

A Decision Support System to Predict Acute Fish Toxicity

Anders L Madsen ALM@HUGIN.COM
HUGIN EXPERT A/S and Department of Computer Science, Aalborg University, Denmark

S. Jannicke Moe JANNICKE.MOE@NIVA.NO
Norwegian Institute for Water Research (NIVA), Oslo, Norway

Thomas Braunbeck BRAUNBECK@UNI-HD.DE
University of Heidelberg, COS, Aquatic Ecology and Toxicology, Heidelberg, Germany

Kristin A. Connors CONNORS.KA@PG.COM
The Procter & Gamble Company, Mason OH, USA

Michelle Embry MEMBRY@HESIGLOBAL.ORG
Health and Environmental Sciences Institute (HESI), Washington DC, USA

Kristin Schirmer KRISTIN.SCHIRMER@EAWAG.CH
Eawag, Swiss Federal Institute of Aquatic Sci & Tech, Dübendorf, Switzerland

Stefan Scholz STEFAN.SCHOLZ@UFZ.DE
Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Raoul Wolf RAOUL.WOLF@NGI.NO
Norwegian Geotechnical Institute (NGI), Oslo, Norway

Adam Lillicrap ADAM.LILICRAP@NIVA.NO
Norwegian Institute for Water Research (NIVA), Oslo, Norway

Abstract

We present a decision support system using a Bayesian network to predict acute fish toxicity from multiple lines of evidence. Fish embryo toxicity testing has been proposed as an alternative to using juvenile or adult fish in acute toxicity testing for hazard assessments of chemicals. The European Chemicals Agency has recommended the development of a so-called weight-of-evidence approach for strengthening the evidence from fish embryo toxicity testing. While weight-of-evidence approaches in the ecotoxicology and ecological risk assessment community in the past have been largely qualitative, we have developed a Bayesian network for using fish embryo toxicity data in a quantitative approach. The system enables users to efficiently predict the potential toxicity of a chemical substance based on multiple types of evidence including physical and chemical properties, quantitative structure-activity relationships, toxicity to algae and daphnids, and fish gill cytotoxicity. The system is demonstrated on three chemical substances of different levels of toxicity. It is considered as a promising step towards a probabilistic weight-of-evidence approach to predict acute fish toxicity from fish embryo toxicity.

Keywords: Bayesian networks; toxicology; weight-of-evidence; real-world application.

1. Introduction

European legislations require Reduction, Replacement or Refinement of animal testing wherever possible (Lillicrap et al., 2016). The use of fish embryos for toxicity testing is considered a promising alternative to the use of juvenile or adult fish. However, fish embryos are not yet

accepted as an alternative for regulatory purposes. European Chemicals Agency (ECHA) has therefore recommended the development of a weight-of-evidence (WOE) approach to evaluate Fish Embryo Toxicity (FET) data in combination with other types of information as a replacement for juvenile fish toxicity data. They have challenged industry and academia to specify the methodology needed for implementation.

In response to this challenge, we have developed a Bayesian network to predict AFT from FET data in combination with other lines of evidence within the project SWiFT (Strengthening weight of evidence for FET data to replace acute fish toxicity). A Bayesian network is a probabilistic graphical model well-suited for integrating uncertain knowledge from multiple sources and reasoning under uncertainty, see e.g., (Kjærulff and Madsen, 2013). This makes Bayesian networks the natural choice of framework for developing a quantitative approach to fish acute toxicity prediction from multiple types of evidence including physical and chemical properties, quantitative structure-activity relationships, toxicity to algae and daphnids, and fish gill cytotoxicity.

In the past, WOE has largely been qualitative in nature in ecotoxicology, and risk and hazard management (Suter et al., 2017a,b). As a result, WOE has been criticised as too vague, non transparent and subjective (Linkov et al., 2016). All the same, to support the development of animal alternatives in toxicity testing, methods are needed for integrating alternative types of evidence and for assessing their uncertainty. In Bayesian methods, WOE has a specific meaning. It is the logarithm of Bayes factor computed as the ratio of the posterior odds to the prior odds (Good, 1960). In the context of environmental assessments, the WOE term is used more generally as a subjective scoring of the evidence impacted by, for instance, relevance, strength and reliability.

Prior to the SWiFT project, Moe et al. (2020) developed a hybrid Bayesian network as a pilot version of a quantitative WOE approach for predicting AFT from FET data. The model performance was evaluated for a selection of substances Lillicrap et al. (2020), and discussed in the more general context of WOE tools for prediction of acute fish toxicity (Belanger et al., 2022). The scientific community of animal alternatives to toxicity testing have received this preliminary model version with high interest, e.g., Paparella et al. (2021), and provided much feedback for improvement of the model.

This paper presents a further developed decision support system using a Bayesian network to integrate and quantify evidence from multiple lines of evidence to predict acute fish toxicity. The model naturally decomposes into a set of components reflecting different types of information. The model is delivered to the user with a flexible web-based user interface. The web-based user interface has been presented to and evaluated by different groups of stakeholders. The main use-case of the system is that a producer of a new chemical substance uses the system to predict fish acute toxicity of the substance and use the results as documentation toward the regulatory authorities such as, e.g., ECHA.

The system was developed using HUGIN software (Andersen et al., 1989; Madsen et al., 2005). and we focus on the technical implementation and functionality.

2. Preliminaries and Notation

A Bayesian network $\mathcal{N} = (\mathcal{X}, G, \mathcal{P})$ over a set of discrete variable \mathcal{X} is an efficient factorization of the joint probability distribution $P(\mathcal{X})$ when the directed, acyclic graph $G = (W, E)$

with vertices W and edges E is sparse. The joint distribution $P(\mathcal{X})$ decomposes into a product of conditional probability distributions (CPDs) $P(X | \text{pa}(X))$ as specified by G such that $P(\mathcal{X}) = \prod_i P(X_i | \text{pa}(X_i))$, where $\mathcal{X} = \{X_1, \dots, X_n\}$, $\text{pa}(X_i)$ are the parents of X_i in $G = (W, E)$. The vertices W of G correspond one-to-one with \mathcal{X} and E is the set of directed edges in G .

The Conditional Linear Gaussian (CLG) Bayesian network, see, e.g., Lauritzen and Jensen (2001), is a hybrid model with discrete and continuous variables where the latter follows a CLG distribution, i.e., $Y \sim N(\alpha(i) + \sum_j \beta_j(i)z_j, \sigma(i)^2)$, where i is a configuration of the discrete parents of Y and z_j is the value of the j th continuous parent Z_j of Y , respectively. The continuous variables are used to capture individual toxicity values, which otherwise would be merged into an average.

A Bayesian network with functions nodes (Madsen et al., 2014) is an extension of Bayesian networks to include function nodes. A function node represents a numerical value computed from a mathematical expression. An expression may include probabilities computed through inference in a Bayesian network and may be used to define constants in mathematical expressions defining the content of a CPD. We use function nodes to parameterize the model and compute values based on the results of inference in the model.

Inference in a Bayesian network is the task for computing posterior marginals $P(X | \epsilon)$ where ϵ is a set of observations (variable instantiations) for all non-observed variables. Inference is performed by message passing in a junction tree structure followed by the evaluation of the expressions associated with function nodes. Function nodes enable the specification of all calculations related to the use of a Bayesian network in a single model, e.g., a function node may take the value one, if a computed value is above a threshold and zero otherwise.

3. Domain of Application

Bayesian networks are gaining popularity in ecotoxicology and ecological risk assessment, because of their ability to integrate different types of data and other information, and to predict the probability of specified states (Kaikkonen et al., 2021).

We have developed a Bayesian network to predict the acute toxicity of a chemical substance to juvenile fish (acute fish toxicity - AFT) based on FET data in combination with different types of evidence, organized into three lines of evidence to represent a WOE approach. Toxicity is measured as either LC50 (lethal concentration for 50% of the test population) or EC50 (effect concentration for 50% of the test population), depending on the assessment endpoint and test type. Note that higher LC50 values represent lower toxicity and vice versa.

The model described in this paper is the only published model developed for this purpose to our knowledge. Therefore, the performance of this model can only be compared with the use of FET data directly as a predictor of AFT.

The long-term aim of the system described in this paper is to provide a conclusive framework that supports the full adoption and acceptance of FET data in combination with other relevant information as an alternative to AFT data, to routinely fulfil the regulatory requirements for AFT data. Relevant knowledge gaps have been identified by ECHA and other stakeholders that currently prevent full acceptance of the FET as a replacement. These in-

clude the domain of applicability for the various inputs (physical-chemical properties, mode of action, all toxicity test lines of evidence, and metabolism considerations) as they relate to FET and AFT outcomes. The intent of the work is to define the widest scope possible, i.e., the group of chemical substances and their properties, for which AFT can safely be replaced by FET data using existing alternative assays and supporting information.

From regulators and society in general there is an ever-increasing demand that the use of animal testing is reduced as much as possible. The aim of the system is to predict AFT LC50 intervals for a chemical of concern from fish embryo toxicity taking multiple types of evidence into consideration thereby reducing the use of animal testing. We have developed a Bayesian network to predict the acute toxicity of a chemical substance to juvenile fish based on multiple source and types of evidence, organized into three lines of evidence.

4. Bayesian Network

The earlier version of this Bayesian network was been developed as a response to the challenge to specify a methodology that is needed for implementing the use of FET data in a WOE approach (Moe et al., 2020), and evaluated for a selection of substances (Lillicrap et al., 2020). This process identified several technical challenges, some of which have been addressed in this paper by further developments of the model. The main structure of the developed Bayesian network is shown in Figure 1. The Bayesian network predicts acute fish toxicity of a chemical substance based on information on the group of substance and observational data from different tests performed on four different assessment endpoints: fish embryo, algae and daphnids, and fish gill toxicity. These endpoints represent three different lines of evidence that are combined in a quantitative WOE approach where the weighting is encoded in the conditional probability distribution of the variable representing acute fish toxicity. In total the Bayesian network has 97 nodes and it should be used to predict acute fish toxicity for one substance at a time.

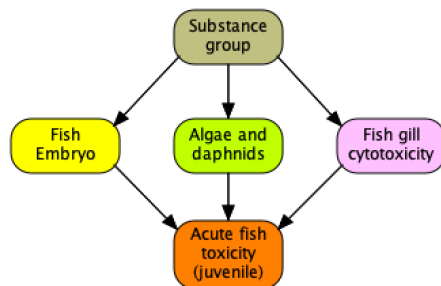


Figure 1: The main components of the Bayesian network.

For the three lines of evidence there is a similar structure supporting the specification of up to ten different test values for each of fish embryo, algae, daphnids, and fish gill cytotoxicity. Figure 2 shows the part of the model responsible for capturing evidence on daphnids and for computing how sensitive the target is to daphnids. We use the (lowest) geometric mean to find the most sensitive endpoint. The evidence is represented as conditional Gaussian variables as the inputs are real values (concentration of the substance

in mg/L). We use continuous variables to represent the toxicity values to better account for certain characteristics of the input data such as the variation among individual toxicity values. A function node and a selector node are responsible for computing the sensitivity of the endpoint. These values are used to identify the most sensitive assessment endpoint, i.e., to identify the species group to which the substance is most toxic. This information is a main criterion for the hazard assessment which is important to communicate back to the user and is computed inside the model. There is a similar structure for the other lines of evidence.

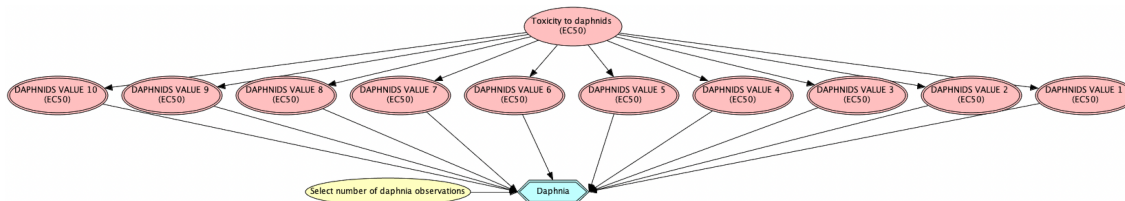


Figure 2: The model component for daphnids with a selector and function node to compute the average toxicity.

Below, we describe a set of model improvements reported in this paper compared to the model reported by Moe et al. (2020): (1) Prior probability distribution of toxicity variables. A set of 27 possible substance groups were defined based on three physical/chemical properties of substances, each discretized into three states: (i) mode of action, (ii) molecular weight and (iii) hydrophobicity (i.e., the inverse of solubility). For each assessment endpoint, a distribution of toxicity values was estimated for each substance group by a hierarchical Bayesian model (ANOVA), using available toxicity data from the EnviroTox database (<https://envirottoxdatabase.org/>). This estimated distribution formed the prior probability distribution of each toxicity node, specific for each substance group.

(2) Determination of substance group for new chemicals. Figure 3 shows the component of the model for determining the substance group when the model is run for a new substance. The substance group is determined by the combination of the three physical/chemical properties of the substance mentioned above, provided by the user. The latter two are entered as real values through function nodes to avoid the additional efforts of conversion to discrete states by the user, in part to support trust in the system. Moreover, the exact molecular weight is needed to transform the input values from concentration unit mg/L to mol/L (explained below).

(3) Conversion of the scale for toxicity values. Although toxicity concentration values are normally reported with the unit mg/L, the toxic potential of a solution is better represented by the count of molecules per volume (i.e., in unit mol/L). Therefore, modelling of toxic properties across substances with different molecular weight (g/mol) is more sensible after conversion to the scale mol/L. Moreover, toxicity values typically follow a log-normal distribution and are therefore log₁₀-transformed before entering the value nodes, which assume a normal distribution. The users typically have the toxicity values in the unit mg/L and the system performs the required transformations, again to avoid the need for conversion by the user. Also, we would like the user to enter their raw values to ensure the highest possible trust in the system. For this reason, both the unit conversion and the log-transformation

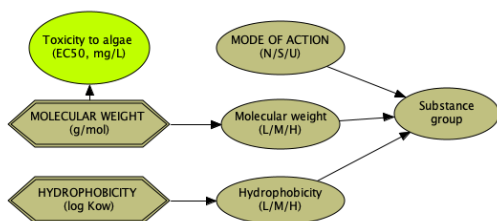


Figure 3: Substance group.

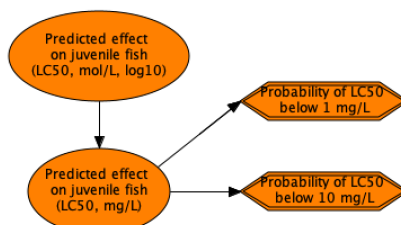


Figure 4: Cumulative probabilities.

are built into the model. The toxicity values were discretised into seven intervals, both in the original and the transformed scale.

(4) Discretization of toxicity values. A resolution of seven states was chosen for the discrete toxicity variables, since this was found to be suitable in a previous version (Moe et al. 2022). Different rules for discretization were used for the two scales of toxicity. In the original scale (mg/L), two concentration values are used as threshold in regulatory processes, 1 and 10 mg/L, and were a natural choice of breakpoints to facilitate interpretation for end-users. Considering that toxicity is often analysed in log-scale, we chose a discretisation with equidistant intervals in log₁₀ scale, extending in both directions from 1 and 10 mg/L. For the transformed scale (mol/L, log₁₀), the discretisation aimed to obtain a more even distribution of observation across the states. An equidistant scale was chosen for the three middle intervals the range -6 to -3 , to capture the majority of the distribution, while wider intervals were used for the tails of the distribution (-9 to -6 and 0 to 3) in this scale.

(5) Back-transformation of the predicted toxicity to the original unit (mg/L). The node molecular weight (g/mol) is a parent of four nodes representing the toxicity of the substance to each line of evidence (Fish embryo, Algae, Daphnids, and Fish gill) measured in mg/L. The purpose is to enable back-transformation of the predicted toxicity (uni mol/L, log₁₀) to the scale of the input values (EC₅₀, mg/L). The distribution of these variables is specified as a mathematical expression involving the molecular weight. This means that these CPTs are generated for each separate substance when a hazard assessment analysis is performed. This model parameterization is a powerful feature of Bayesian networks with function nodes.

Figure 4 shows the computation of cumulative probabilities for the predicted effect on juvenile fish back-transformed from mol/L (log₁₀) to mg/L. The function nodes calculate the probability toxicity level (LC₅₀) being below the commonly used regulatory thresholds, 1 and 10 mg/L, which is relevant information to include in the hazard assessment results.

5. Web-Interface

The system was developed using the HUGIN Web Service (WS) Application Programming Interface (API) (Madsen et al., 2013). It is an architecture for web deployment of probabilistic graphical models including Bayesian networks. Figure 5 illustrates the structure of the system and the flow of data and information through the system. On the left-hand side of the figure the input data in the form of a Microsoft Excel xlsx-file is shown and on the right-hand side of the figure the output data in the form of a PDF report is shown. In between the input and the output, the core structure of the system is shown.

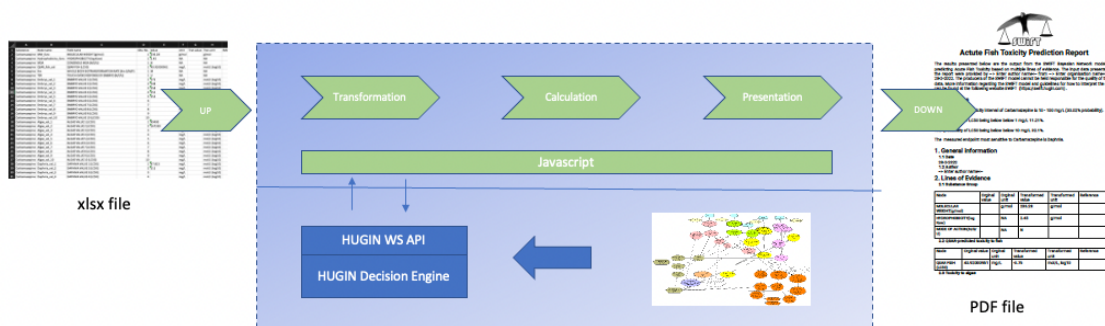


Figure 5: Underlying architecture of the system.

The system is implemented as a client-server architecture enabling multiple users to access the system through a web-browser. On the server side (bottom part of Figure 5), the HUGIN WS API is hosting the Bayesian network model to be accessed from the client side using a REST API in combination with a set of web widgets. On the client side (top part of Figure 5), the browser is running JavaScript code implementing three main steps *Transformation*, *Calculation*, and *Presentation* and interacting with the server.

Field name	Value	Unit
MOLECULAR WEIGHT (g/mol)	236.28	g/mol
HYDROPHOBICITY (log Kow)	2.45	NA
CONSENSUS MOA (N/S/U)	U	NA
QSAR FISH (LC50)	40.92000961	mg/L
WHOLE BODY BIOTRANSFORMATION RATE (Km S/M/F)	M	NA
TOUCH-EVOKE RESPONSE OF EMBRYO (N/Y/U)	U	NA
EMBRYO VALUE 1 (LC50)	157	mg/L
EMBRYO VALUE 2 (LC50)	148	mg/L
EMBRYO VALUE 3 (LC50)	158	mg/L
EMBRYO VALUE 4 (LC50)	150	mg/L
EMBRYO VALUE 5 (LC50)	155	mg/L
EMBRYO VALUE 6 (LC50)	136	mg/L
EMBRYO VALUE 7 (LC50)	156	mg/L
EMBRYO VALUE 8 (LC50)	153	mg/L
EMBRYO VALUE 9 (LC50)	153	mg/L
EMBRYO VALUE 10 (LC50)	150	mg/L
ALGAE VALUE 1 (LC50)	167.33	mg/L
ALGAE VALUE 2 (LC50)	49.4	mg/L
DAPHNIA VALUE 1 (LC50)	97.815	mg/L
DAPHNIA VALUE 2 (LC50)	111	mg/L

The most probable toxicity interval of Carbamazepine is 10 - 100 mg/L (55.12% probability).
 The cumulative probability of LC50 being below 1 mg/L is 0.03%.
 The cumulative probability of LC50 being below 10 mg/L is 18.59%.
 The measured endpoint most sensitive to Carbamazepine is algae.
 The model predicted mean of LC50 is 180.14 mg/L.

Predicted effect on juvenile fish (LC50, mg/L)	Probability	Interval
0.00%	0 - 0.01	
0.00%	0.01 - 0.1	
0.03%	0.1 - 1	
18.55%	1 - 10	
55.12%	10 - 100	
25.37%	100 - 1000	
0.93%	1000 - inf	

Figure 6: (left) data input example and (right) extract of the conclusions part of the website for Carbamazepine.

Data is uploaded as a Microsoft Excel xlsx-file containing the information on the substance including information on the source of each finding. Part of the content of the input file for the chemical substance Carbamazepine is shown in Figure 6 (left). The first lines specify information to determine the substance group and information used in the domain of applicability assessment (not described further here). After this follows the information related to each line of evidence. In the example, there are ten observations of fish embryo, two algae values, and two daphnids values. There are no fish gills values in the data for Carbamazepine. An important feature of the system and underlying model is the ability to manage different number of observations for each line of evidence and missing values.

Figure 6 (right) shows the *Conclusions* part of the website. This part presents the results of the analysis. There are four main results computed using the posterior distribution of

Predicted effect on juvenile fish (LC50, mg/L), see Figure 4, and the nodes reflecting the toxicity to juvenile fish of each endpoint. The rightmost part of Figure 5 shows the first of the automatically generated PDF report that can be produced by the user once the hazard assessment of the chemical substance is completed. The PDF report includes general information on the substance, the original values and the transformed values used by the support, as well as any references and other information included in the input data file).¹

The system has been developed through a virtuous cycle of interaction between model developers, system developers, users, knowledge providers, domain experts and legislators. The system has been presented for different stakeholder groups such as, for instance, the Animals Alternatives Interest Group of SETAC (Society of Environmental Toxicology and Chemistry), People for the Ethical Treatment of Animals (PETA), ECHA, The European Chemical Industry Council (CEFIC), Environment Canada, the OECD validation management group (OECD vmgECO), the Norwegian Environment Agency and the Norwegian Food Safety Authority. Valuable feedback and constructive critique have been collected through both in-person and online presentations and demonstrations of the systems.

The web front-end of the system and the three chemical substances considered in the next section are available from <https://swift.hugin.com/models/FET/>.

6. Results for Three Substances

Model performance was assessed by running the Bayesian network with input data from several substances and comparing the outcome, i.e., the predicted acute toxicity of selected chemical substances to juvenile fish, to the experimentally measured toxicity of the same substances to this endpoint. The evaluation compared the most probable toxicity interval of the predicted versus the observed AFT nodes under various combinations of criteria for data selection and precision.

A more in-depth analysis and fine-tuning of the performance of the Bayesian network is ongoing and will be reported elsewhere. The current results on more than 150 different chemicals show that the Bayesian network can accurately identify the most probable toxicity interval of each substance in 75-80% of the cases. Moreover, in most of the wrongly identified cases the model predicts a higher toxicity of the substance than the observed, which means that the hazard assessment is over-protective rather than under-protective. While we aim for the model to be unbiased, an-overprotective outcome is preferable to under-protection from a risk management perspective.

In this section, we analyse three substances of different levels of toxicity. The substances are in order of decreasing toxicity to fish embryo Endrin, Triclosan, and Carbamazepine.

6.1 Endrin

*Endrin is an organochloride primarily used as an insecticide, as well as a rodenticide and piscicide.*² Figure 7 shows the input data for the chemical substance Endrin and the conclusion including the posterior distribution. The measured acute fish toxicity of Endrin

1. A complete sample report is available here https://swift.hugin.com/huginprog/models/FET/data/swift_PGM22_sample_report.pdf

2. <https://en.wikipedia.org/wiki/Endrin> accessed 31 May 2022.

is 0.0012 mg/L (average of 57 values), which corresponds to the toxicity level labelled *extremely high toxicity* (LC50 interval 0-0.01 mg/L).

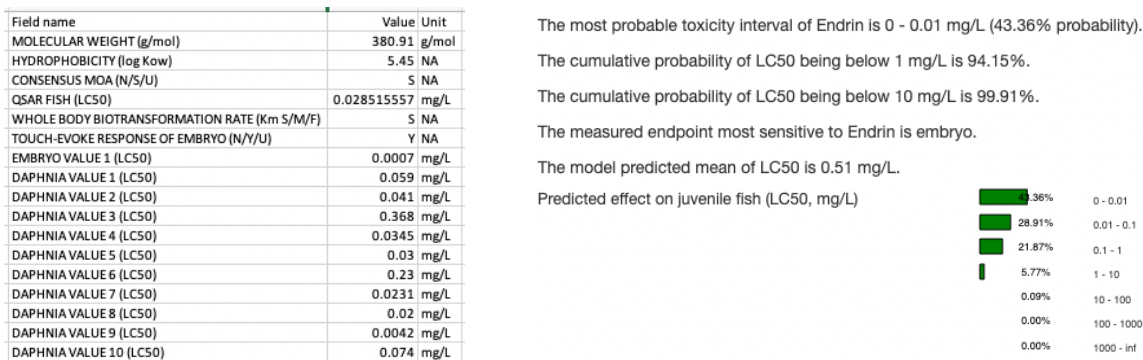


Figure 7: For chemical Endrin (left) data input and (right) conclusions.

The system correctly identifies the most probable toxicity interval of the substance with a high probability (43%) compared to the other intervals (all below 29%). For this substance only one fish embryo toxicity value was available (0.0007 mg/L). The supporting toxicity evidence was 10 daphnid toxicity values in the range 0.0042-0.37 mg/L, corresponding to extremely high to high toxicity. The one FET data point alone would also have given a correct prediction of extremely high toxicity. Nevertheless, the supporting information on toxicity to daphnids resulted in a more precise prediction (lower SD). The high consistency of toxicity to fish embryo and daphnids is in accordance with the purpose of this chemical, which is to kill both fish and insects (to which daphnids are closely related).

The endpoint most sensitive to Endrin is fish embryo, which suggests that fish in general will be more sensitive to this chemical than other species groups (represented by algae and daphnids). Consequently, the predicted toxicity to fish will overrule the toxicity to other endpoints, according to principles for chemical hazard assessment. Therefore, an accurate prediction of toxicity is crucial for cases like this, where fish is the most sensitive endpoint.

6.2 Triclosan

*Triclosan is an antibacterial and antifungal agent present in some consumer products, including toothpaste, soaps, detergents, toys, and surgical cleaning treatments.*³ Figure 8 shows the input data for the chemical substance Triclosan and the conclusion including the posterior distribution. The measured acute fish toxicity of Triclosan is 0.50 mg/L (average of 21 values), which corresponds to *high toxicity* (0.1-1.0 mg/L).

The system again correctly identifies the most probable toxicity interval of the substance with a 43% probability. In this case, there are also toxicity data available for algae, to which triclosan has only medium toxicity. However, the line of evidence that combines evidence from algae and daphnids gives higher weight to daphnids (as closer relatives to fish than algae), therefore the lower toxicity to algae does not prevent an accurate prediction for this substance. The measured endpoint most sensitive to Triclosan is algae, which means that

3. <https://en.wikipedia.org/wiki/Triclosan> accessed 31 May 2022.

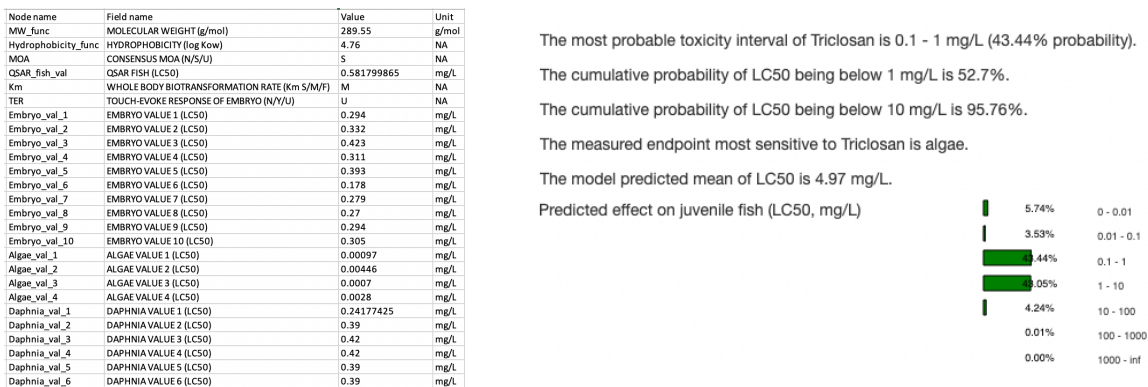


Figure 8: For chemical Triclosan (left) data input and (right) conclusions.

an accurate prediction of fish toxicity interval is crucial for the hazard assessment also in this case.

6.3 Carbamazepine

*Carbamazepine is an anticonvulsant medication used primarily in the treatment of epilepsy and neuropathic pain.*⁴ Figure 6 (left) shows the input data for the chemical substance Carbamazepine while Figure 6 (right) shows the conclusion and the posterior distribution. The measured acute fish toxicity of carbamazepine is 40 mg/L (average of 4 values), which corresponds to *low toxicity* (10-100 mg/L). The measured toxicity values for fish embryo are all in the range *very low toxicity* (100-1000 mg/L). However, since the posterior probability distribution is also influenced by the prior distribution, which is less extreme than the observed values, the predicted toxicity for this node is still *low toxicity*.

Therefore, the system correctly identifies the most probable toxicity interval of the substance to fish with a high probability (55%). The measured endpoint most sensitive to Carbamazepine is algae, which will therefore determine the conclusion of the hazard assessment instead of the fish. For algae there are only two measured values representing low toxicity (49 mg/L) and very low toxicity (167 mg/L), respectively. Again, the posterior probability distribution is influenced by the prior distribution which is less extreme than the observations, therefore the predicted most probable interval is low toxicity (55%). This example illustrates the importance of the prior probability distributions in this system, which can have a high influence on the posterior probabilities for example in cases where there are few measured values. The modelling of prior probability distributions will be further elaborated in other papers.

7. Conclusion and Future Work

The developed system is considered as a promising step towards a probabilistic and quantitative WOE approach to predict acute fish toxicity from fish embryo in combination with

4. <https://en.wikipedia.org/wiki/Carbamazepine> accessed 31 May 2022.

other types of evidence. Although traditional WOE assessments frameworks require a more comprehensive evaluation of the information available, this model supports the process by efficiently providing a prediction of toxicity, as well as an objective assessment of the consistency of the evidence. While the purpose of the system is to predict acute fish toxicity, the approach can be relevant in a more general setting for evaluating animal alternatives in regulatory toxicity testing, or even for other types of environmental assessments.

The paper describes how a Bayesian network is turned into a decision support system by combining different modelling approaches and linking the model to an interactive web interface. The Bayesian network was developed combining several different modelling approaches into a single parameterized model. The model is parameterized by the information on the substance group supplied by the user using function nodes and makes use of function nodes as well as discrete and continuous variables on different scales to achieve its purpose of predicting acute fish toxicity.

Future work will aim to improve the model performance by (1) refining the definition of substance group in order to provide more informative prior probability distributions of toxicity, (2) exploring the use of additional information sources such as predicted toxicity from chemical models (quantitative structure-activity relationships), (3) increasing the precision of the model predictions by assuming lower uncertainty for continuous value variables, (4) exploring the use of machine learning methods for optimising the weights of the lines of evidence, and (5) better characterising the applicability domain of the model.

Acknowledgments

The work was performed as part of the SWiFT project (Strengthening Weight of evidence for FET data to replace acute Fish Toxicity), see <http://niva.no/swift>). The SWiFT project was funded by the European Chemical Industry Council (CEFIC) as part of its Long-range Research Initiative (LRI) programme, project ECO51. The authors would like to thank the project's Monitoring Team for useful feedback as well as the reviewers for their insightful comments that have helped improve the quality of the paper.

References

- S. K. Andersen, K. G. Olesen, F. V. Jensen, and F. Jensen. HUGIN — a Shell for Building Bayesian Belief Universes for Expert Systems. In *Proc. of IJCAI*, pages 1080–1085, 1989.
- S. E. Belanger, A. D. Lillicrap, S. J. Moe, R. Wolf, K. Connors, and M. R. Embry. Weight of evidence tools in the prediction of acute fish toxicity. *Integrated Environmental Assessment and Management*, n/a(n/a), 2022.
- I. J. Good. Weight of evidence, corroboration, explanatory power, information and the utility of experiment. *J. R. Stat. Soc.*, B22:319–331, 1960.
- L. Kaikkonen, T. Parviainen, M. Rahikainen, L. Uusitalo, and A. Lehtikainen. Bayesian Networks in Environmental Risk Assessment: A Review. *Integrated Environmental Assessment and Management*, 17(1):62–78, 2021.
- U. B. Kjærulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer, 2nd edition, 2013.

- S. L. Lauritzen and F. Jensen. Stable local computation with mixed Gaussian distributions. *Statistics and Computing*, 11(2):191–203, 2001.
- A. Lillicrap, S. Belanger, N. Burden, D. D. Pasquier, M. R. Embry, M. Halder, M. A. Lampi, L. Lee, T. Norberg-King, B. A. Rattner, K. Schirmer, and P. Thomas. Alternative approaches to vertebrate ecotoxicity tests in the 21st century: A review of developments over the last 2 decades and current status. *Environmental Toxicology and Chemistry*, 35(11):2637–2646, 2016.
- A. Lillicrap, S. J. Moe, R. Wolf, K. A. Connors, J. M. Rawlings, W. G. Landis, A. Madsen, and S. E. Belanger. Evaluation of a Bayesian Network for Strengthening the Weight of Evidence to Predict Acute Fish Toxicity from Fish Embryo Toxicity Data. *Integrated Environmental Assessment and Management*, 16(4):452–460, 2020.
- I. Linkov, O. Massey, J. Keisler, I. Rusyn, and T. Hartung. From "weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX*, 1(32):3–8, 2016.
- A. L. Madsen, F. Jensen, U. B. Kjærulff, and M. Lang. HUGIN - The Tool for Bayesian Networks and Influence Diagrams. *International Journal on Artificial Intelligence Tools* 14, 3:507–543, 2005.
- A. L. Madsen, M. Karlsen, G. Barker, A. Garcia, J. Hoorfar, F. Jensen, and H. Vigre. A Software Package for Web Deployment of Probabilistic Graphical Models. In *Frontiers in Artificial Intelligence and Applications, Volume 257, Twelfth SCAI*, pages 175–184, 2013.
- A. L. Madsen, F. Jensen, M. Karlsen, and N. Sønderberg-Jeppesen. Bayesian Networks with Function Nodes. In *Proc. of PGM*, pages 286–301, 2014.
- S. J. Moe, A. L. Madsen, K. A. Connors, J. M. Rawlings, S. E. Belanger, W. G. Landis, R. Wolf, and A. D. Lillicrap. Development of a hybrid Bayesian network model for predicting acute fish toxicity using multiple lines of evidence. *Environmental Modelling and Software*, pages 1–17, 2020.
- M. Paparella, S. Scholz, S. Belanger, T. Braunbeck, P. Bicherele, K. Connors, C. Faßbender, M. Halder, A. Lillicrap, R. Liska, K. Schirmer, G. Stoddart, P. Thomas, and S. Walter-Rohde. Limitations and uncertainties of acute fish toxicity assessments can be reduced using alternative methods. *ALTEX - Alternatives to animal experimentation*, 38(1):20–32, Jan. 2021.
- G. Suter, S. Cormier, and M. Barron. A weight of evidence framework for environmental assessments: Inferring qualities. *Integrated Environmental Assessment and Management*, 13(6):1038–1044, 2017a.
- G. Suter, S. Cormier, and M. Barron. A weight of evidence framework for environmental assessments: Inferring quantities. *Integrated Environmental Assessment and Management*, 13(6):1045–1051, 2017b.