpose. To assess the importance of validation it is important to investigate how critical the impact of the (Q)SAR prediction is on the risk assessment outcome.

According to the OECD principles for validation of (Q)SAR models, (Q)SARs should be associated with a defined endpoint, an unambiguous algorithm to ensure transparency, a defined domain of applicability, appropriate measures of internal performance (as represented by goodness-of-fit and robustness) and predictivity (as determined by external validation) and, when possible, the mechanistic interpretation should be presented [3,4].

This implies that full details of the training set, processed data and the basic statistical performance need to be provided. Validation statistics measure the error or accuracy of the prediction for the validation range and are expressed as goodness-of-fit, robustness and predictivity. The goodness-of-fit expresses whether the outcome of the model is statistically significant. The r^2 is used for information about the goodness-of-fit. A model's robustness indicates how sensitive the model parameters are to changes in the training set. The model robustness can be checked for instance by internal cross-validation (Q^2) [5]. The accuracy (or predictivity) of the model is acceptable if the predicted value is within a factor of three (0.5 log units) of the measured value.

External validation analyses the predictivity for substances outside the training set. The process of external validation includes the selection of an adequate number of test structures; the identification of the test structures; the selection procedures of test structures; a statistical analysis of the predictive performance of the model; and a comparison of the predictive performance of the model with previously defined acceptability criteria [6].

To validate the model, data independent of the training set are needed, i.e. compounds in the external validation set are not used for model development and do not affect the model. Further, the chemical domains of the training and validation sets must be similar and should cover the whole chemical domain [7]. The domain of a (Q)SAR is the range of properties of the substances for which the model is valid. The training set consists of the data used to construct the model, and the validation set is the set of data used to validate the model and is independent of the training set.

For validation of QSAR models, four strategies are usually adopted: internal validation or cross-validation; data randomization or Y-scrambling; validation by dividing the data set into training and test compounds; and true external validation by application of the model on external data.

The main aim of this paper is the true external validation of a QSAR model for long-term exposure of fathead minnow (*Pimephales promelas*) to the chemical neutral organics class as developed for the ECOSAR^{TM} program [8].

2. Materials and methods

2.1 *Selection of the validation set*

To validate the QSAR, independent data were needed that meet the properties of the substances used in the relationship. The main data source that was explored to collect

effect concentrations is ECOTOX 4.0 (http://cfpub.epa.gov/ecotox/). ECOTOX 4.0 is a database with information per species (*Pimephales promelas*) and effect category (e.g. growth and reproduction), and contains peer-reviewed literature as main source of data. The OECD Screening Information Data Set (SIDS) for High Production Volume (HPV) chemical (http://www.inchem.org/pages/sids.html) contains separate studies sorted by substance, and the European chemical Substances Information System (ESIS) (http://ecb.jrc.ec.europa.eu/esis/) provides only limited information. Other substance information such as molecular weight, substance name and structure and an estimate of the log Kow was obtained from ECOSAR™ v1.00 [8]. ECOSAR™ v1.00 has been incorporated in EPI Suite v4.0 [9], which integrates a number of estimation models to predict environmental and physical/chemical properties. As the long-term QSAR for fish was developed using log Kow values calculated with EPI Suite v4.0, no experimental values were collected for the external validation set.

2.2 *Neutral organics*

Neutral organic chemicals belong to a class of chemicals that are non-ionizable and non-reactive and act via simple non-polar narcosis generally thought of as a reversible, drug-induced loss of conscience (general anaesthesia). This general narcosis is often referred to as baseline toxicity [10]. The types of chemicals that are known to present general narcosis include, but are not limited to, alcohols, acetals, ketones, ethers, alkyl halides, aryl halides, aromatic hydrocarbons, aliphatic hydrocarbons, many cyanates, sulfides and di-sulfides [11].

Neutral organic compounds tend to be hydrophobic, i.e. they are less soluble in water than in organic solvents. Exceptions include organic compounds that contain ionizable groups as well as low molecular weight alcohols (e.g. methanol and ethanol, but not propanol), amines and carboxylic acids where hydrogen bonding occurs, and therefore exhibit excess toxicity [11].

2.3 *ECOSAR™*

ECOSAR™ (Ecological Structure Activity Relationships) is a program to perform a quantitative effect assessment for aquatic organisms [8]. The QSARs presented in ECOSAR™ are used to predict the aquatic toxicity of substances based on data available on substances in the same chemical class.

(Q)SARs have been used by the U.S. Environmental Protection Agency since 1981 to predict the aquatic toxicity of new industrial chemicals in the absence of test data. The program predicts the short-term and, sometimes, long-term toxicity of chemicals to aquatic organisms such as fish, invertebrates and algae, but in some cases also for terrestrial organisms, such as earthworms. If a chemical cannot be identified in a specific class, it is categorized as a neutral organic [8].

2.3.1 *Training set*

The training set for fish consists of 46 experimental data points from 31 substances (Table 1). Most of the substances are benzene, chloroethane and propane and their analogues.

$$\log 30\text{-}d\ ChV(mmol/l) = -0.8508 \log Kow + 0.6063$$
$$r^2 = 0.7393,\ n = 46$$

Table 1. Training set ($n = 28$) of chemicals for the long-term QSAR for fish, training chemical identification number, CAS number, chemical name, log Kow and experimental and predicted toxicity.

					log ChV (mmol/l)		
ID	*CAS-no.*	*Chemical name*	*MW*[a]	*log Kow*[a]	*Experimental–ECOSAR*	*Experimental–adapted*	*Predicted*
O1	1122-54-9	Ethanone, 1-(4-pyridinyl)-	74.1	0.49	−1.07	omitted	
T1	78-83-1	1-Propanol, 2-methyl-	72.1	0.77	0.75	0.75	0.22
O2	110-86-1	Pyridine	100.2	0.8	−0.45	omitted	
T2	109-99-9	Furan, tetrahydro-	68.1	0.94	0.85	0.85	0.06
T3	108-10-1	2-Pentanone, 4-methyl-	88.2	1.2	−0.11	−0.11	−0.18
T4	110-00-9	Furan	99.0	1.4	−0.83	−0.83	−0.37
T5	1634-04-4	Methyl tertiary-butyl ether	133.4	1.4	0.44	0.44	−0.37
T6	107-06-2	1,2-Dichloroethane	138.2	1.8	−0.39	−0.39	−0.74
			167.9	1.8	−0.38		
T7	79-00-5	1,1,2-Trichloroethane	113.0	2.01	−0.68	−0.92	−0.94
			113.0	2.01	−1.15		
T8	150-78-7	*p*-Dimethoxybenzene	92.1	2.2	−0.81	−0.81	−1.12
T9	79-34-5	Ethane, 1,1,2,2-tetrachloro-	112.6	2.2	−1.83	−1.84	−1.12
			130.2	2.2	−1.84		
T10	78-87-5	Propane, 1,2-dichloro-	202.3	2.3	−1.14	−1.14	−1.21
			182.2	2.3	−1.14		
T11	142-28-9	Propane, 1,3-dichloro-	147.0	2.3	−1.29	−1.15	−1.21
			147.0	2.3	−1.01		
T12	108-88-3	Benzene, methyl-	153.1	2.5	−1.02	−1.02	−1.40

O3	88-73-3	Benzene, 1-chloro-2-nitro-	180.6	2.5	−2.62	omitted	
T13	108-90-7	Benzene, chloro-	161.0	2.6	−4.23	−4.23	−1.49
T14	111-87-5	1-Octanol	181.5	2.8	−1.26	−1.26	−1.60
T15	76-01-7	Ethane, pentachloro-	236.7	3.1	−2.09	−2.16	−2.20
			154.2	3.1	−2.01		
			178.2	3.1	−2.26		
			215.9	3.1	−2.27		
T16	119-61-9	Methanone, diphenyl-	226.0	3.2	−2.13	−2.13	−2.06
T17	541-73-1	Benzene, 1,3-dichloro-	232.0	3.3	−1.99	−1.99	−2.15
			74.1	3.3	−1.99		
T18	106-46-7	Benzene, 1,4-dichloro-	72.1	3.3	−2.28	−2.29	−2.15
			100.2	3.3	−2.29		
T19	822-86-6	1,2-Dichlorohexane-trans	68.1	3.5	−2.30	−2.30	−2.34
T20	98-56-6	Benzene, 1-chloro-4-(trifluoromethyl)-	88.2	3.6	−2.32	−2.32	−2.43
T21	95-75-0	Benzene, 1,2-dichloro-4-methyl-	99.0	3.8	−3.17	−3.17	−2.62
T22	120-82-1	Benzene, 1,2,4-trichloro-	133.4	3.9	−2.42	−2.43	−2.71
			138.2	3.9	−2.43		
T23	67-72-1	Ethane, hexachloro-	167.9	4	−3.34	−3.32	−2.81
			113.0	4	−3.30		
T24	83-32-9	Acenaphthylene, 1,2-dihydro-	113.0	4.2	−2.57	−2.78	−2.99
			92.1	4.2	−2.89		
			112.6	4.2	−2.89		
T25	85-01-8	Phenanthrene	130.2	4.4	−4.45	−4.45	−3.18
T26	634-66-2	Benzene, 1,2,3,4-tetrachloro-	202.3	4.6	−3.12	−2.98	−3.37
			182.2	4.6	−2.83		
T27	CBI	CBI	147.0	6.1	−4.01	−4.01	−4.78
T28	CBI	CBI	147.0	6.2	−4.25	−4.25	−4.87

[a]Calculated with ECOSAR™ [8].

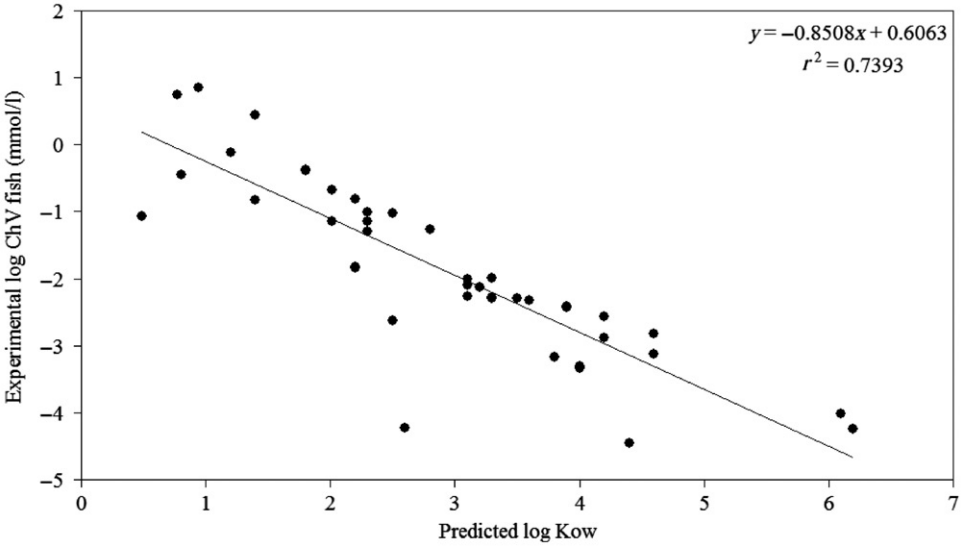

Figure 1. QSAR for Neutral Organics, Fish ChV.

The long-term QSAR for fish is shown in (Figure 1). The adapted long-term QSAR for *Pimephales promelas* (Figure 2) obtained by omitting substances with excess toxicity and averaging the multiple data points used for further validation has the following equation:

$$\log 30\text{-}d\,ChV(mmol/l) = -0.9382 \log\, Kow + 0.9459$$
$$r^2 = 0.7679,\ n = 28$$

2.3.2 *Validation set*

Data for the test set are selected on the basis of 'best available data'. First a collection was made of all substances that were tested with the species for which the QSAR applies (*Pimephales promelas*), and for different effect categories that represent long-term effects such as growth and reproduction. A further selection was made by only using tests with a minimum test duration of 21 days or more.

Only flow-through or static renewal tests with analytical monitoring and only no-observed effect concentration (NOEC), no-observed effect level (NOEL) and maximum allowable toxicant concentration (MATC) values were taken into account, with NOEC and NOEL values being preferred over MATC values if more than one endpoint per CAS-number was available.

Next, the data were filtered for the specific group of substances, the neutral organics, using ECOSAR™ which categorizes the substances based on their structure. Furthermore, as ECOSAR™ does not always categorize chemicals correctly from their set of substances [12], all pesticides (e.g. carbamates and organophosphates) and other substances with one or more functional groups that have a specific mode

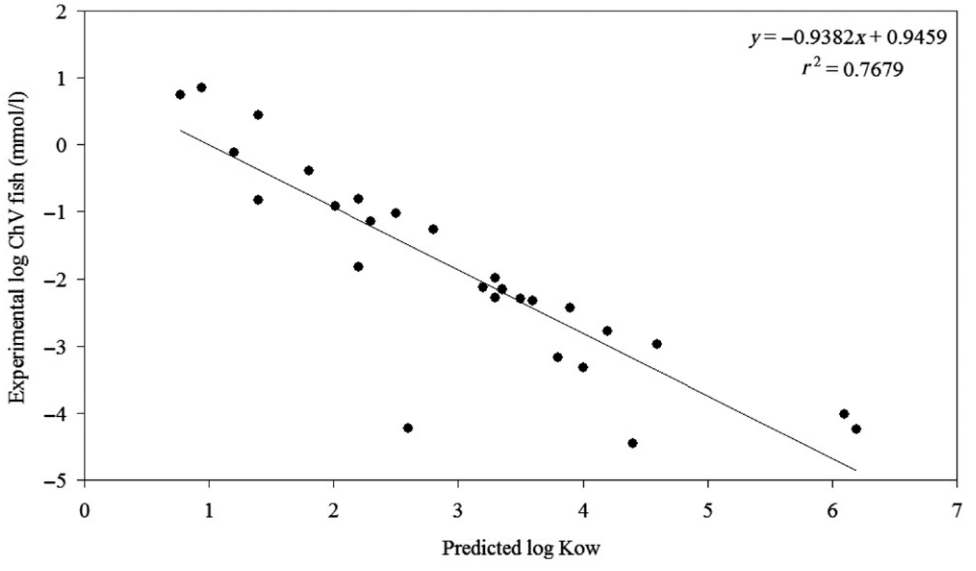

Figure 2. Adapted QSAR for Neutral Organics, Fish ChV.

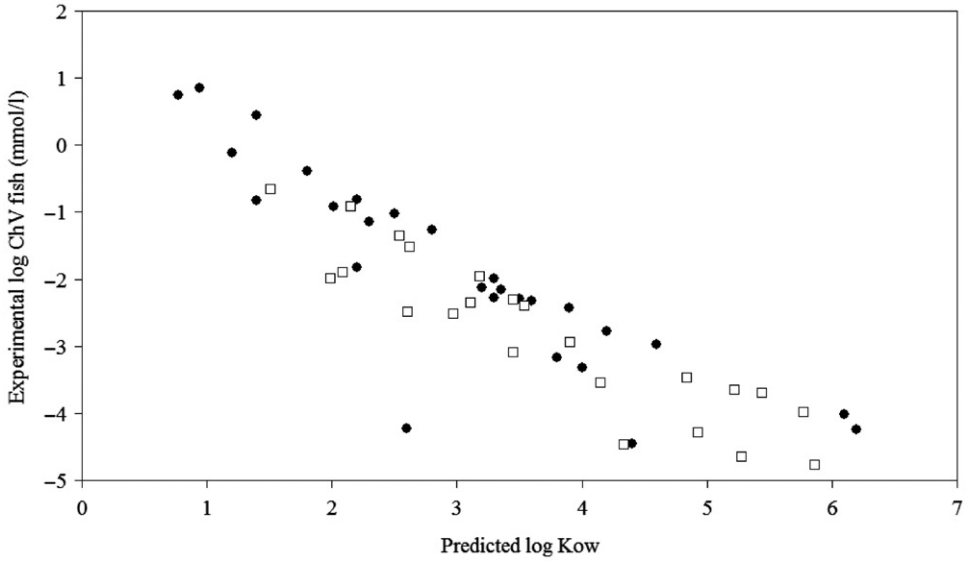

Figure 3. Training ($n = 28$, closed circles) and validation ($n = 23$, open squares) set for long-term fish QSAR.

of action (e.g. nitroso compounds or amines) were discarded from the validation set as they have access toxicity (toxicity additional to narcotic toxicity).

The spread of the training and validation set for the long-term toxicity QSAR for fish is shown in (Figure 3). The final validation set for fish is given in Table 2 [13–28].

Table 2. Validation set ($n = 23$) of chemicals for the long-term QSAR for fish, validation chemical identification number, CAS number, chemical name, number of exposure days, effect concentration, molecular weight (MW), log Kow, experimental and predicted toxicity, residual, hat value and validation ratio.

								log ChV				
ID	*CAS No.*	*Chemical Name*	*Days*	*Concentration (μg/L)*	*Reference*	*MW*[a]	*log Kow*[a]	*Experimental*	*Predicted*	*Residual*	*Hat*	*Validation ratio*
V1	71-43-2	Benzene	32	800	[13]	78.1	1.99	−1.99	−0.92	−1.07	0.056	0.46
V2	76-01-7	Ethane, pentachloro-	32	900	[14]	202.3	3.11	−2.35	−1.97	−0.38	0.036	0.84
V3	78-59-1	2-Cyclohexen-1-one, 3,5,5-trimethyl-	32	4,200	[15]	138.2	2.62	−1.52	−1.51	−0.01	0.039	1.00
V4	80-46-6	Phenol, 4-(1,1-dimethylpropyl)-	30	188	[16]	164.3	3.91	−2.94	−2.72	−0.22	0.053	0.93
V5	83-32-9	Acenaphthene	32	44	[17]	154.2	4.15	−3.54	−2.95	−0.60	0.063	0.83
V6	87-86-5	Phenol, pentachloro-	30	56	[18]	266.3	4.84	−3.47	−3.59	0.12	0.105	1.04
V7	88-06-2	Phenol, 2,4,6-trichloro-	30	970	[19]	197.5	3.45	−2.31	−2.29	−0.02	0.040	0.99
V8	95-49-8	Benzene, 1-chloro-2-methyl-	30	1,400	[20]	126.6	3.18	−1.96	−2.04	0.08	0.036	1.04
V9	95-95-4	Phenol, 2,4,5-trichloro-	28	160	[18]	197.5	3.45	−3.09	−2.29	−0.80	0.040	0.74
V10	105-67-9	Phenol, 2,4-dimethyl-	32	398	[21]	122.2	2.61	−2.49	−1.50	−0.98	0.039	0.60
V11	106-44-5	Phenol, 4-methyl-	32	1,350	[22]	108.1	2.09	−1.90	−1.01	−0.89	0.052	0.53
V12	108-88-3	Benzene, methyl-	32	4,000	[23]	92.1	2.54	−1.36	−1.44	0.08	0.040	1.06
V13	108-95-2	Phenol	32	20,200	[21]	94.1	1.51	−0.67	−0.47	−0.20	0.081	0.70
V14	118-74-1	Benzene, hexachloro-	32	4.76	[14]	284.8	5.86	−4.78	−4.55	−0.22	0.203	0.95
V15	120-12-7	Anthracene	21	6.08	[24]	178.2	4.34	−4.47	−3.13	−1.34	0.072	0.70
V16	127-18-4	Ethene, tetrachloro-	32	500	[14]	165.8	2.97	−2.52	−1.84	−0.68	0.036	0.73
V17	140-66-9	Phenol, 4-(1,1,3,3-tetramethylbutyl)-	124	4.6	[25]	206.3	5.28	−4.65	−4.01	−0.64	0.142	0.86
V18	150-78-7	Benzene, 1,4-dimethoxy-	31	16,600	[26]	138.2	2.15	−0.92	−1.07	0.15	0.050	1.16
V19	206-44-0	Fluoranthene	32	10.4	[27]	202.3	4.93	−4.29	−3.68	−0.61	0.112	0.86
V20	608-93-5	Benzene, pentachloro-	55	4,433	[14]	250.3	5.22	−3.66	−3.95	0.29	0.136	1.08
V21	822-86-6	1,2-Dichlorohexane-trans	31	612	[26]	153.1	3.54	−2.40	−2.38	−0.02	0.042	0.99
V22	3547-04-4	1,1-Bis(*p*-chlorophenyl) ethane	21	50	[25]	251.2	5.44	−3.70	−4.02	0.32	0.157	1.09
V23	84852-15-3	Phenol, 4-nonyl-, branched	33	23	[28]	220.4	5.77	−3.98	−4.47	0.49	0.192	0.46

[a]Calculated with ECOSAR™ [8].

2.4 *Validation*

2.4.1 *Internal validation*

As the model was adapted, an internal validation was performed. The cross-validated squared correlation coefficient (Q^2) [3,4] is calculated according to the formula:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_{i/i}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2},$$

where y_i indicates the experimental value of the ith object, $\hat{y}_{i/i}$ indicates the predicted value of the ith object without using the ith object, and $\bar{y}$ is the average value of the dependent variable of the entire training set; the summations run over all substances in the training set. A model is considered acceptable when the value of Q^2 exceeds 0.5 [29].

2.4.2 *External validation*

One of the most important characteristics of Q(SAR) models is their predictive power. The true predictive power must be estimated by comparing the predicted and observed activities of an (sufficiently large) external test set of compounds that were not used in the model development. A model that is externally predictive should also be robust.

To validate the models, an estimate of the log ChV was made by the model for the validation data set, and these estimates were compared with the measured log ChV values. The correlation between the measured and estimated values reveals the model performance. The external explained variance [3,4] is calculated as:

$$Q^2_{\text{ext}} = 1 - \frac{\sum_{i=1}^{n_{\text{ext}}} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n_{\text{ext}}} \left(y_i - \bar{y}\right)^2},$$

where $\hat{y}_i$ is the predicted value of the ith object; the summations cover all the substances in the validation set.

Another criteria is the validation ratio (predicted toxicity value/measured toxicity value). If the ratio is 1.0, perfect accuracy would be demonstrated. Ratios of less than 1.0 indicate over-prediction and ratios greater than 1.0 toxicity under-prediction. The Spearman's rank correlation ρ coefficient is a measure to which extent the simulations follow the observations.

2.4.3 *Applicability domain*

The applicability domain was verified by the leverage approach. A simple measure of a chemical being too far from the applicability domain of the model is its leverage in the original variable space, h_{ii} [30], which is defined as:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n} (x_i - \bar{x})^2},$$

where x_i is the descriptor value of the ith object, and $\bar{x}$ is the average value of the descriptor in the training set, and n is the number of substances in the training set. There are different 'rules of thumb' for cut-off values. For small data sets, a warning leverage, h^*, is considered large when the $h_{ii} > 3(k+1)/n$. A leverage greater than the warning leverage h^* means that the predicted response is the result of substantial extrapolation of the model

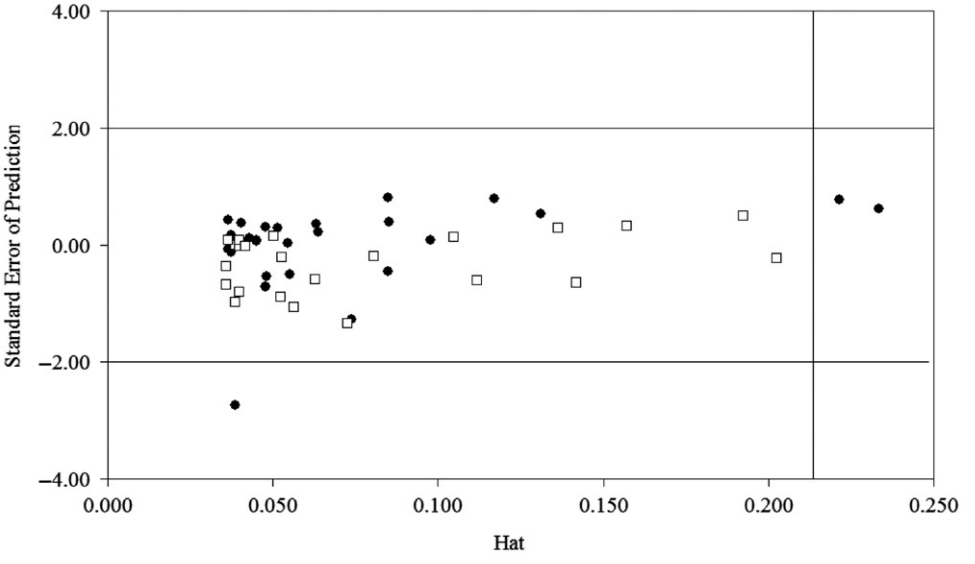

Figure 4. Projection of the training set chemicals (closed circles) and the validation test set chemicals (open squares) in the Williams plot. The vertical line is $h^* = 0.214$, the warning value for the X descriptor space, and the horizontal lines are $\pm 2\sigma$, the cut-off value for the Y dependent space.

Table 3. Internal and external validation of long-term QSARs for neutral organics.

Q^2_{LOO}	Q^2_{LMO}	r^2_{int}	Q^2_{ext}	r^2_{ext}	*Validation ratio*
0.77	0.76	0.77	0.85	0.84	0.88

and, therefore, may not be reliable, so the predicted value must be used with great care. Only predicted data for chemicals belonging to the chemical domain of the training set should be proposed and used. The Williams plot was used to visualize influential chemicals, i.e. chemicals with leverage greater than h^*, as well as outlier chemicals, i.e. chemicals with standardized cross-validated residual greater than $\pm 2\sigma$, the cut-off values for the prediction (Figure 4).

The model's performance was described by using statistical parameters related to fitting power (r^2), model robustness (Q^2, Q^2_{LOO}, Q^2_{LMO}) and model predictivity (Q^2_{ext}).

3. Results and discussion

The robustness of the adapted long-term toxicity QSAR for fish was confirmed by leave-one-out internal cross-validation (Q^2_{LOO}). The Q^2_{LOO} for the long-term toxicity QSAR for fish as taken from ECOSAR™ is 0.75, and the Q^2_{LOO} for the adapted long-term toxicity QSAR for fish is 0.77 (see Table 3). Even though the Q^2_{LOO} was slightly lower, the model is considered acceptable as the value of Q^2 exceeds 0.5 [29]. The $Q^2_{LMO(18\%)}$ for the adapted long-term toxicity QSAR for fish was 0.76.

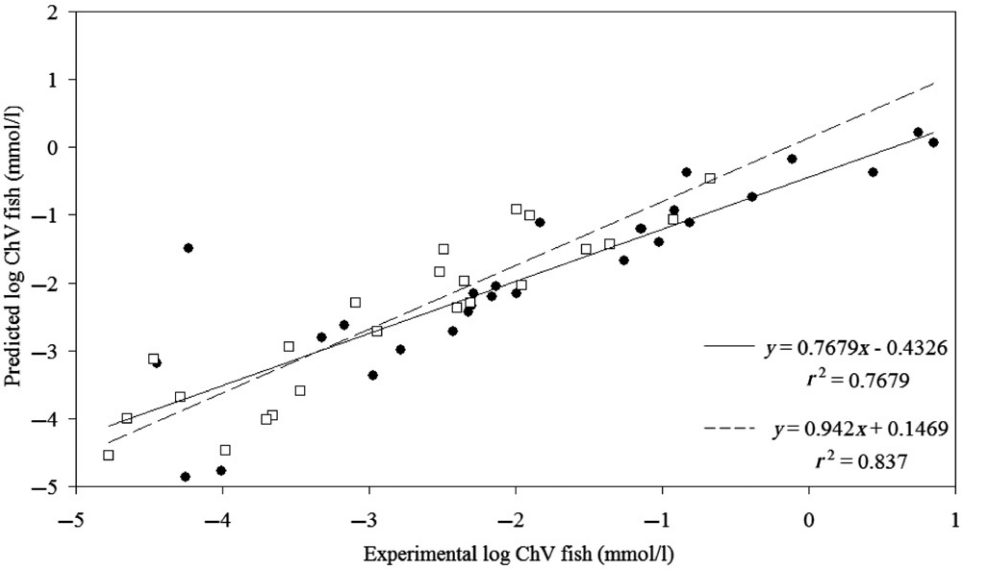

Figure 5. Regression line of the model equation from the predicted log ChV fish (in mmol/l) vs. the measured log ChV fish (in mmol/l) for both the training set ($n = 28$, closed circles, solid line) and the validation set chemicals ($n = 23$, open squares, dashed line).

Cross-validation provides a reasonable approximation of ability with which the QSAR predict the activity values of new compounds. However, supported by other authors, Golbraikh and Tropsha [31] suggest that only external validation gives the ultimate proof of the true predictability of a model. To estimate the predictive power of a QSAR model they recommended the use of following three statistical characteristics of the test set: the correlation coefficient r between the predicted end observed values; coefficients of determination r^2; and the slopes α of the regression lines through the origin. In addition, the external test set must contain at least five compounds, representing the whole range of the log Kow [30].

The adapted long-term toxicity QSAR for fish seems to be able to provide predictions for neutral organic chemicals with a satisfactory predictivity, evaluated by an external explained variance in prediction (Q^2_{ext}) of 0.85. The r^2 is 0.84 (Figure 5), thus the predictions follow the experimental observations. The validation ratio is 0.88, indicating a good accuracy of the model with the model having the tendency to slightly over-predict the long-term toxicity in fish.

A parallel study was performed in which the long-term QSAR for daphnids and algae for the neutral organics chemical class was also validated [32]. For the long-term QSAR for daphnids the internal validity (Q^2_{LOO}) was much higher than the long-term QSAR for fish, 0.93 and 0.75 respectively. The internal validity of the long-term QSAR for algae was lowest (0.70). The external validation, on the other hand, was much better for fish (0.89) than for daphnids (0.79) and algae (0.65). Also, the predictive power of the long-term QSAR for fish is better, as the validation ratio is 0.95, resulting in a slight over-prediction of the results; the validation ratio for daphnids is much lower (0.55), which means a larger over-prediction of the toxicity, whereas the predictive power of the long-term QSAR for algae is good with a validation ratio of 0.93.

(Q)SARs are depended on accurate and valid environmental test data. In order to improve and extend (Q)SAR analysis to all chemical classes, there is a need for standard testing methods and greater knowledge about the physical/chemical properties of chemicals prior to testing. In addition, all test data should be made available in an accessible database.

Only a few studies are available that have been focussed on the validity of the long-term toxicity QSAR for fish in ECOSAR^{TM}. An early study by Nabholz et al. [33] on the first version of the long-term QSAR for fish in ECOSAR^{TM} shows a low mean validation ratio of 0.2 for the class of neutral organics.

The validity of ECOSAR^{TM} was also evaluated by Hulzebos and Posthumus [12] with the Setubal principles. The validation of ECOSAR^{TM} for three 'valid' classes results in a predictivity of $\geq$64% for fish toxicity. Their conclusion is that ECOSAR^{TM} fulfils the first two principles: the endpoint is clear and the algorithm is easily applicable. However, the description of the structural fragments and the modulatory effects are missing. Also, the standard error for the regression lines is missing. Furthermore, Hulzebos and Posthumus [12] showed that for neutral organics, 1 out of 3 (33%) of the long-term estimations for fish were wrong, but they concluded that the QSAR was reliable. In the study of Hulzebos and Posthumus [12] a prediction was performed only for three substances for the neutral organic class. This low number of data might have led to the poor results in their study as opposed to the results in the present study, which was performed with a larger data set.

However, it must be stated that for both studies the validation set was not available, and it could not be tested whether the substances fell within the defined applicability domain of the model, either by defined log Kow or by the type of the substance. As shown by Hulzebos and Posthumus [12], sometimes ECOSAR^{TM} does not categorize chemicals correctly from their set of substances. Therefore it is advised to carefully consider whether the substances falls within the set of compounds on which the QSAR is built, and whether additional functional groups are present with a more specific mode of toxic action.

The applicability domain of the model was analysed using a Williams plot (Figure 4), where the vertical line is the warning leverage value (h^*) which describes the descriptor applicability domain, and the horizontal lines are the $\pm 2\sigma$ the cut-off values for the prediction. The model is characterized by the presence of some influential training chemicals, with leverage values greater than the 'warning leverage' (h^*) fixed at $3(k+1)/n$ (0.214). These chemicals greatly influence the regression line: the regression line is forced near the observed value and their residual values are small, i.e. they are well predicted.

From the Williams plot, it can be observed that for the validation set there is one substance just falling within the applicability domain, and one substance is identified as an outlier of the model domain in the Williams plot according to its high leverage, but the prediction is still reliable as can be seen from the low standard error of prediction. Thus, the model provides predictions that are expected to be reliable.

The upper limit of the applicability domain of the QSAR is set at a log Kow of 5.80. Caution has to be taken when applying (Q)SARs to chemicals where the used property value lies on the edges of the model, and validation should endeavour to keep the domain as broad as possible. Carefully considering the abovementioned arguments and, according to the results from the internal and external testing, the QSAR passed all validation criteria and meets all of the OECD principles for (Q)SAR validation.

The QSAR has a defined endpoint, which is long-term toxicity to fish. The QSAR has an unambiguous algorithm to ensure transparency, which is the adapted equation as described in this study. The QSAR should have a defined domain of applicability,

here neutral organic substances with a log Kow up to 5.80 and a molecular weight higher than 68.1 g/mol. The QSAR should have appropriate measures of internal performance (as represented by goodness-of-fit and robustness) and predictivity (as determined by external validation), which are presented in the current study (see also Table 3 for a summary of the validation results). And, when possible, the mechanistic interpretation should be presented, which here is the reversible, drug-induced loss of consciousness (general anaesthesia), better known as baseline toxicity [10]. Therefore the long-term toxicity QSAR for fish can be applied with high certainty of a correct prediction within the limits of inherent uncertainty of the model, in cases where the substance falls within the applicability domain.

4. Conclusion

The purpose of this study was to validate a long-term toxicity QSAR model for fish to evaluate its predictivity in order to increase the reliability of the predictions of this long-term QSAR model for neutral organics.

The long-term toxicity QSAR model for fish could be used for hazard assessment of chemicals by filling data gaps for the classification of environmental toxicity, thereby reducing animal testing, or by providing a priority list of chemicals for aquatic toxicity testing.

The predicted toxicity values for the substances in the validation set were considered reliable only for those substances belonging to the descriptor domain of the model, as judged by the leverage approach.

In conclusion, the adapted long-term QSAR model for fish meets all of the OECD principles for QSAR validation and could therefore be considered for use in the regulatory assessment of chemicals. The model was developed for a clearly defined endpoint measured in a specific experimental system. It is associated with an unambiguous and therefore transparent algorithm. The applicability domain of the model was re-defined, and the model exhibits a satisfactory internal performance and predictivity. Finally the model has a tentative mechanistic interpretation since the descriptors used in the model are considered relevant to the endpoint. Further, it can be stated that good internal validation results do not automatically imply that external validation results are also good, and vice versa, as can be seen from the comparison of this validation study with the study on validation on the long-term QSAR for fish, daphnids and algae.

Moreover, this study emphasizes the importance of identifying properly the model applicability domain. Caution has to be taken when applying (Q)SARs to chemicals where the used property value lies on the edges of the model, and validation should endeavour to keep the domain as broad as possible. Further validation of other (Q)SARs will improve insight into environmental effect assessment and increase the use and applicability of these models.

References

[1] *Regulation (EC) No 1907/2006 of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC of the European Parliament and of the Council and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94, as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.*

[2] OECD. *The Principles for Establishing the Status of Development and Validation of (Quantitative) Structure-Activity Relationships [(Q)SARs], ENV/JM/TG(2004)27*. Organization for Economic Cooperation and Development, Paris, France, 2004.
[3] OECD. *Guidance document on the validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] models. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69*. Organization for Economic Cooperation and Development, Paris, France, 2007.
[4] P. Gramatica, *Principles of QSAR models validation: Internal and external*, QSAR Comb. Sci. 26 (2007), pp. 694–701.
[5] M. Pavan, T.I. Netzeva, and A.P. Worth, *Validation of a QSAR model for acute toxicity*, SAR QSAR Env. Res. 17 (2006), pp. 147–171.
[6] E. Hulzebos, D. Sijm, T. Traas, R. Posthumus, and L. Maslankiewicz, *Validity and validation of expert (Q)SAR systems*, SAR QSAR Env. Res. 16 (2005), pp. 385–401.
[7] A. Golbraikh, M. Shen, Z. Xiao, Y. Xiao, K Lee, and A. Tropsha, *Rational selection of training and test sets for the development of validated QSAR models*, J. Comput.-Aided Mol. Des. 17 (2003), pp. 241–253.
[8] *ECOSAR^{TM} v1.00*, U.S. Environmental Protection Agency, 2009 a; software available at http://www.epa.gov/opptintr/exposure/pubs/episuite.htm.
[9] *EPI Suite^{TM} v4.00*, U.S. Environmental Protection Agency, 2009 b, software available at http://www.epa.gov/opptintr/exposure/pubs/episuite.htm.
[10] G.D. Veith and S. Broderius, *Rules for distinguishing toxicants that cause types I and type II narcosis syndromes*, Environ. Health Perspect. 87 (1990), pp. 207–211.
[11] K. Mayo-Bean, J.V. Nabholz, R. Clements, M. Zeeman, T. Henry, D. Rodier, K. Moran, B. Meylan, and P. Ranslow *Estimating toxicity of industrial chemicals to aquatic organisms using the ECOSAR (ECOlogical Structure Activity Relationship) Class Program*. MS-Windows Version 1.00. U.S. Environmental Protection Agency (2009). Available at http://www.epa.gov/oppt/newchems/tools/ecosartechfinal.pdf
[12] E.M. Hulzebos and R. Posthumus, *(Q)SARs: Gatekeepers against risk on chemical?*, SAR QSAR Env. Res. 14 (2003), pp. 285–316.
[13] C.L. Russom, and S.J. Broderius, *A chronic aquatic toxicity database for development of predictive toxicology models for industrial organic chemicals*. U.S. EPA, Environmental Research Laboratory-Duluth, MN, USA. Deliverable No. 8477. PPA: L104/G/2013, 1991.
[14] N. Ahmad, D. Benoit, L. Brooke, D. Call, A. Carlson, D. Defoe, J. Huot, A. Moriarity, J. Richter, P. Shubat, and G. Veith, *Toxicity of selected uncoupling and acetylcholinesterase-inhibiting pesticides to the fathead minnow (*Pimephales promelas*)*. In *Pesticides in Terrestrial and Aquatic Environments*, D.L.Weigmann, ed., Proc. Natl. Res. Conf., Virginia Polytechnic Inst. and State Univ., Blacksburg, VA, US, 1989, pp. 317–336.
[15] A.E. Lemke, E. Durhan, and T. Felhaber, *Evaluation of a fathead minnow* Pimephales promelas *embryo-larval test guideline using acenaphthene and isophorone*. U.S. EPA Environmental Research Laboratory-Duluth, MN, USA. U.S. EPA Report No. 600/3-83-062, 1983.
[16] G.H. Panter, T.H. Hutchinson, K.S. Hurd, J. Bamforth, R.D. Stanley, S. Duffell, A. Hargreaves, S. Gimeno, and C.R. Tyle, *Development of chronic tests for endocrine active chemicals: Part 1. An extended fish Early-Life Stage test for oestrogenic active chemicals in the fathead minnow (*Pimephales promelas*)*, Aquat. Toxicol. 77 (2006), pp. 279–290.
[17] The Academy of Natural Sciences, *Early Life Stage studies using the fathead minnow (*Pimephales promelas*) to assess the effects of isophorone and acenaphthene,* Div. Limnol. Ecol., Philadelphia, PA, Manuscript Submission to U.S. EPA, Cincinnati, OH, USA, 1981.
[18] A.D. Arthur and D.G. Dixon, *Effects of rearing density on the growth response of juvenile fathead minnow (*Pimephales promelas*) under toxicant-induced stress*, Can. J. Fish. Aquat. Sci. 51 (1994), pp. 365–371.

[19] G.A. LeBlanc, *Comparative structure-toxicity relationships between acute and chronic effects to aquatic organisms*, in *QSAR in Environmental Toxicology*, K.L.E. Kaiser, ed., D. Reidel Publ. Co., Dordrecht, Holland, 1984, pp. 235–260.
[20] EG & G, Bionomics, *The toxicity of o-chlorotoluene to fathead minnow (*Pimephales promelas*) embryos and larvae*. Toxicity test report submitted to Hooker Chemicals & Plastics Corporation, Niagara Falls, New York. Report #BW-82-7-1235. E.G. & G., Bionomics, Wareham, Massachusetts, 1982.
[21] C.L. Russom, *Acute and Early Life Stage toxicity data*. June 21 Memo from U.S. EPA, Duluth, MN. to R. Spehar U.S. EPA, Duluth, 1993.
[22] M.G. Barron and I.R. Adelman, *Nucleic acid, protein content, and growth of larval fish sublethallity exposed to various toxicants*, Canad. J. Fish. Aquat. Sci. 41 (1984), pp. 141–150.
[23] E.W. Devlin, J.D. Brammer, and R.L. Puyear, *Acute toxicity of toluene to three age groups of fathead minnows (*Pimephales promelas*)*, Bull. Environm. Contam. Toxicol. 29 (1982), pp. 12–17.
[24] A.T. Hall, *Reproductive and behavioral toxicity of anthracene in the fathead minnow (*Pimephales promelas*)*, Ph.D. Thesis, Miami Univ. OH, USA, 1993.
[25] J.V. Brian, C.A. Harris, M. Scholze, A. Kortenkamp, P. Booy, M. Lamoree, G. Pojana, N. Jonkers, A. Marcomini, and J.P. Sumpter, *Evidence of estrogenic mixture effects on the reproductive performance of fish*, Environ. Sci. Toxicol. 41 (2007), pp. 337–344.
[26] D.J. Call, L.T. Brooke, M.L. Knuth, S.H. Poirier, and M.D. Hoglund, *Fish subchronic toxicity prediction model for industrial organic chemicals that produce narcosis*, Environ. Toxicol. Chem. 4 (1985), pp. 335–341.
[27] R.L. Spehar, S. Poucher, L.T. Brooke, D.J. Hansen, D. Champlin, and D.A. Cox, *Comparative toxicity of fluoranthene to freshwater and saltwater species under fluorescent and ultraviolet light*, Arch. Environ. Contam. Toxicol. 37 (1999), pp. 496–502.
[28] T.J. Ward and R.L. Boeri, *Early Life Stage toxicity of nonylphenol to the fathead minnow,* Pimephales promelas, Final Report of the Chemical Manufacturers Association, Washington, DC, USA, 1991.
[29] P.P. Roy, S. Paul, I. Mitra, and K. Roy, *On two novel parameters for validation of predictive QSAR models*, Molecules, 14 (2009), pp. 1660–1701.
[30] A.C. Atkinson, *Plots, transformations and regression*, Clarendon Press, Oxford, 1985.
[31] A. Golbraikh and A. Tropsha, *Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection*, J. Comput.-Aided Mol. Des. 16 (2002), pp. 357–369.
[32] E.M. de Haas and T. Bouwman, *Internal and external validation of the long-term QSARs for neutral organics to fish, daphnids and algae*. TNO Report V9176. TNO Quality of Life, The Netherlands, 2010.
[33] J.V. Nabholz, R.G. Clements, M.G. Zeeman, K.C. Osborn, and R. Wedge, *Validation of structure activity relationships used by the U.S. EPA's office of pollution prevention and toxics for the environmental hazard assessment of industrial chemicals*, Environ. Toxicol. Risk Assess. 2 (1993), pp. 571–590.