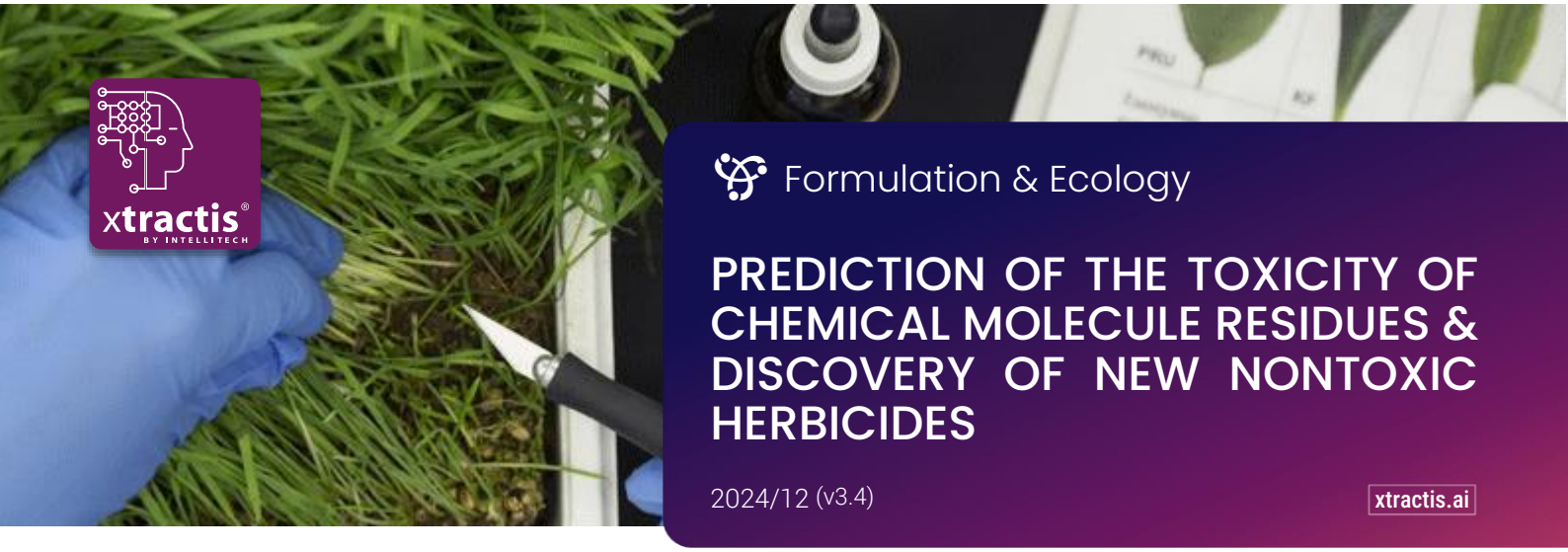


Formulation & Ecology

PREDICTION OF THE TOXICITY OF CHEMICAL MOLECULE RESIDUES & DISCOVERY OF NEW NONTOXIC HERBICIDES

2024/12 (v3.4)

xtractis.ai

PROBLEM DEFINITION

GOALS

Design an AI-based decision system that accurately predicts the toxicity of residues of chemical molecules on animals in a rational and explainable way.

Quickly discover the formulation of new molecules that have minimum impact on wildlife.

PROS & BENEFITS

- Help agrochemical manufacturers to measure the toxicity of existing molecules and check compliance with environmental regulations.
- Help chemists design new bioactive molecules with less impact on small animals.
- Reduce testing on laboratory animals.

REFERENCE DATA

Source:
US EPA,
German BBA,
EU-project SEEM

Variables to Predict

6 variables; the effectiveness of the molecule as an herbicide, and the toxicity level $\text{Toxicity} = \log_{10}(\frac{1}{\text{lethal\_dose}})$ for 5 animal species: Trout Toxicity ∈ [-2.33, 7.74] $\log_{10}$(l/mmol), Daphnia Toxicity ∈ [-0.952, 7.673] $\log_{10}$(l/mmol), Oral* Quail Toxicity ∈ [-1.322, 2.278] $\log_{10}$(kg/mmol), Diet** Quail Toxicity ∈ [-2.265, 1.435] $\log_{10}$(kg/mmol), Bee Toxicity ∈ [- 1.007, 5.527] $\log_{10}$(bee/µmol).

*when the molecule is put down the quail's throat, **when it is put in the quail's food.

Predictive Variables

Each molecule has a chemical profile characterized by 99 potential predictors [2D/3D chemical descriptors, geometry, topology, electro-topology, physicochemistry, constitution, etc.] with 58.78% of missing values.

Observations

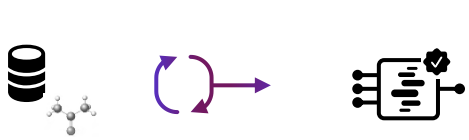
358 chemical molecules which toxicity was tested on different animal species: 229 molecules on trout, 220 on water flea (daphnia), 96 on quail, and 88 on bee.

MODEL TYPE

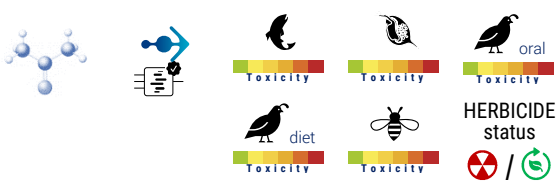
Regression | Multinomial Classification | Binomial Classification | Scoring

XTRACTIS SOLUTION – PHASE 1: INDUCE A MODEL FOR EACH SPECIES

STEPS



Reference Data — INDUCTION — XTRACTIS Top-Models



New Molecules — DEDUCTION — Automated Decision

SOFTWARE ROBOTS

XTRACTIS® **REVEAL**
Delivers the decision systems + their Structure & Performance Reports

XTRACTIS® **PREDICT**
Delivers the decisions + the Prediction Reports explaining each reasoning

OUTCOME

☑ **Intelligible Model, Explainable Decisions**
We get 6 decision systems; each system is composed of gradual rules without chaining, and each rule uses some of the variables that XTRACTIS identified as predictors.
The most intelligible model uses 14 rules combining 29 predictors and the less intelligible one uses 11 rules sharing 39 predictors.

☑ **High Predictive Capacity**
Overall, models have good Real Performances in Test.

☑ **Ready to Deploy**
Models compute real-time predictions up to 70,000 decisions/sec., off/online (API).

TOP-MODELS INDUCTION

INDUCTION PROCESS

Powered by:

XTRACTIS® REVEAL v13.0.46434

1. We launch 2,000 inductive reasoning strategies for each variable to predict. Each strategy is applied to the Training 70% / Validation 30% dataset to get a reliable assessment of the descriptive and predictive performances. The Test 15% dataset is used to evaluate the real performance of the models.
2. Each strategy generates a unitary model called **Individual Virtual Expert (IVE)**.
3. For each variable to predict, the top-IVE is the one that has the best predictive performance, close to its descriptive performance, and with the fewer predictors and rules, among the 2,000 induced IVEs. We obtain 6 top-IVEs that have 4 to 14 rules sharing 20 to 39 predictors, depending on the model.

Total number of induced unitary models (for all models)	Criterion for the induction optimization	Validation criterion for the top-models selection
12,000 IVEs	Regression: RMSE Binomial Classification: F_1-Score	Regression: RMSE Binomial Classification: F_1-Score

TOP-MODELS STRUCTURE

The 6 top-IVE models have a poor to very good intelligibility for complex phenomena and for a dataset that includes about 59% of missing values!

Model	type	Number of rules	Number of predictors	Number of predictors per rule | on average
TROUT Toxicity	Regression	8	32	3 to 14 | 7.8
DAPHNIA Toxicity	Regression	13	20	1 to 9 | 4.8
ORAL QUAIL Toxicity	Regression	5	27	6 to 13 | 8.8
DIET QUAIL Toxicity	Regression	11	39	5 to 28 | 10.3
BEE Toxicity	Regression	14	29	1 to 8 | 4.6
HERBICIDE Status	Binomial Classification	4	28	8 to 16 | 12.3

Each of their Structure Reports reveals all the internal decision logic and ensures that the human expert understands each model. These decision systems are transparent models that can be audited by the domain expert and certified by the regulator before being deployed to end-users.

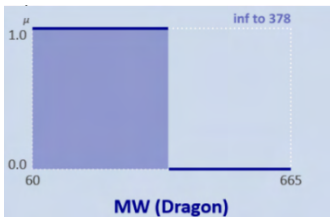
For example, here is the structure of the "TROUT Toxicity" Model:

TROUT Toxicity

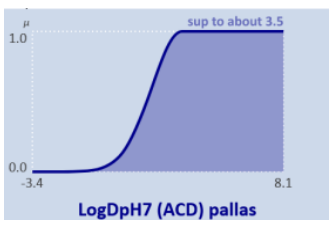
PREDICTORS

- 32 predictors
- Ranked by impact significance:
  3 strong, 25 medium, and 4 weak signals
  (#1 Mp (Dragon) /#2 LogDpH7 (ACD) pallas #3 …)
- Labeled by binary and fuzzy classes.
  Examples:

binary interval "*inf to 378*"



fuzzy interval "*sup to about 3.5*"



RULES

- 8 connective fuzzy rules without chaining
- 3 to 14 predictors per rule (on average 7.8 predictors per rule)
- Example:
  fuzzy rule R3 uses 8 predictors and concludes {-1.760}.
  7 other fuzzy rules complete this model.

IF	D/Dr09 (Dragon)	IS	inf to about 0
AND	LogDpH7 (ACD) pallas	IS	sup to about 3.5
AND	Mp (Dragon)	IS	inf to about 0.601
AND	MW (Dragon)	IS	inf to 378
AND	nHDon (Dragon)	IS	about 0.00
AND	C-032	IS	sup to about 2.33
AND	BEHm8 (Dragon)	IS	inf to about 0.00
AND	nN (Dragon)	IS	sup to about 3.88
THEN	Toxicity TROUT	IS	-1.760

TOP-MODELS PERFORMANCE

The top-IVE performances, measured in Training/Validation/Test, guarantee the models' predictive and real performances.

Model	Dataset	DESCRIPTIVE Performance	PREDICTIVE Performance	REAL Performance
		70% Training	15% Validation	15% Test
TROUT Toxicity	RMSE	**0.718 (5.94%)**	**0.721 (5.96%)**	**0.914 (7.56%)**
	Correlation	0.881	0.902	0.866
DAPHNIA Toxicity	RMSE	**0.817 (7.94%)**	**0.824 (8.00%)**	**1.045 (10.15%)**
	Correlation	0.887	0.900	0.850
ORAL QUAIL Toxicity	RMSE	**0.388 (8.99%)**	**0.390 (9.04%)**	**0.407 (9.42%)**
	Correlation	0.910	0.909	0.871
DIET QUAIL Toxicity	RMSE	**0.293 (6.61%)**	**0.293 (6.60%)**	**0.419 (9.44%)**
	Correlation	0.925	0.956	0.760
BEE Toxicity	RMSE	**0.584 (7.45%)**	**0.598 (7.63%)**	**0.649 (8.28%)**
	Correlation	0.913	0.908	0.878
HERBICIDE Status	F_1-Score	**80.68%**	**81.08%**	**75.68%**
	Classif. Error	34 (13.65%)	7 (12.96%)	9 (16.67%)

We can now use these models with XTRACTIS PREDICT to perform a virtual screening and find a new herbicide formulation with the lowest toxicity. As these 6 models share 86 predictors out of the 99 potential predictors, it would take an infinite time to gather all predictions for all possible combinations of inputs, given the huge dimension of the decision space. And if we proceed randomly, we will probably discover only sub-optimal solutions. This is why the use of the XTRACTIS OPTIMIZE abductive robots ensures the discovery of the most optimal formulations of new herbicides minimizing toxicity for all species.

EXPLAINED PREDICTION FOR 1 CASE FROM THE TEST SET (Trout Toxicity Model)

CASE
(not used in Training/Validation)

Profenofos MOLECULE

actual value = 4.250	
C-031 (Dragon)	0.00
D/Dr03 (Dragon)	0.0
D/Dr09 (Dragon)	0
Log P (Cache)	5.04
LogDpH7 (ACD) pallas	4.6
…	…
C-032	0.00
X1Av	*Missing Value*
T(S..S)	*Missing Value*
JGI6	*Missing Value*
BEHm8 (Dragon)	2.51
Kfldx (Codessa)	7.1
nN (Dragon)	0.00
TI2 (Dragon)	3.10
AMW	11.3
O-058	1.00
DFG0047	No chain
nOHPh (Dragon)	No chain
nxch3 (MDL)	No chain
nNR2Ph	No chain

Real Time

DEDUCTIVE INFERENCE OF RULES

For this molecule, 7 rules are triggered with different firing degrees, to conclude 4.826 $\log_{10}$(l/mmol)

# Rule	Conclusion	Firing Degree
R1 →	-2.287	0.074
R3 →	-1.760	1.09E-05
R4 →	0.080	0.045
R5 →	2.482	0.527
R6 →	3.410	0.379
R7 →	5.725	>0.999
R8 →	7.654	0.571
Prediction	4.826	

Powered by: XTRACTIS® PREDICT v13.0.46434

AUTOMATED DECISION

NUMBER OF TRIGGERED RULES
7 / 8

FUZZY PREDICTION
{ 5.725 | >0.999,
7.654 | 0.571,
2.482 | 0.527,
3.140 | 0.379,
-2.287 | 0.074,
0.080 | 0.045,
-1.760 | 1.09E-05 }

FINAL PREDICTION
{ 4.826 }

The system delivers a prediction of 4.826 $\log_{10}$(l/mmol), rather close to the value of 4.250 $\log_{10}$(l/mmol) measured in the experiment:

This molecule's toxicity on the trout is very high

XTRACTIS SOLUTION – PHASE 2: DISCOVER THE MOST OPTIMAL SOLUTIONS

REQUESTS

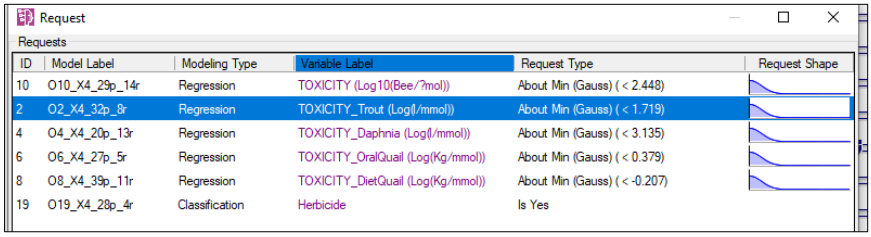
☑ **Min and Max bounds Research**.

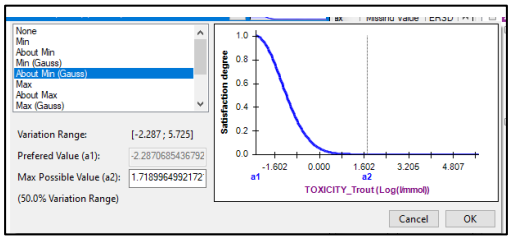
The first step is to discover the mimum and the maximum of each regression model, thanks to the XTRACTIS OPTIMIZE abductive robots.

☑ **Fuzzy Multi-Objective Request**.

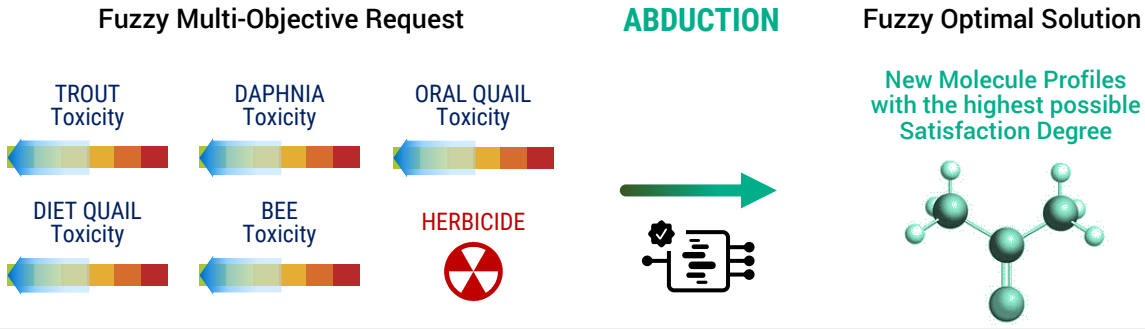
We define a collection of objectives on the 6 variables to predict. Each Toxicity objective is defined by a fuzzy request, i.e., a flexible request which accepts a continuous satisfaction degree between 0 and 1, except for HERBICIDE whose objective is defined by a binary request.

In this scenario, we want to **discover a new herbicide with minimal toxicity for the 5 species**. The global request is thus defined by the fuzzy conjunction of 6 elementary requests. "*About Min*" means that we try to be the nearest possible to the minimum of each Toxicity model.





PROCESS



SOFTWARE ROBOTS

XTRACTIS® **OPTIMIZE**

OUTCOME

☑ **Fuzzy Abduction**.

In total, the 6 models share 86 predictors which define the molecular profile. The XTRACTIS OPTIMIZE robots explore the 86-dimension piecewise continuous decision space of possible molecules to satisfy simultaneously the 6 objectives with the highest possible degree. The robots develop their abductive reasoning collectively, first in competition, then in co-operation and finally they evolve their reasoning strategies.

☑ **Fuzzy Optimal Solution**.

For this strongly constrained problem, the binary satisfaction of the 6 requests is impossible, but the advantage of fuzzy mathematics is to obtain a fuzzy optimal solution, with the highest satisfaction degree for each request.

☑ **New Molecule Profile**.

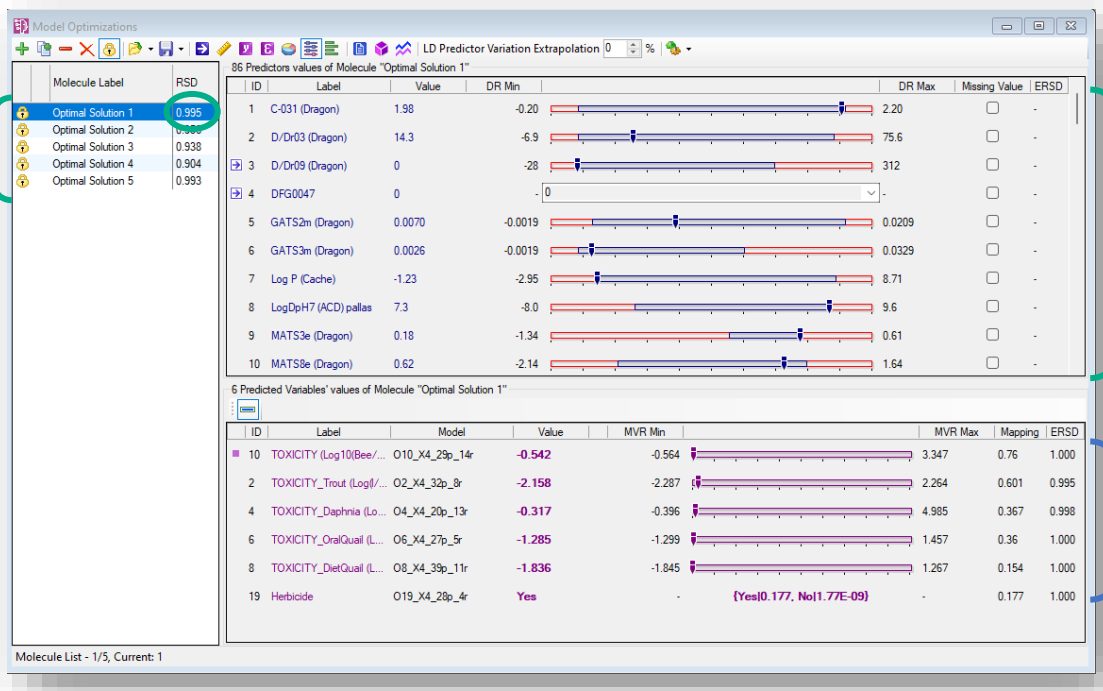
XTRACTIS OPTIMIZE discovers **a new molecule of herbicide with a global satisfaction degree of 0.995**. This is quite impressive given the fact that the most nontoxic herbicide already known in the training dataset achieves an overall satisfaction degree of only 0.063!

Powered by:

XTRACTIS® OPTIMIZE
v13.0.46434

The best new herbicide discovered by XTRACTIS OPTIMIZE has a Request Satisfaction Degree (RSD) = 0.995

Molecules discovered by XTRACTIS OPTIMIZE, ranked according to their RSD.

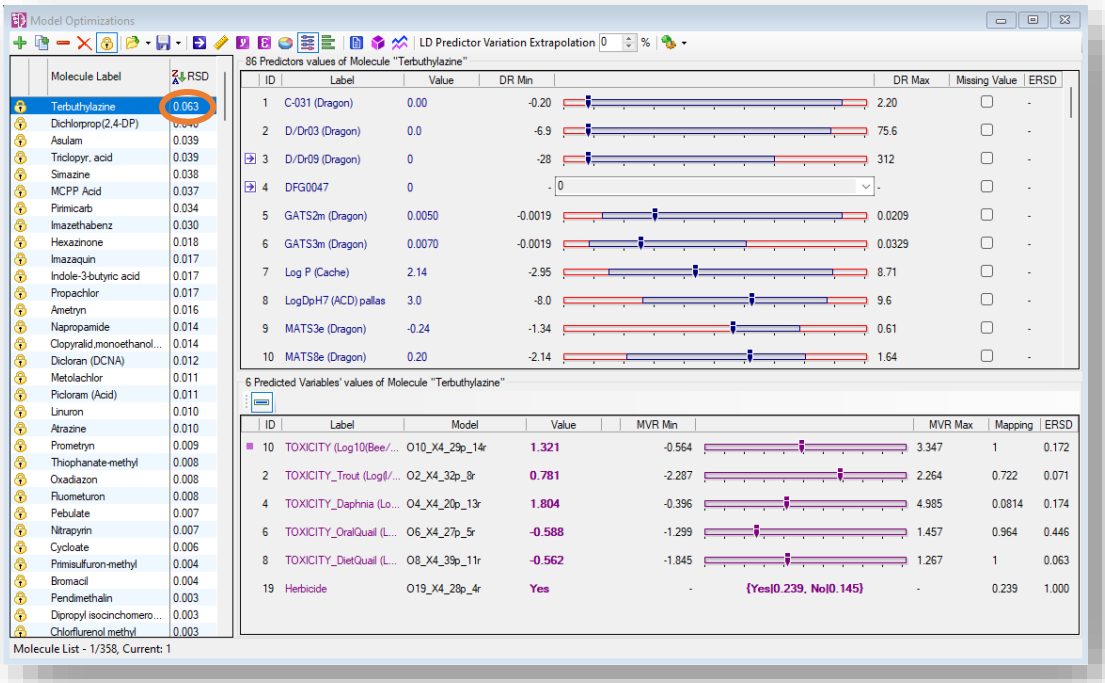


9 first values of the molecular profile

Predictions of the molecule's toxicities and its status as an herbicide

Terbuthylazine is the best herbicide from the Training dataset with a Request Satisfaction Degree (RSD) = 0.063

List of molecules from the reference dataset, ranked according to their RSD.



You can access a live demonstration replay of the Optimization process for this Use Case, including comments and detailed explanations. If you wish to watch this video, please visit our xtractis.ai website and request a credential by clicking on "Log in to watch Demos" from the header menu.