"Structural Alerts A New Classification Model to Discriminate Excess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay"

P.C. Von Der Ohe, R. Kunhe, R. Ebert, R. Altemburger, M. Liess, and G. Schuurmann

Daniel Dubois

Giordano Tamburrelli

Outline

- Motivation
- Excess Toxicity Ratio
- Classification Models
- Classification Performances
- Validation of Results
- Combined Approach
- Conclusions and future work
- SE Considerations

Motivation

- 48h Acute Daphnid Toxicity test used as a method to evaluate aquatic toxicity in the European legislation.
- Lack of analysis of QSAR models related to this problem.
- No classification scheme to discriminate narcotic effect range and excess toxicity in daphnids.

What are Daphnids?

- Daphnids (we consider *Daphnia Magna*):
  - Small crustacean, 0.5-5mm in length

Methodology

1. Excess Toxicity Ratio (T_e) Definition
2. Classification through T_e of compounds in the dataset
3. Structural description derivation and CMs definition
4. CMs performance evaluation
5. CMs validation

Toxicity

"Toxicity is the degree to produce illness or damage to an exposed organism"

- One possible Toxicity metric: LC_{50}, concentration of a compound that kills 50% of the exposed population.
- Toxicity tests:
  - Acute (one exposure);
  - Chronic (multiple exposures).
- Toxicity classes:
  - Narcotic Effect;
  - Excess Toxicity.
- In this study we consider only Acute Toxicity Tests.
- Problem: How to discriminate the two toxicity classes?

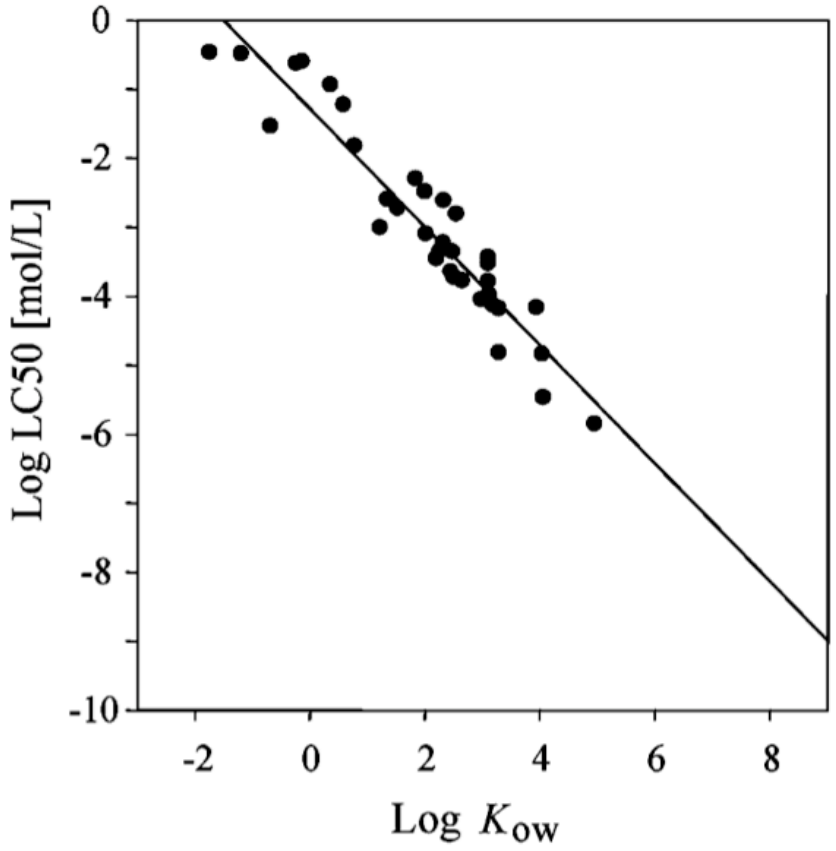
Excess Toxicity Ratio

$$T_e = \frac{LC_{50}\,(\text{baseline})}{LC_{50}\,(\text{exp})}$$

- LC_{50}(***baseline***) is the arithmetic mean of 36 compounds reported in literature as narcotics. We call this: *baseline toxicity*.
- LC_{50}(***exp***) is the experimental toxicity.
- T_e ($\log T_e$) **meaning**:
  - <100 (<2): the compound is **narcotic level**
  - ≥100 (≥2): the compound is **excess toxic**
- Why a cut-off value of 100?
  - A factor 10 covers data uncertainty due to the heterogeneous sources of experimental data.
  - Another factor of 10 is related to polar/nonpolar issues in molecules.

Baseline Toxicity QSAR Model

Log LC_{50}(mol/L) = -0.857(±0.049)LogK_{ow}-1.281(±0.125)



K_{ow} is the octanol-water partition coefficient.

n=36 (number of compounds considered as narcotics in literature)
R^2=0.9.

Solid line is the baseline toxicity.

Classification Models

- Goal: predict the toxicity of a compound without using experimental tests.
- Recipe: classification models based on structural descriptors.
- Model Performance evaluation:

$$\text{concordance} = \frac{1}{N}\sum_{i=1}^{2} n_{\text{cpred}}(\text{i})$$

$$\text{sensitivity} = \frac{n_{\text{cpred}}}{n_{\text{exp}}} \qquad \text{predictivity} = \frac{n_{\text{cpred}}}{n_{\text{pred}}}$$

Training Set

- A dataset of 264 compounds has been used as training set.
- Chemicals that contain metals, the ones that have LC_{50}>$Solubility_{water}$, and finally inorganic ones have been excluded.
- Classification results using Excess Toxicity Ratio:
  - 78 excess toxic compounds;
  - 186 narcotic level compounds.
- Three Classification Models (CM1, CM2, CM3) have been derived:
  - visual inspection of the chemical structures;
  - identification of common patterns.

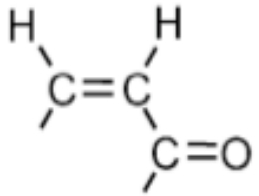
Classification Model 1

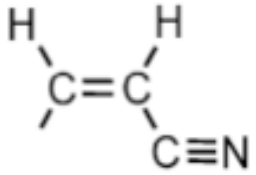
- Narcotic Level Condition:
  - Organic chemicals that contain any combination of C, H, O, and halogens, excluding α,β-unsaturated carbonyl groups as electrophilic functionalities.
- Excess Toxicity Condition:
  - All the compounds that are not included above.

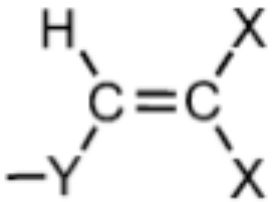
Classification Model 2

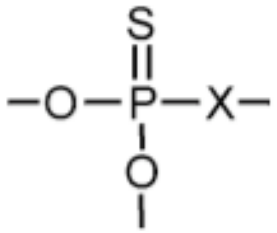
- Narcotic Level Condition:
  - All the compounds that are not excess toxic.
- Excess Toxic Condition:
  - All the compounds with the following *Structural Alerts*...

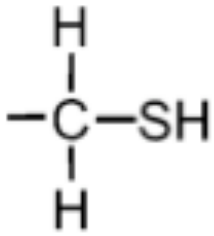
Structural Alerts for CM2

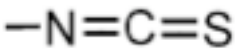


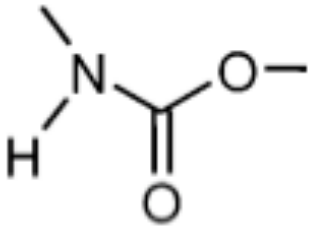


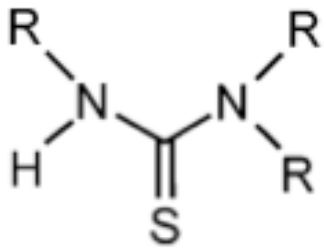


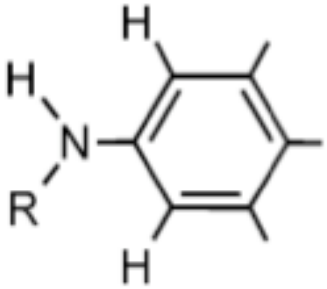


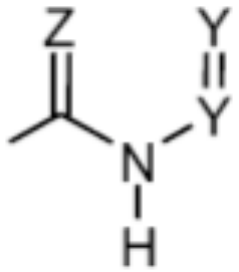








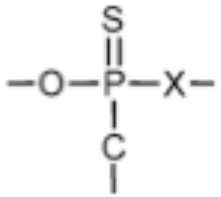


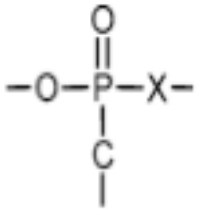


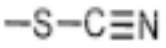
Classification Model 3

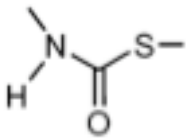
- Extension of CM2 with different Structural Alerts.
- Narcotic Level Condition:
  - All the compounds that are not excess toxic.
- Excess Toxic Condition:
  - All the compounds with the following *Structural Alerts*...

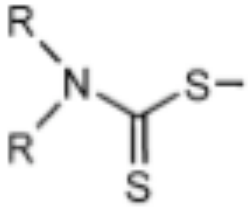
Additional Structural Alerts for CM3

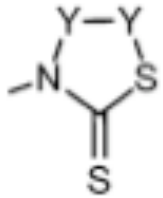


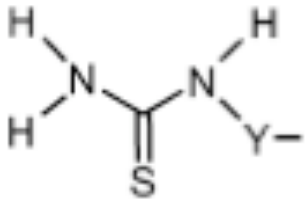


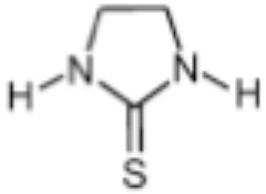












Alternative Classification Models

- The following models are taken from literature and have been derived using different approaches.
  - 4-Mode Of Action CM
    - *Excess toxicity*: reactive chemicals and specifically acting chemicals.
    - *Narcotic Effect level*: inert chemicals and less inert chemicals.
  - 7-Mode Of Action CM
    - *Excess toxicity*: oxidative phosphorylation uncouplers, electrophiles and proelectrophiles, AChE inhibitors, CNS seizure agents.
    - *Narcotic effect level*: baseline narcotics, polar narcotics, ester narcotics.
  - 2-Mode of Action CM
    - *Excess toxicity*: electrophiles, proelectrophiles.
    - *Narcotic Effect Level*: all other organic material.

Outline

- Motivation
- Excess Toxicity Ratio
- Classification Models
- Classification Performances
- Validation of Results
- Combined Approach
- Conclusions and future work
- SE Considerations

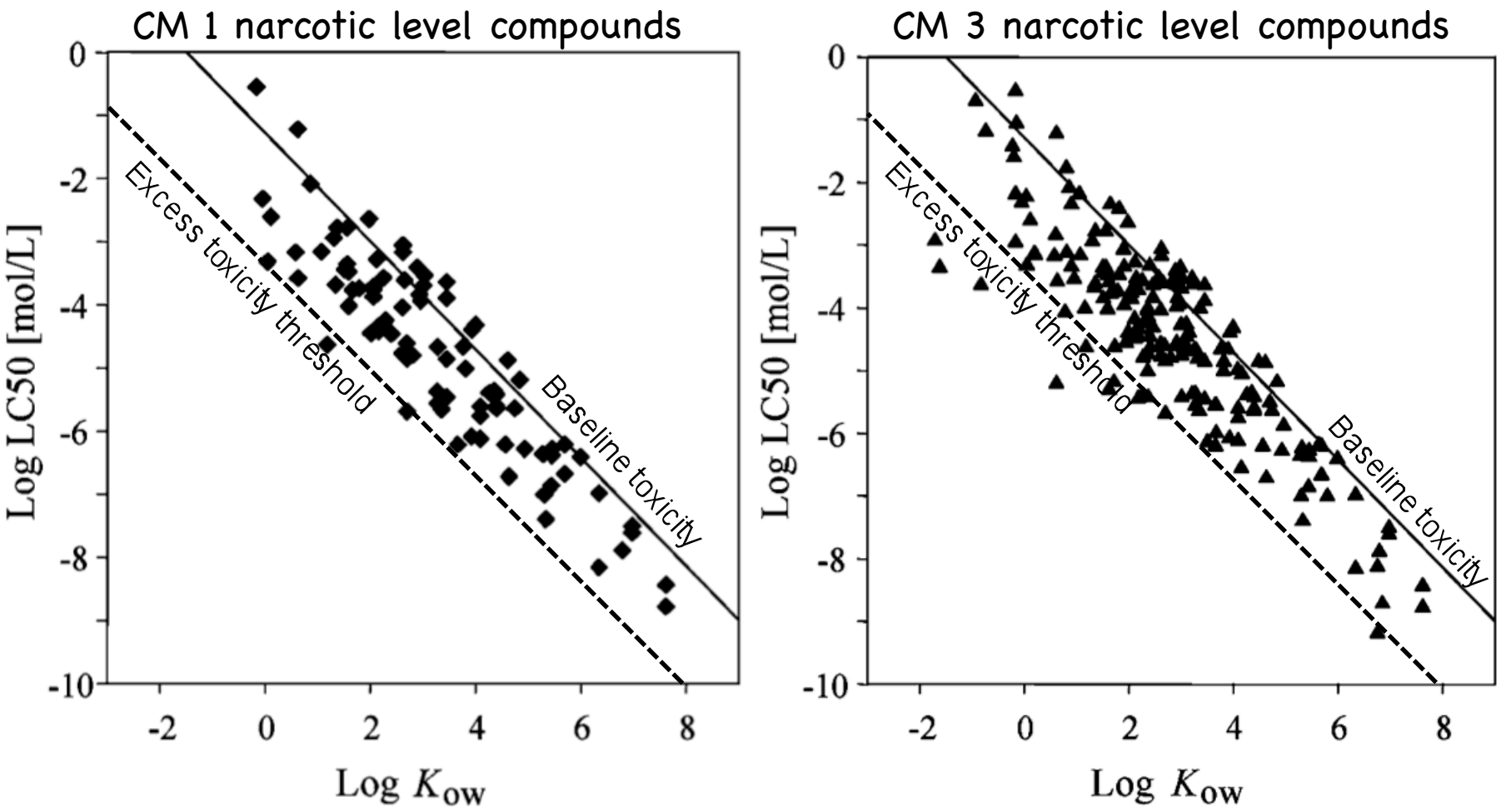
Classification Performances (1)

models	predicted category	narcotic effect level	excess toxicity	total
CM1	narcotic effect level	89	2	91
	excess toxic	97	76	173
	total	186	78	264
CM2	narcotic effect level	179	26	205
	excess toxic	7	52	59
	total	186	78	264
CM3	narcotic effect level	179	14	193
	excess toxic	7	64	71
	total	186	78	264
4-MOA CM	narcotic effect level	59	11	70
	excess toxic	44	25	69
	total	103	36	139
7-MOA CM	narcotic effect level	125	45	170
	excess toxic	61	33	94
	total	186	78	264
2-MOA CM	narcotic effect level	169	65	234
	excess toxic	17	13	30
	total	186	78	264

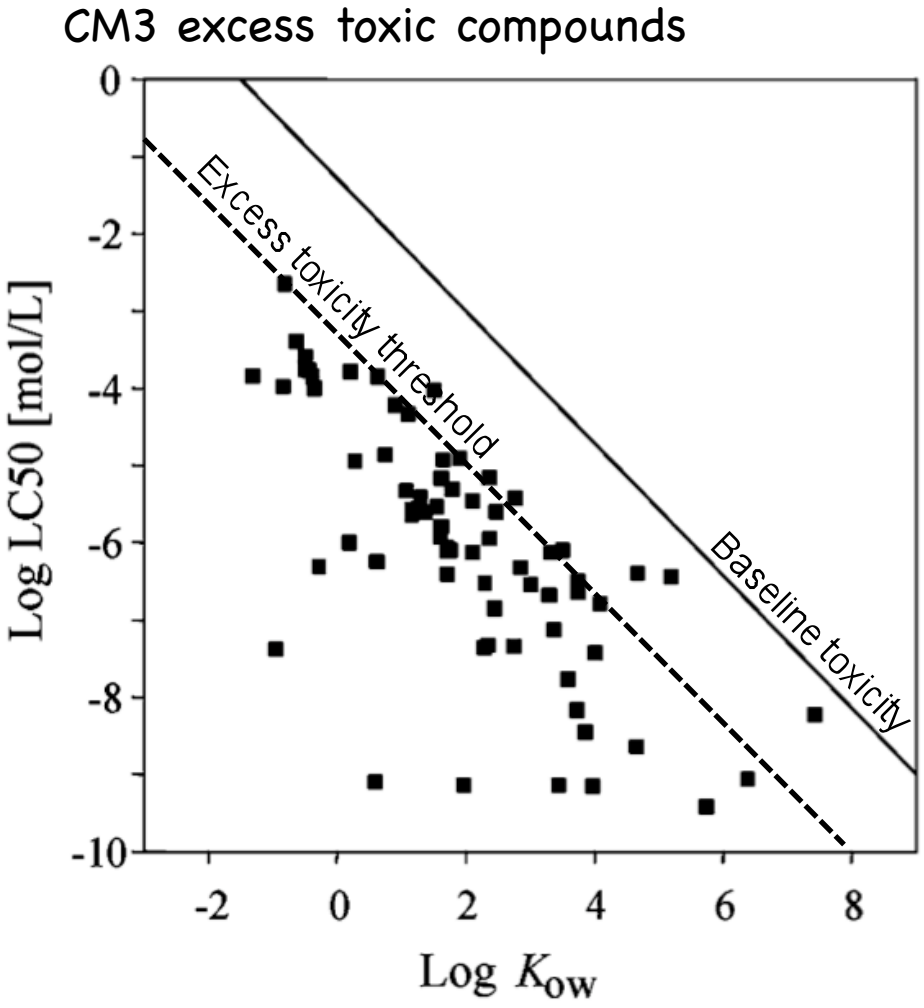
Classification Performances (2)

model	concordance	category	sensitivity	predictivity
CM1	0.625	narcotic effect level	0.478	0.978
		excess toxicity	0.974	0.439
CM2	0.875	narcotic effect level	0.962	0.873
		excess toxicity	0.666	0.881
CM3	0.920	narcotic effect level	0.962	0.927
		excess toxicity	0.820	0.901
4-MOA CM	0.604	narcotic effect level	0.573	0.843
		excess toxicity	0.694	0.362
7-MOA CM	0.598	narcotic effect level	0.672	0.735
		excess toxicity	0.423	0.351
2-MOA CM	0.689	narcotic effect level	0.909	0.722
		excess toxicity	0.167	0.433

QSAR Modeling of Narcotic Level Toxicity



QSAR Modeling of Excess Toxicity



Validation Results

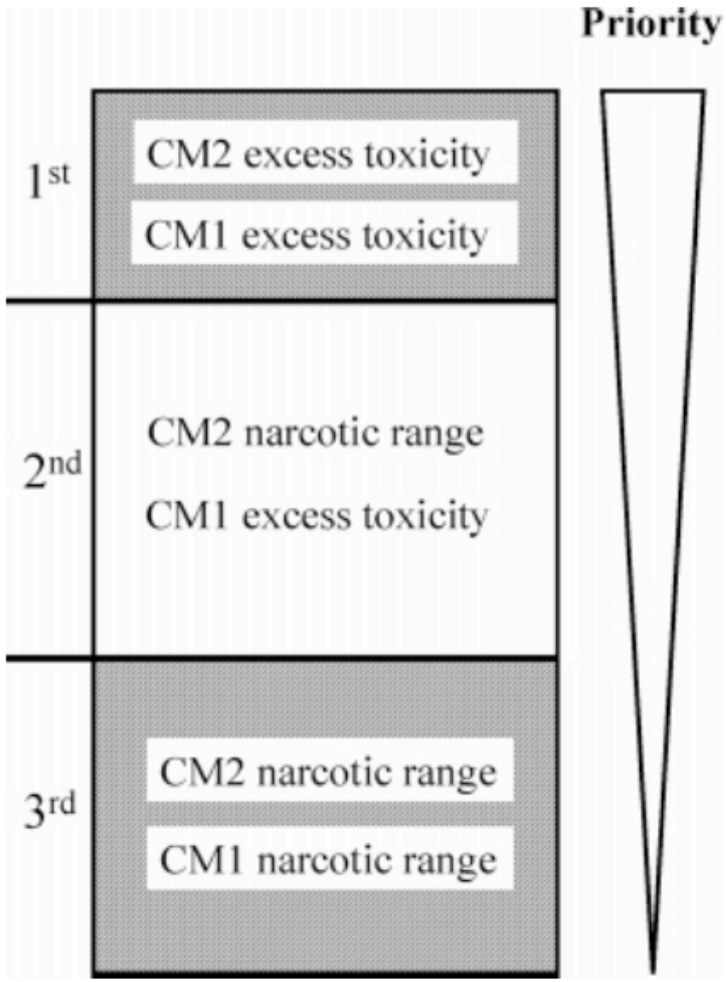
- Need of a compound set different from the training set.
- New validation sets of 74 compounds using different Daphnids:
  - **Validation set 1**: all 74 compounds.
  - **Validation set 2**: 27 compounds (the former 74, minus the 47 compounds that belong also to the Daphnia Magna training set).

model	concordance	category	sensitivity	predictivity
validation set 1 (74 compounds)				
CM1	0.662	narcotic effect level	0.574	0.940
		excess toxicity	0.900	0.439
CM2/CM3	0.932	narcotic effect level	0.981	0.930
		excess toxicity	0.800	0.941
validation set 2 (27 compounds)				
CM1	0.481	narcotic effect level	0.435	0.909
		excess toxicity	0.750	0.188
CM2/CM3	0.963	narcotic effect level	1.000	0.958
		excess toxicity	0.750	1.000

Statistical Considerations

- CM1 is strong in:
  - identifying excess toxicity (*recognition power);*
  - predicting narcotic effectlevels (*prediction power*).
- CM2 shows an overall performance between CM1 and CM3.
- CM3 is the best-performing scheme, but its structural rules require more experimental data for confirmation.
- The result suggests a properly combined use of CM1+CM2+CM3.

Combined Two-Step Approach



Conclusions

- The paper presented an experimental method to distinguish excess toxic compounds from narcotic ones.
- Then it derived some classification models to predict the acute toxicity of compounds for daphnia magna.
- Finally it compared the results with the ones obtained with different models and with different types of daphnids.

Future Work

- Use of a more complete dataset;
- Extension of this QSAR-supported risk assessment to other invertebrate species and other taxonomic groups;
- Do some further experimentation on combined techniques, especially for the CM3 model.

A SE Oriented Consideration…

- Software clones can be represented using abstract syntax trees representation.
- A similar approach working on abstract syntax trees of InChi (or SMILES) could find automatically the structural alerts.
- More generally: the methodology of this paper can be improved by using automatic techniques for the recognition of the structural alerts (also other more traditional approaches like neural networks and data mining).

Questions?