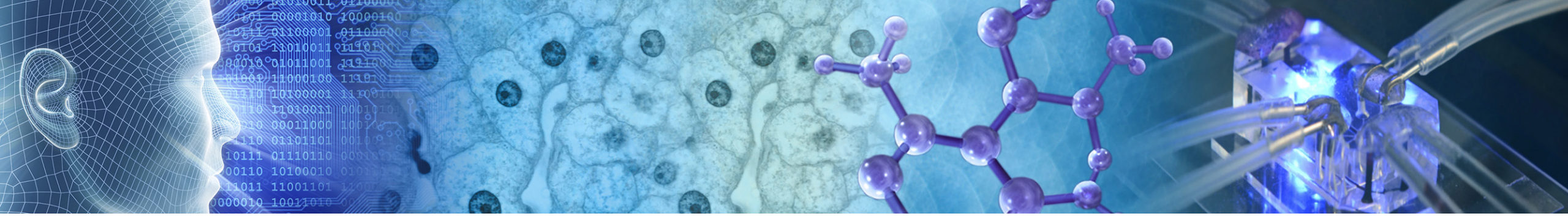


Collaborative modeling efforts for the AcuteTox endpoints

Kamel Mansouri

Predictive Models for Acute Oral Systemic Toxicity Workshop
April 11-12, 2018

Disclaimer: ILS and ScitoVation staff provide technical support for NICEATM, but do not represent NIEHS, NTP, or the official positions of any federal agency.

Agency for Toxic Substances and Disease Registry • Consumer Product Safety Commission • Department of Agriculture
Department of Defense • Department of Energy • Department of the Interior • Department of Transportation
Environmental Protection Agency • Food and Drug Administration • National Institute for Occupational Safety and Health
National Institutes of Health • National Cancer Institute • National Institute of Environmental Health Sciences
National Library of Medicine • Occupational Safety and Health Administration

Outline:

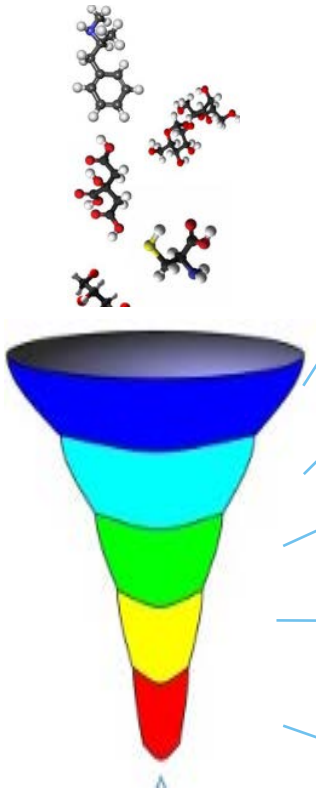
- Preparation for modeling
  - Available data overview
  - Chemical structures standardization and processing
  - Training set/evaluation set split
  - Prediction set preparation
- Modeling efforts
  - Participants
  - Submissions
  - Evaluation
  - Analysis
- Consensus
  - Analysis
  - Combining predictions
  - Analysis of the consensus predictions
  - Implementation of the consensus models to screen new chemicals
- Conclusions

Available data for modeling

15,688 chemicals total
21,200 LD50 values

- Very toxic: 11886
- Nontoxic: 11871
- EPA: 11755
- GHS: 11845
- LD50: 8908

QSAR-ready standardization workflow

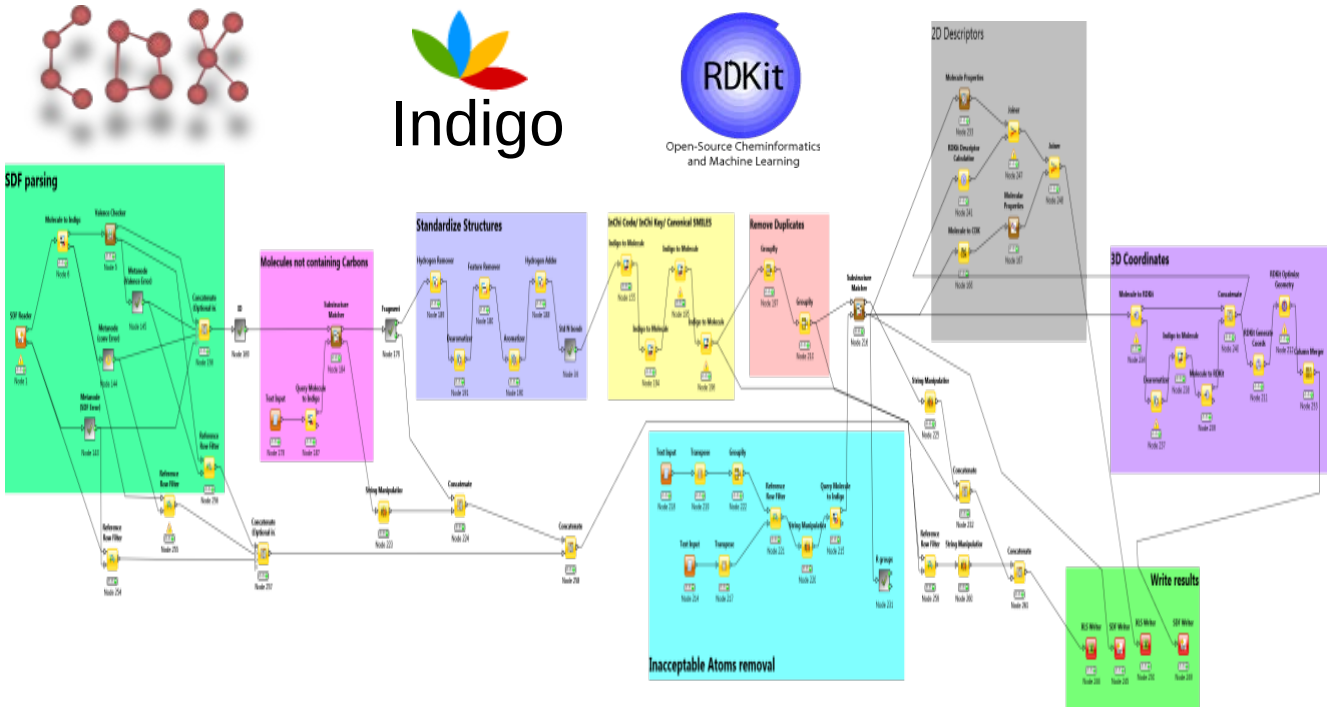


QSAR-ready structures

- Remove inorganics and mixtures
- Clean salts and counterions
- Normalize of tautomers
- Remove of duplicates
- Final inspection

Aim of the KNIME workflow:

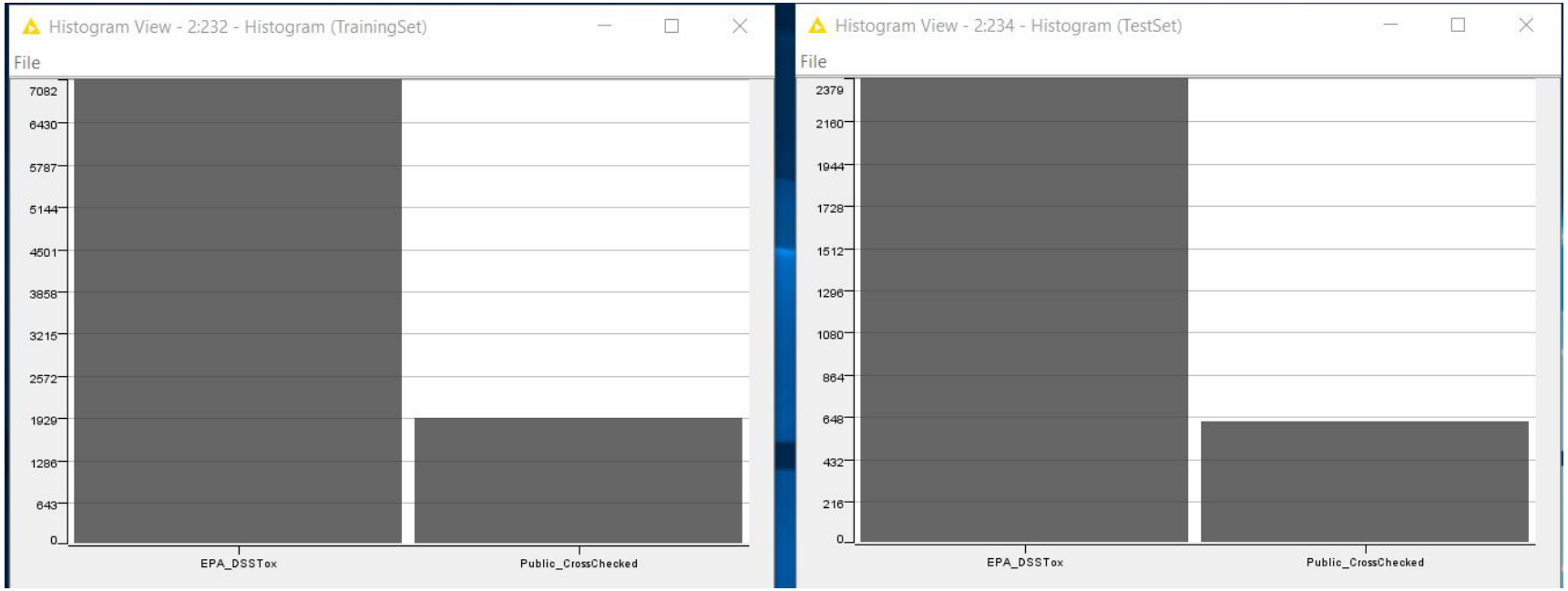
- Combine different procedures and ideas
- Minimize the differences between the structures used for prediction
- Produce a flexible free and open source workflow to be shared



Mansouri et al. (http://ehp.niehs.nih.gov/15-10267/)

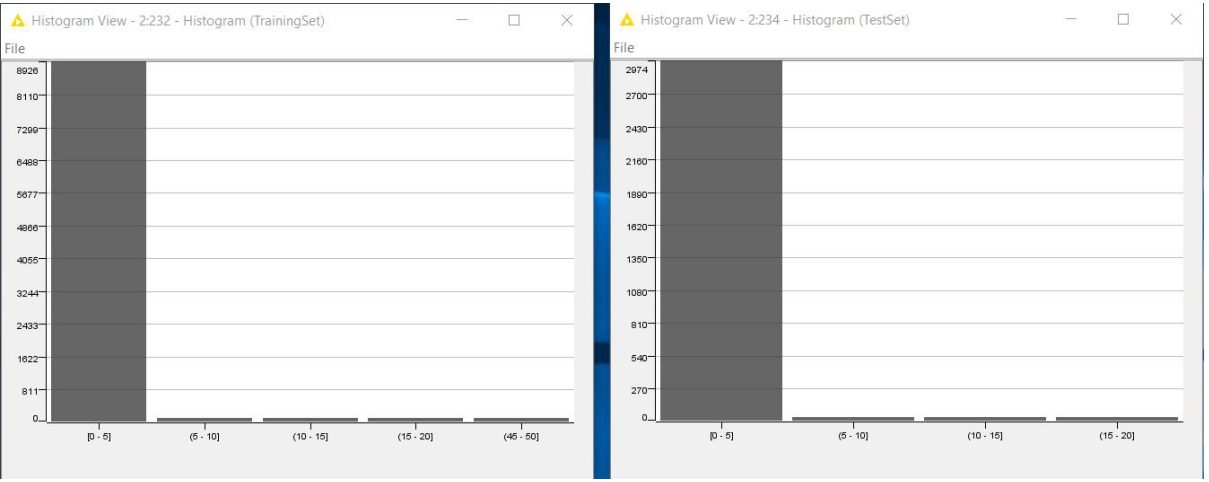
Training set/evaluation set split

- The same training and test chemicals across all endpoints
- Split into training (75%) and evaluation (25%)
- Similar distributions and variability for values and categories
- Similar distribution of chemical structures sources

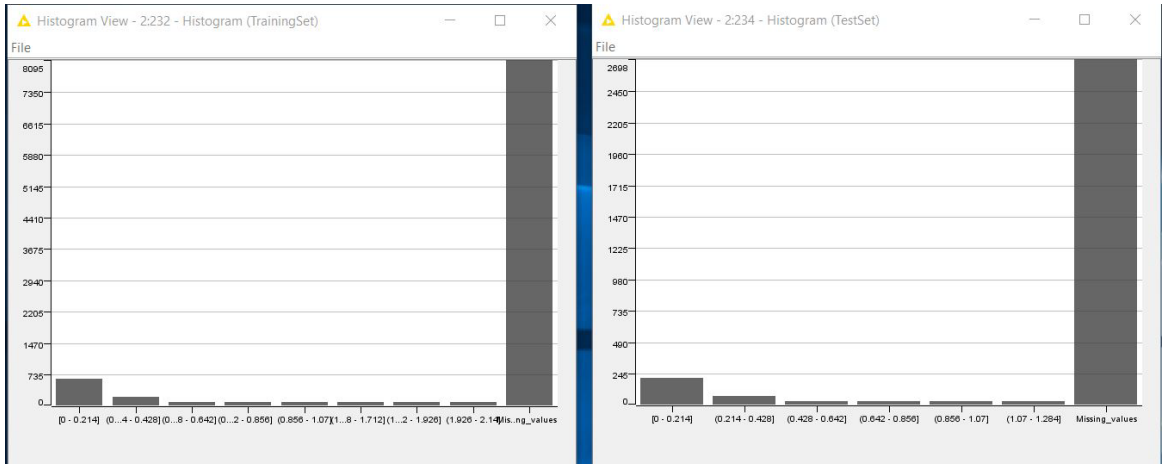


Training set: 9888 chemicals Evaluation set: 2888 chemicals

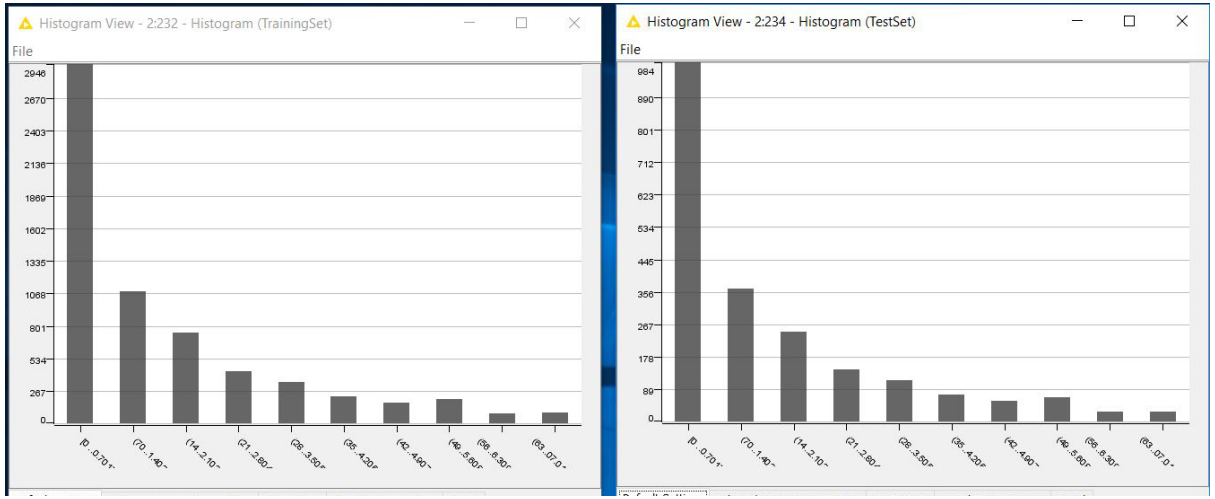
Similar distribution of values and variability (LD50)



Replicates distribution between training and test set

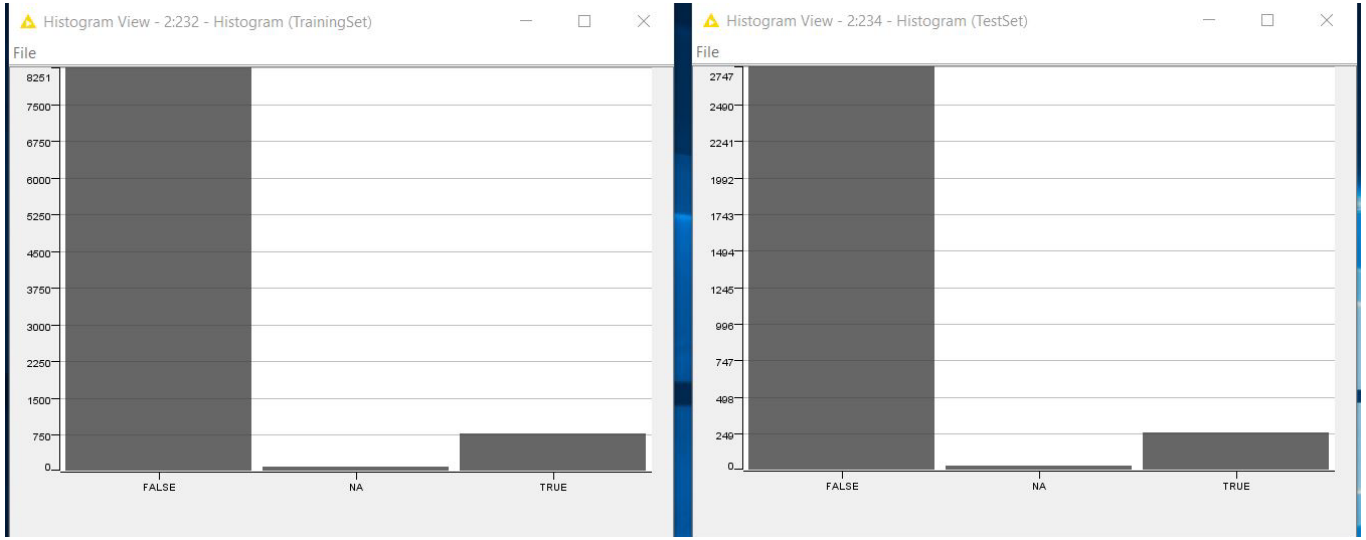


Stdev distribution between training and test set

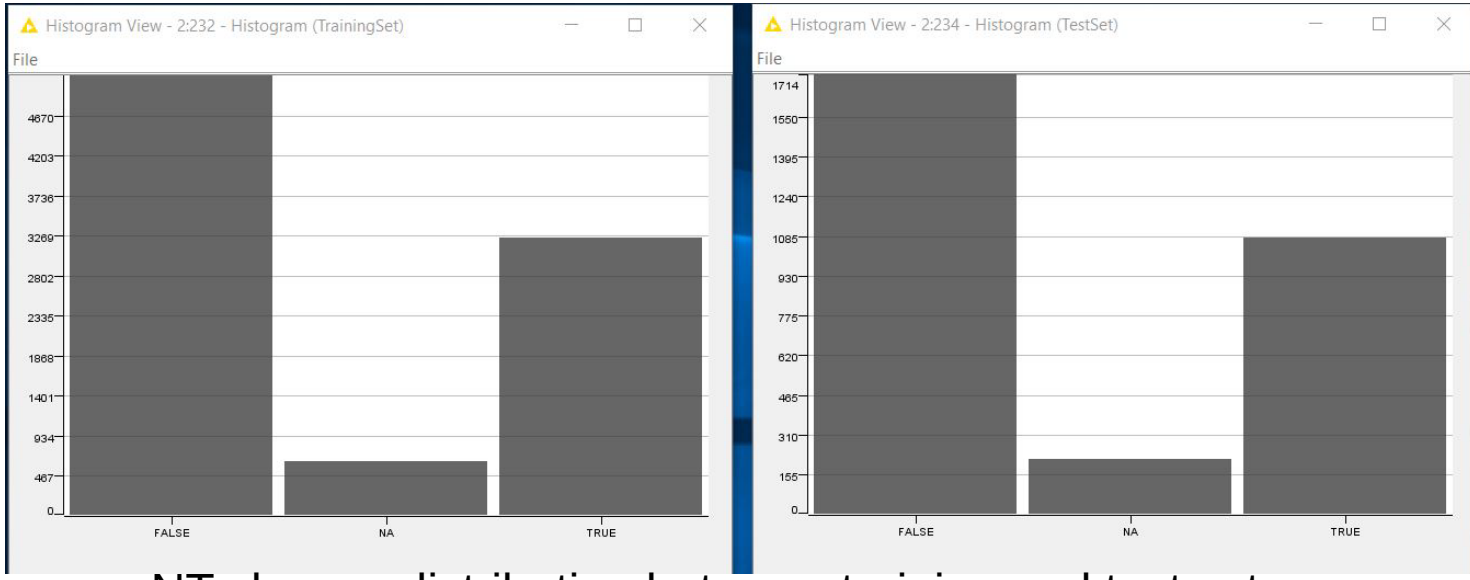


LD50 values distribution between training and test set

Similar distribution for true and false (NT, VT)

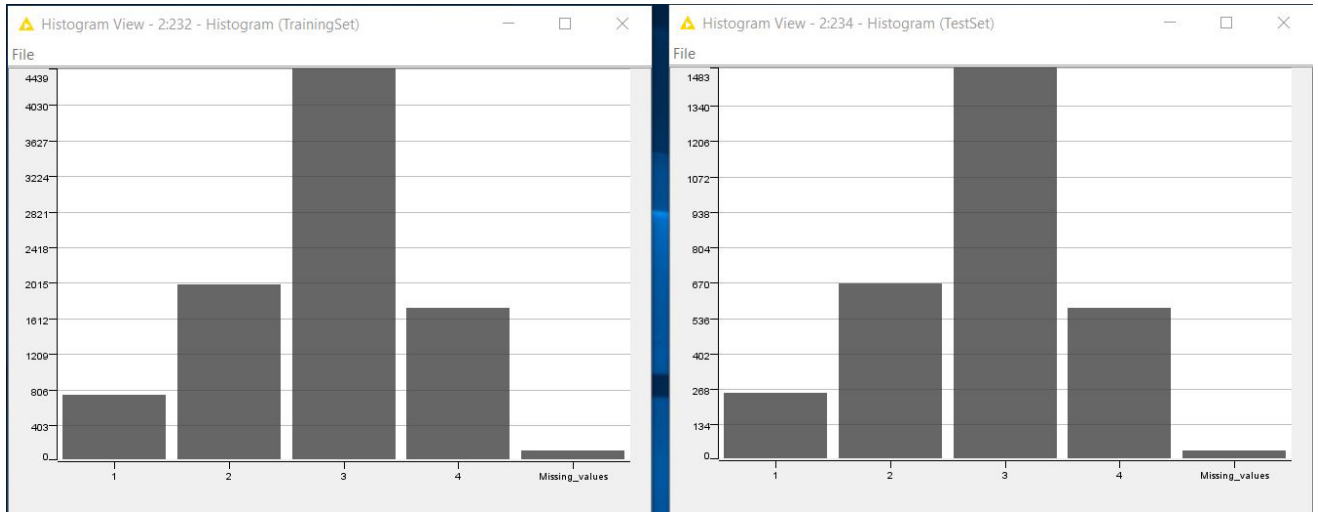


VT classes distribution between training and test set

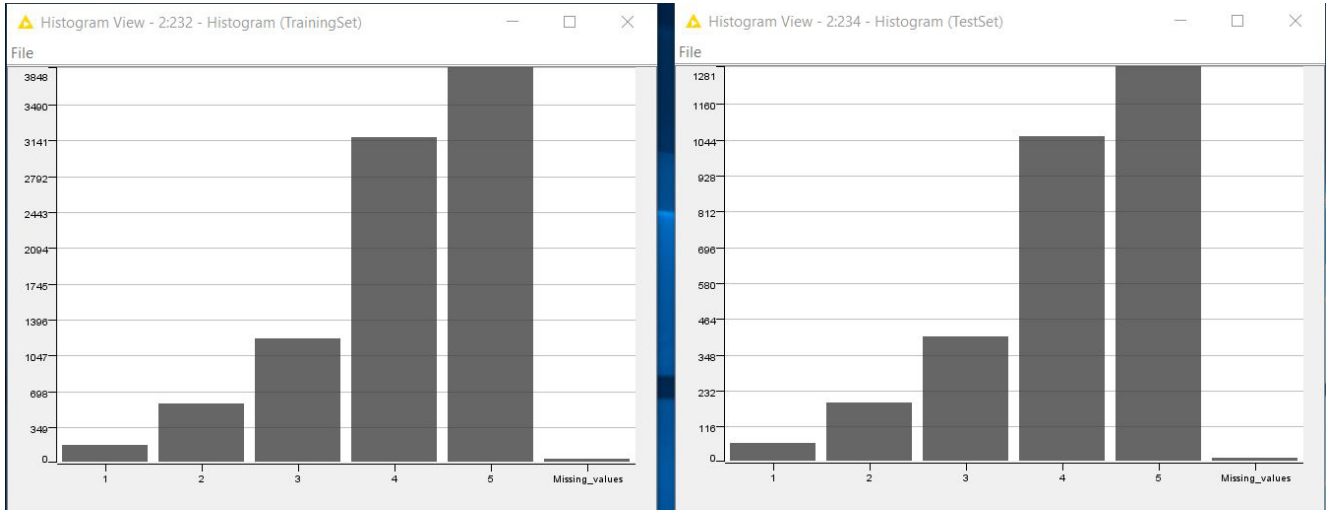


NT classes distribution between training and test set

Similar distribution of categories (EPA, GHS)



EPA categories distribution between training and test set



GHS categories distribution between training and test set

Prediction set

Lists:

- ToxCast/Tox21
- EDSP
- TSCA
- Substances on the market (EPA Dashboard list)

After QSAR-ready standardization:

48137 structures to be predicted

Modeling efforts, participants

Group ID	Group_short	Affiliation	Location
1	UNIMIB	Università degli Studi di Milano - Bicocca	Italy
2	USAFSAM	Henry M Jackson Foundation at USAFSAM	USA
3	UNIBARI	Università degli Studi di Bari	Italy
4	ECUST	East China University of Science and Technology	China
5	LSINC	Leadscope, Inc.	USA
6	UNISTRA	Universite de Strasbourg	France
7	USEPA_NRMRL	US EPA NRMRL	USA
8	IRCCS_1	IRCCS – Istituto di Ricerche Farmacologiche Mario Negri	Italy
9	IRCCS_2	IRCCS - Istituto di Ricerche Farmacologiche Mario Negri	Italy
10	IRCCS_3	IRCCS - Istituto di Ricerche Farmacologiche Mario Negri	Italy
11	IRCCS_4	IRCCS - Istituto di Ricerche Farmacologiche Mario Negri	Italy
12	UNICAMB	University of Cambridge	UK
13	IRCCS_5	IRCCS - Istituto di Ricerche Farmacologiche Mario Negri	Italy
14	NCSTATE	North Carolina State University	USA
15	COLPHA	Collaborations Pharmaceuticals, Inc.	USA
16	LOREAL	L'OREAL Research and Innovation	France
17	UNC	University of North Carolina	USA
18	PNNL	Pacific Northwest National Laboratory	USA
19	UL	Underwriters Laboratories	USA
20	RUTC	Rutgers University-Camden	USA
21	HZM	Helmholtz Zentrum München	Germany
22	SIMPLUS	Simulations Plus Inc	USA
23	NCATS	National Center for Advancing Translational Sciences (N	USA
24	KU	Kyoto University	Japan
25	FUG	Federal University of Goias	Brazil
26	RUT	Rutgers University	USA
27	DOW	Dow Chemical Company	USA
28	USEPA_NCCT	US EPA NCCT	USA
29	MSU	Michigan State University	USA
30	DOW_AGRO	Dow Agrosciences	USA
31	ROSETTAC	Rosettastein Consulting	Germany
32	DUT	Dalian University of Technology	China

Previous collaborations:

CERAPP

Collaborative Estrogen Receptor Activity Prediction Project

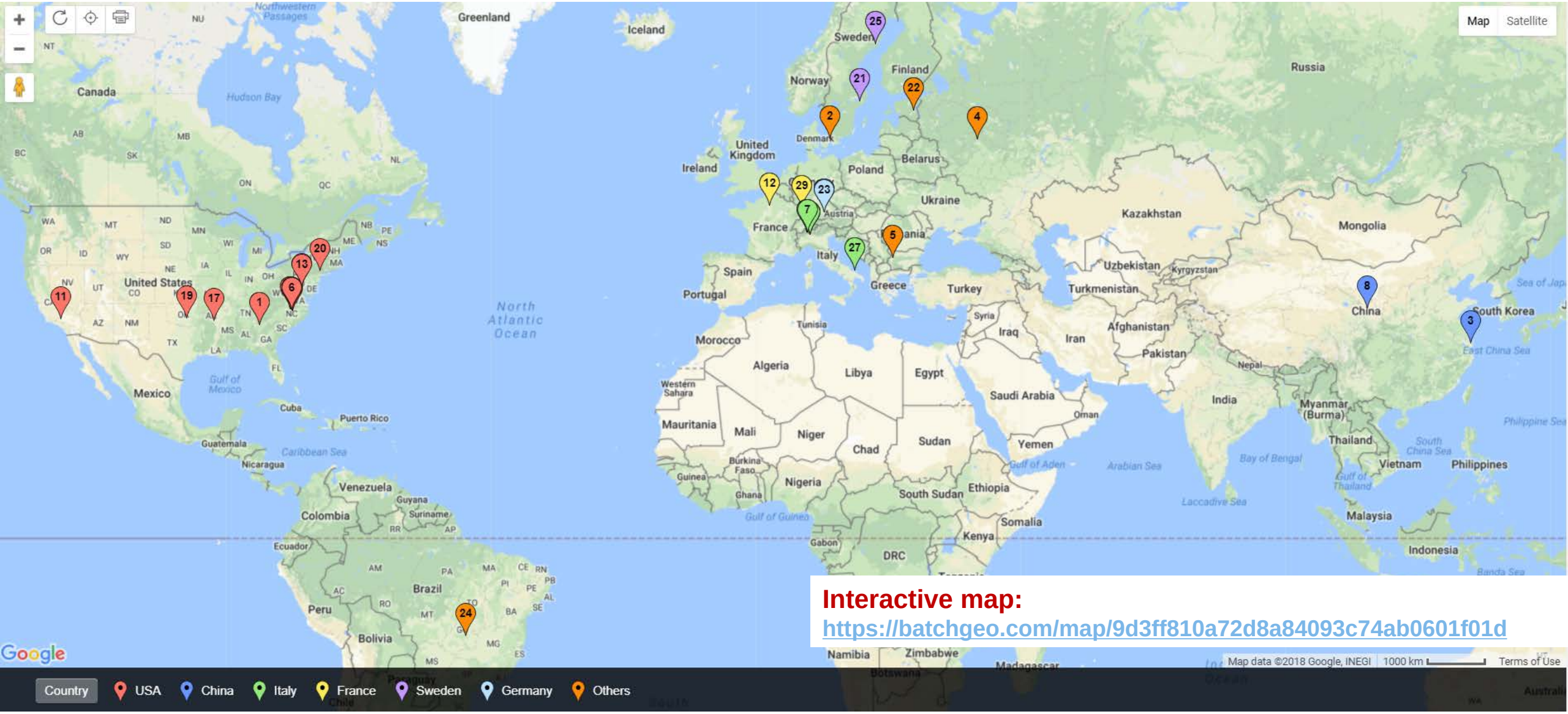
Mansouri et al. (http://ehp.niehs.nih.gov/15-10267/)

CoMPARA

Collaborative **M**odeling **P**roject for **A**ndrogen **R**eceptor **A**ctivity

Mansouri et al. (DOI: 10.13140/RG.2.2.19612.80009)

Participant groups locations

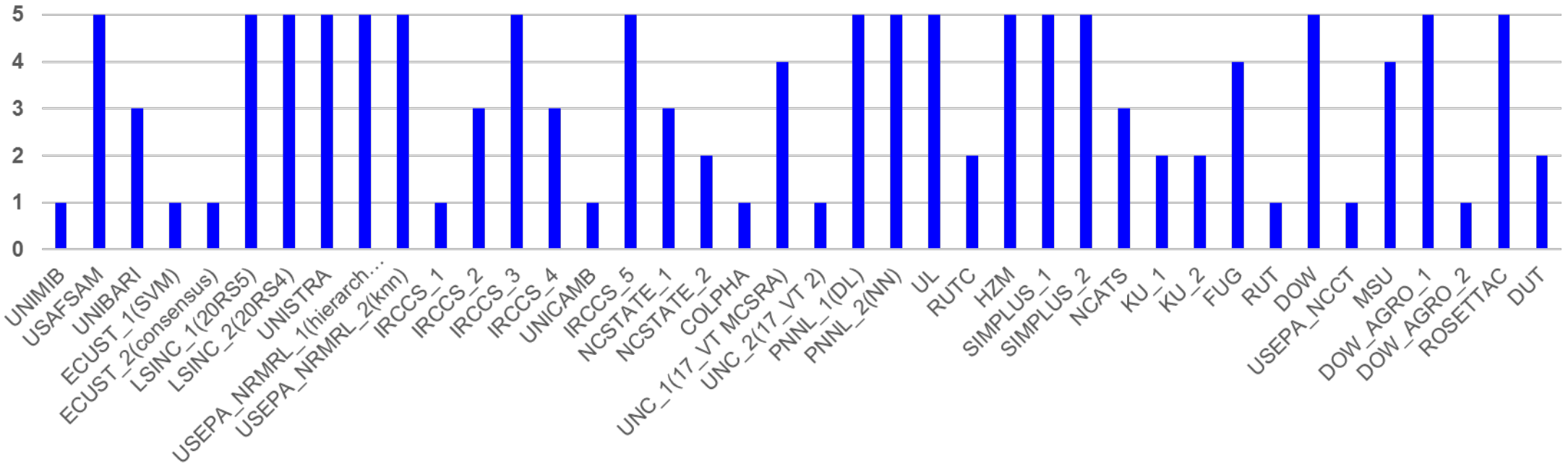


Interactive map:
https://batchgeo.com/map/9d3ff810a72d8a84093c74ab0601f01d

Received models

- Very Toxic: 31 models
- Non-toxic: 32 models
- EPA categories: 24 models
- GHS categories: 21 models
- LD50: 24 models

Total: 132 models



Evaluation procedure:

Qualitative evaluation:

- Documentation
- Defined endpoint
- Unambiguous algorithm
- Availability of code
- Applicability domain definition
- Availability of data used for modeling
- Mechanistic interpretation

Quantitative evaluation:

- Goodness of fit: training statistics
- Evaluation set predictivity: statistics on the evaluation set
- Robustness: balance between (Goodness of fit) & (Test set predictivity)

$$S = 0.3 * (Goodness\ of\ fit) + 0.45 * (Test\ set\ predictivity) + 0.25 * (Robustness)$$

Categorical models (binary and multi-class):

$$Goodness\ of\ fit = 0.7 * (BA_{Tr}) + 0.3 * \left(1 - |\widetilde{Sn_{Tr} - Sp_{Tr}}|\right)$$

$$Test\ set\ predictivity = 0.7 * (BA_{Tst}) + 0.3 * \left(1 - |\widetilde{Sn_{Tst} - Sp_{Tst}}|\right)$$

$$Robustness = 1 - |BA_{Tr} - BA_{Tst}|$$

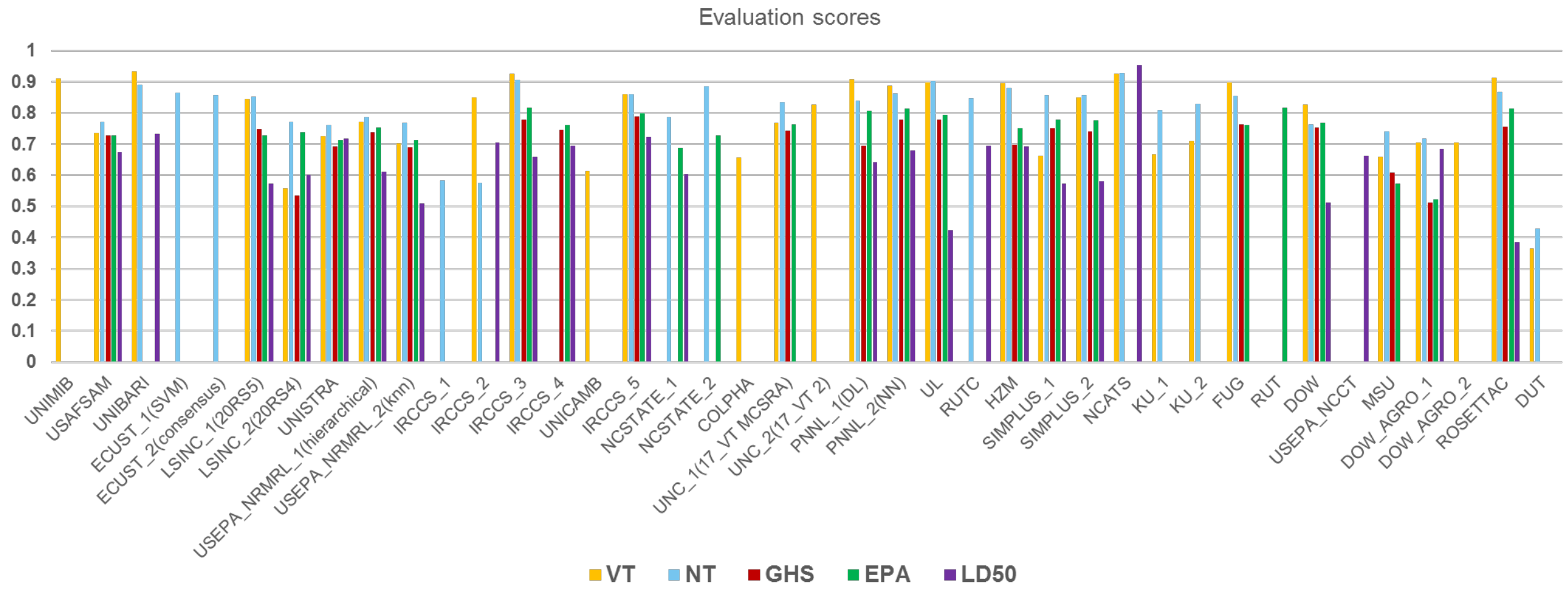
Continuous models:

$$Goodness\ of\ fit = R^2_{Tr}$$

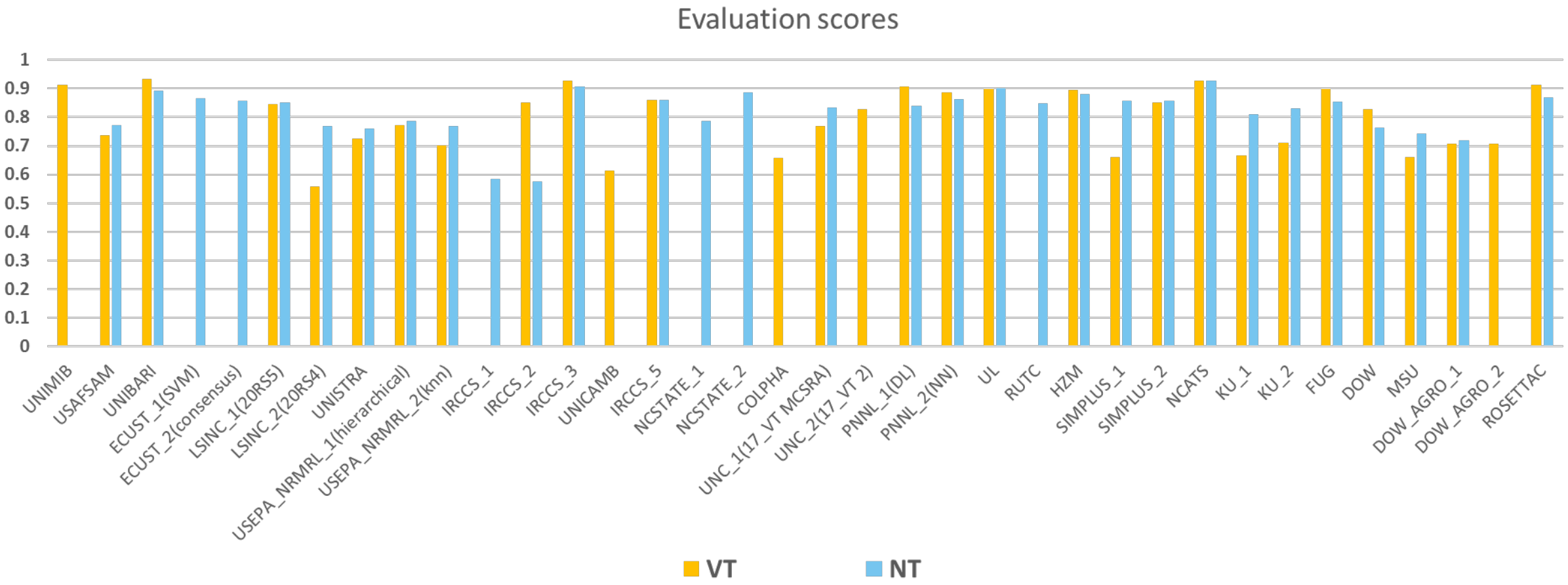
$$Test\ set\ predictivity = R^2_{Tst}$$

$$Robustness = 1 - |R^2_{Tr} - R^2_{Tst}|$$

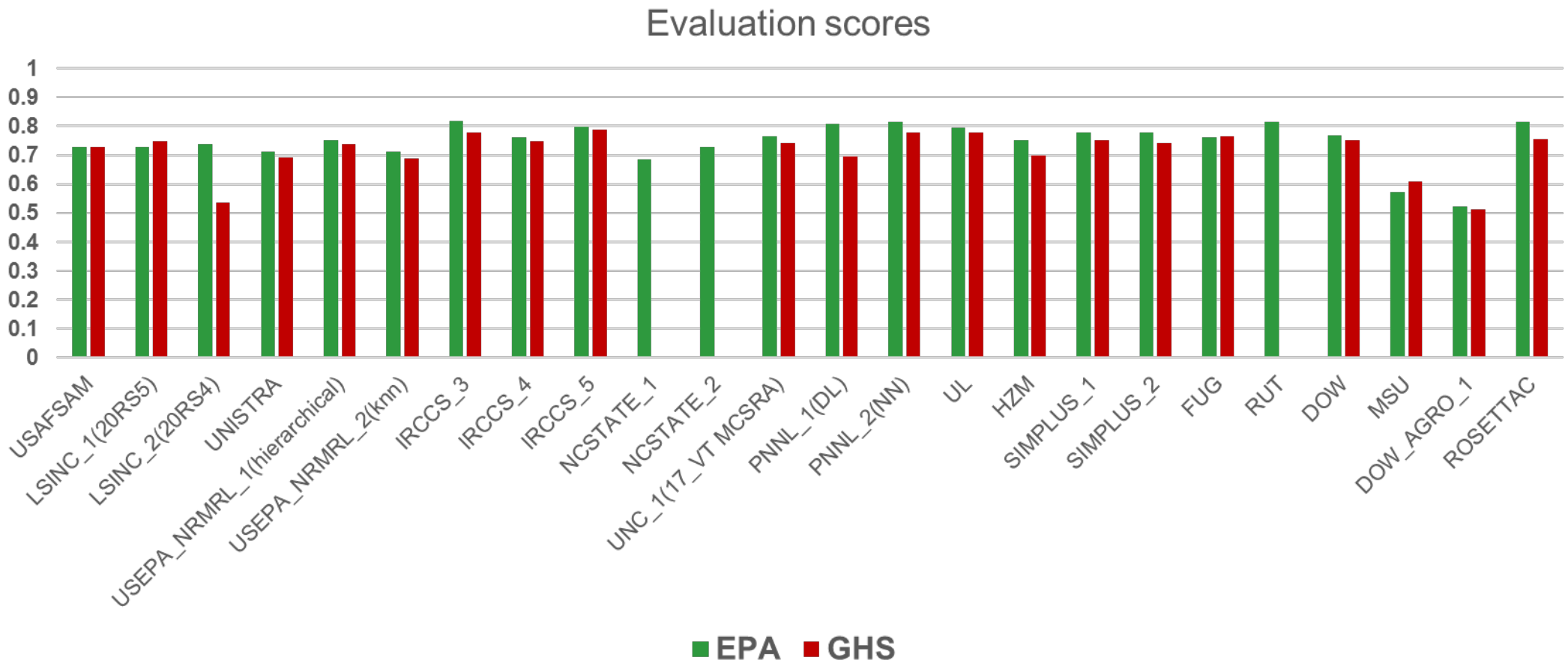
Evaluation results



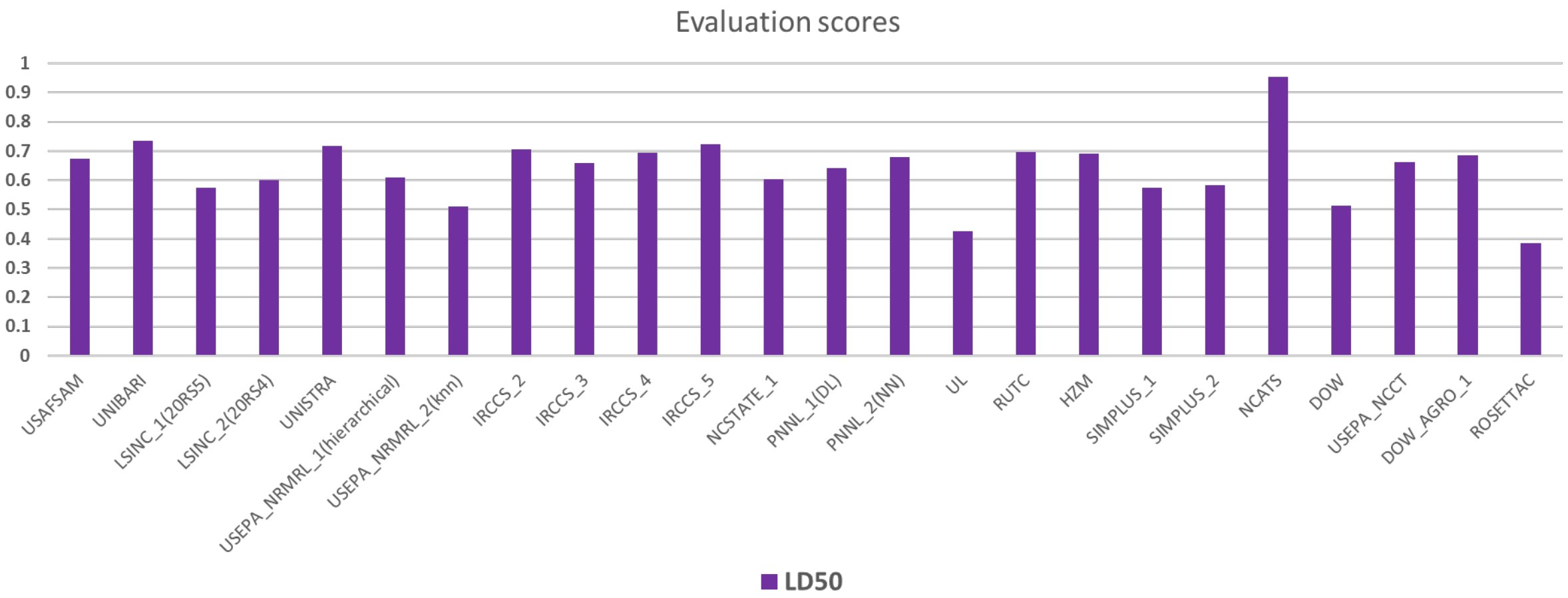
Evaluation of the VT and NT models



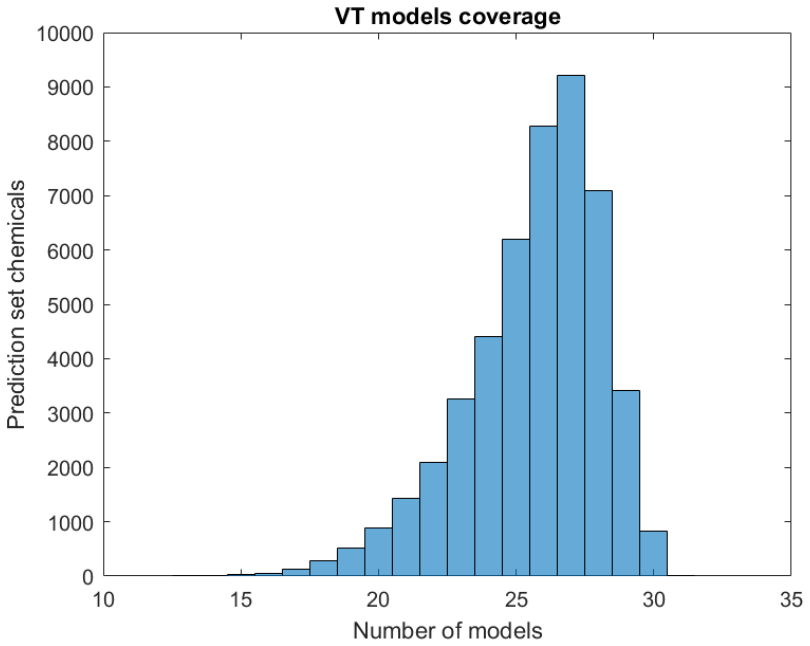
Evaluation of the EPA and GHS models

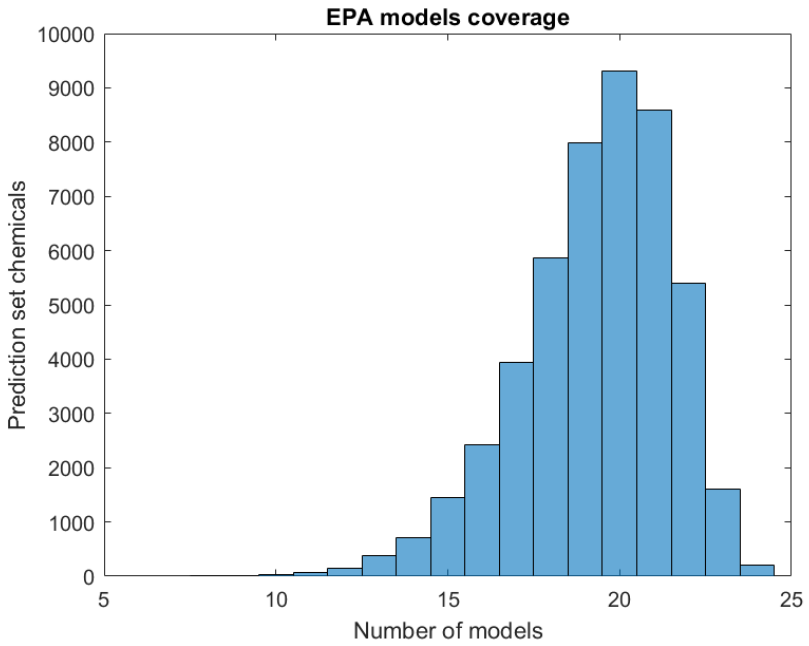


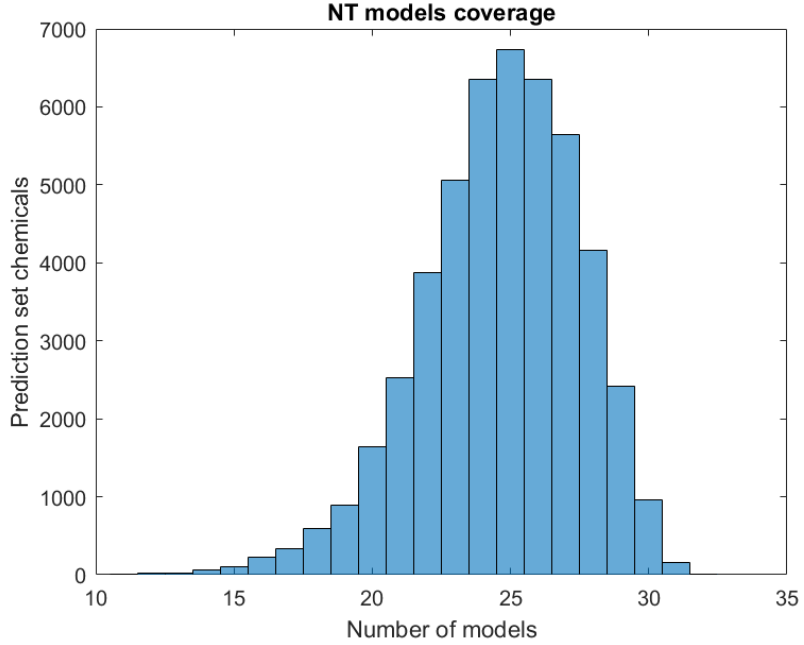
Evaluation of the LD50 models

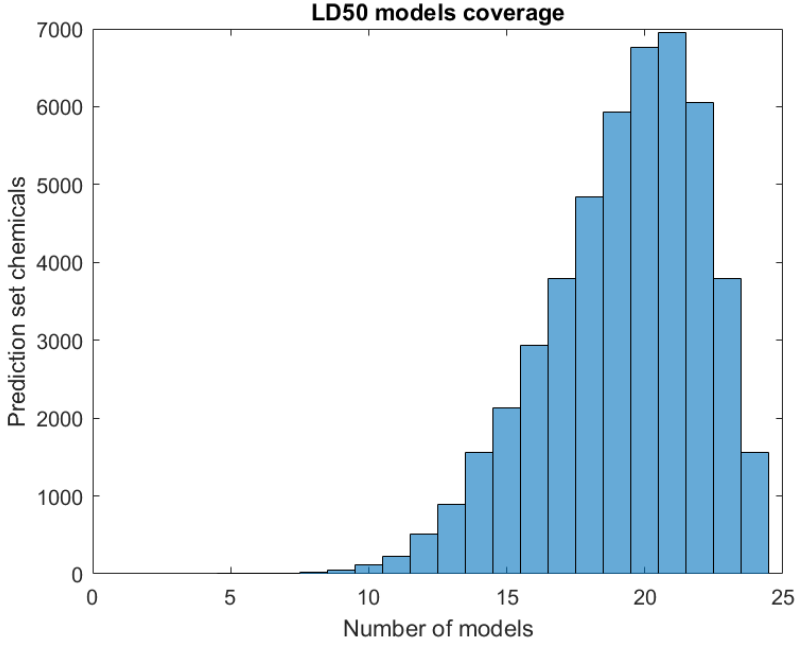


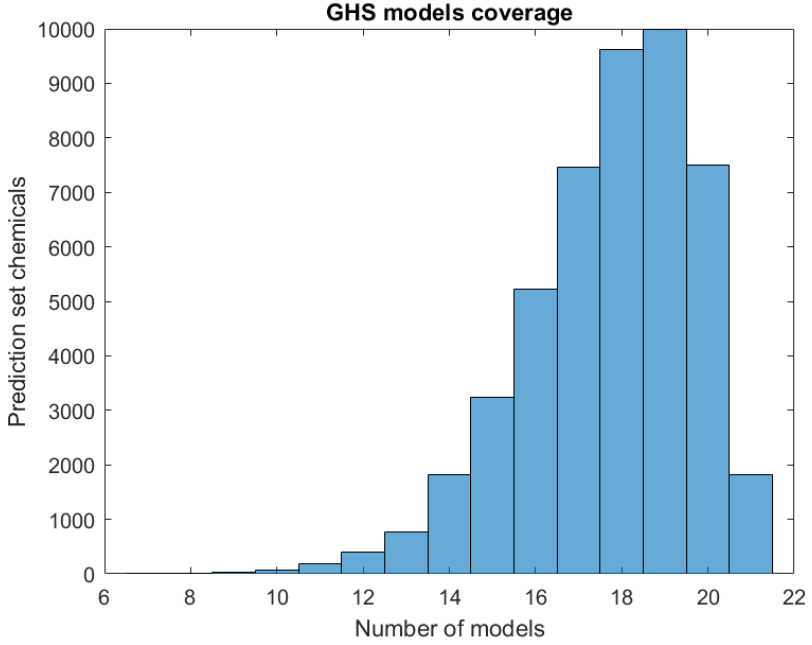
Coverage of the models











Consensus

- **Categorical models:**

Weighted majority rule

- **Continuous models:**

Weighted average

The predicted consensus value (*C*) of the chemical *i* is calculated as:

$$C_i = \sum_{j=1}^{n} w_j \cdot P_j$$

where *n* is the number of models that provided predictions for the chemical i, and P_j is the prediction of each one of them. The weight (*w*) of each model *j* is calculated as:

$$w_j = S_j \Big/ \sum_{k=1}^{n} S_k$$

So that the sum of weights is equal to 1.

For each chemical of the prediction set, the weights and consensus are calculated based on predictions within the AD only.

Consensus results: Binary and LD50

	VT Train	VT Test	NT Train	NT Test
Sn	0.87	0.67	0.93	0.70
Sp	0.94	0.96	0.96	0.88
BA	0.93	0.81	0.94	0.79

	LD50 Train	LD50 Test
R2	0.84	0.64
RMSE	0.32	0.51

The balanced accuracy of the replicate animal data for predicting VT and NT categories was 81% and 89%, respectively.

The reproducibility of the replicate animal data for predicting LD50 had R2 of 0.8 and RMSE of 0.42.

Consensus results: EPA and GHS

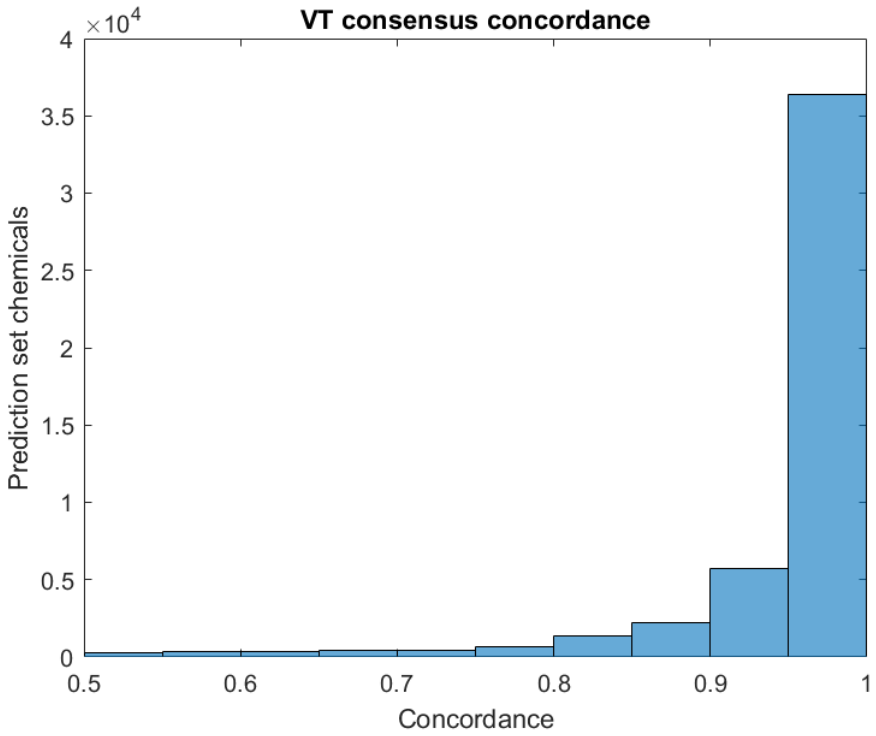
	EPA Train	EPA Test	GHS Train	GHS Test
Median Sn	0.73	0.5	0.63	0.45
Median Sp	0.96	0.91	0.91	0.92
BA	0.83	0.71	0.77	0.68

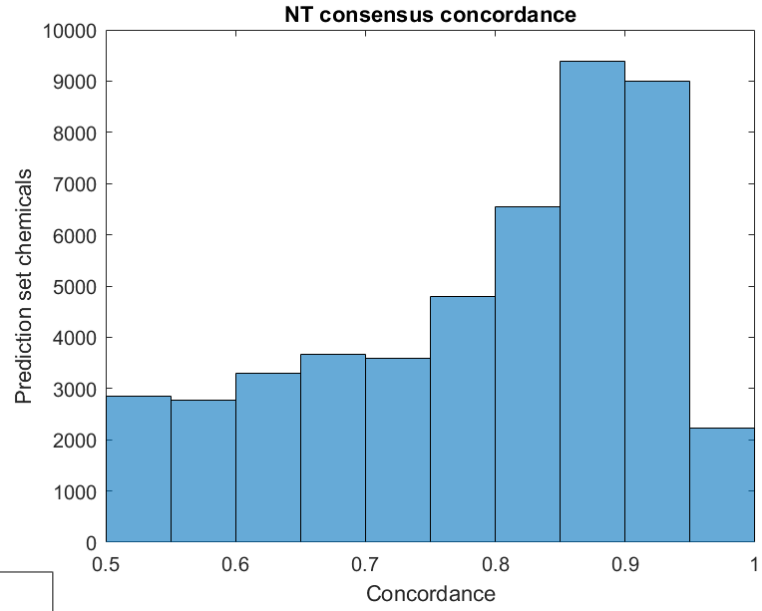
	EPA Train Cat 1	EPA Train Cat 2	EPA Train Cat 3	EPA Train Cat 4	EPA Test Cat 1	EPA Test Cat 2	EPA Test Cat 3	EPA Test Cat 4
Sn	0.55	0.83	0.92	0.65	0.45	0.54	0.80	0.38
Sp	1	0.94	0.75	0.98	0.98	0.86	0.59	0.96

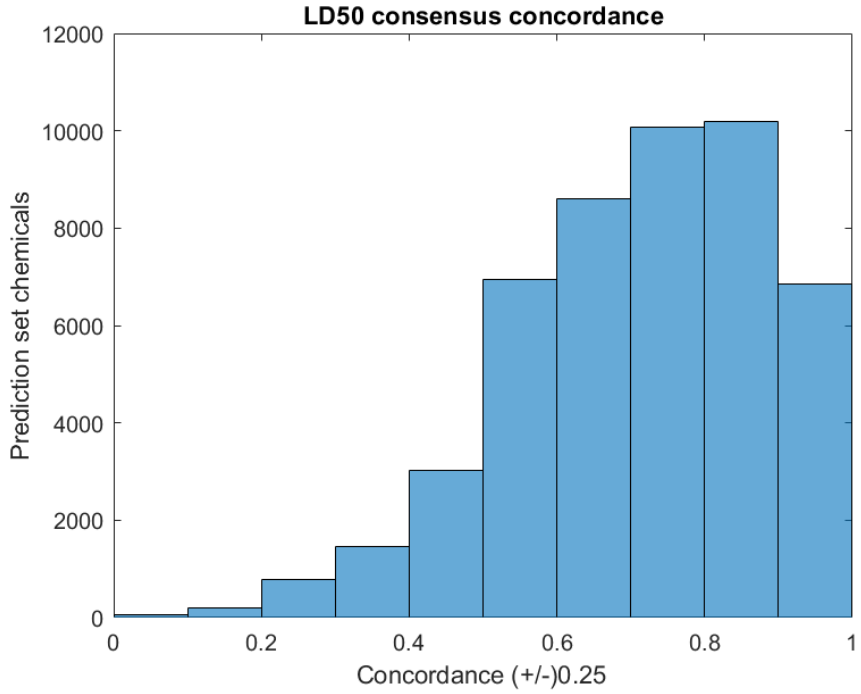
	GHS Train Cat 1	GHS Train Cat 2	GHS Train Cat 3	GHS Train Cat 4	GHS Train Cat 5	GHS Test Cat 1	GHS Test Cat 2	GHS Test Cat 3	GHS Test Cat 4	GHS Test Cat 5
Sn	0.34	0.48	0.63	0.91	0.69	0.18	0.43	0.44	0.76	0.53
Sp	1	1	0.95	0.71	0.98	1	0.96	0.91	0.61	0.92

The accuracy of the animal data for predicting EPA and GHS categories was 78% and 74%, respectively.

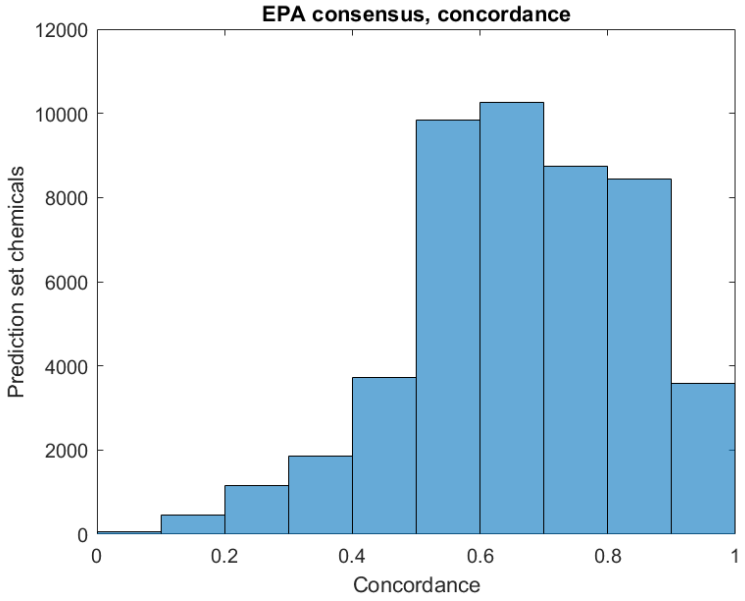
Model concordance

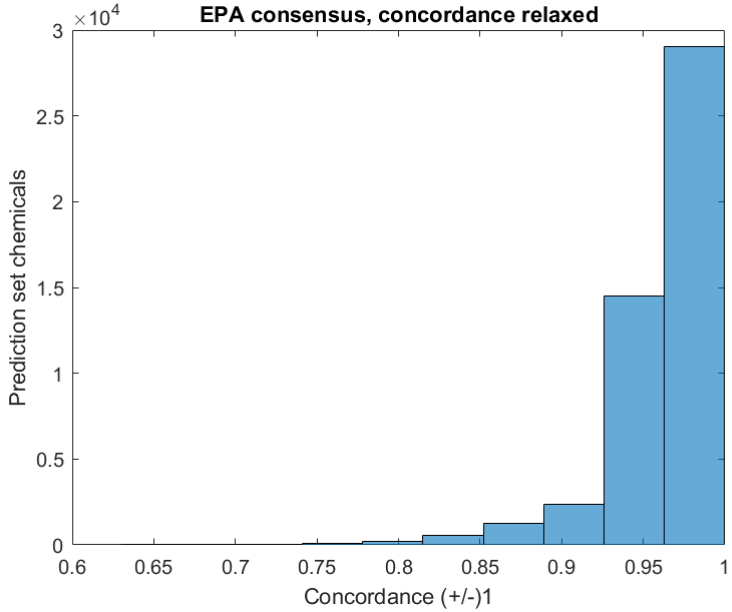




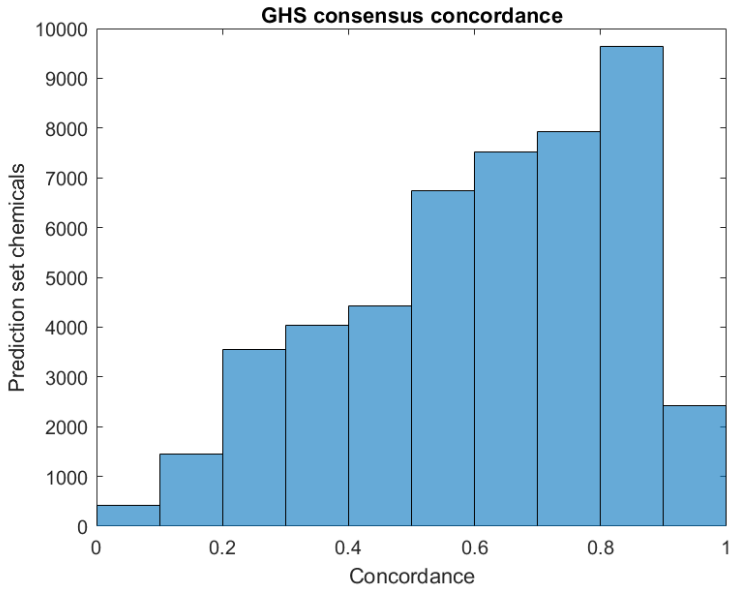


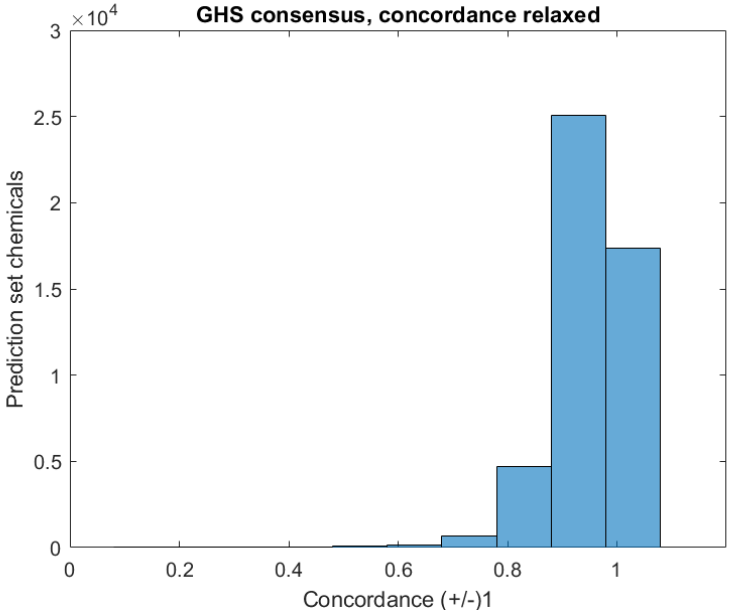
Model concordance





Concordance relaxed:
fraction of models predicting the consensus category +/- 1 category.

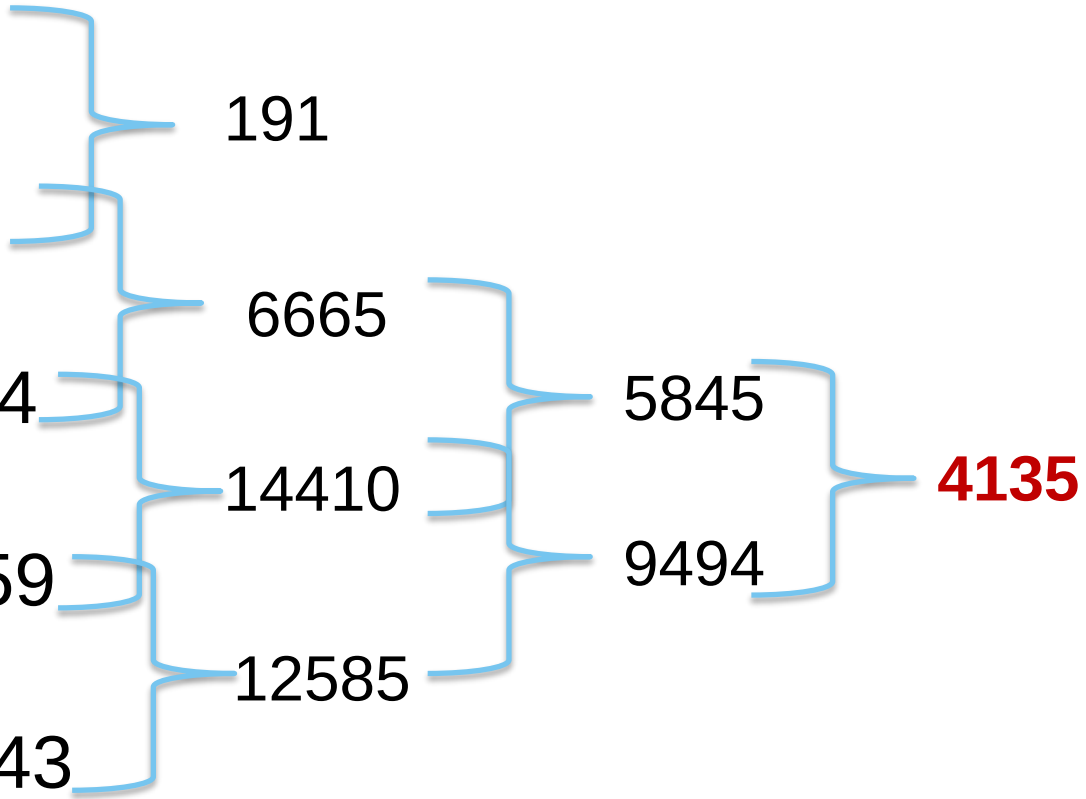




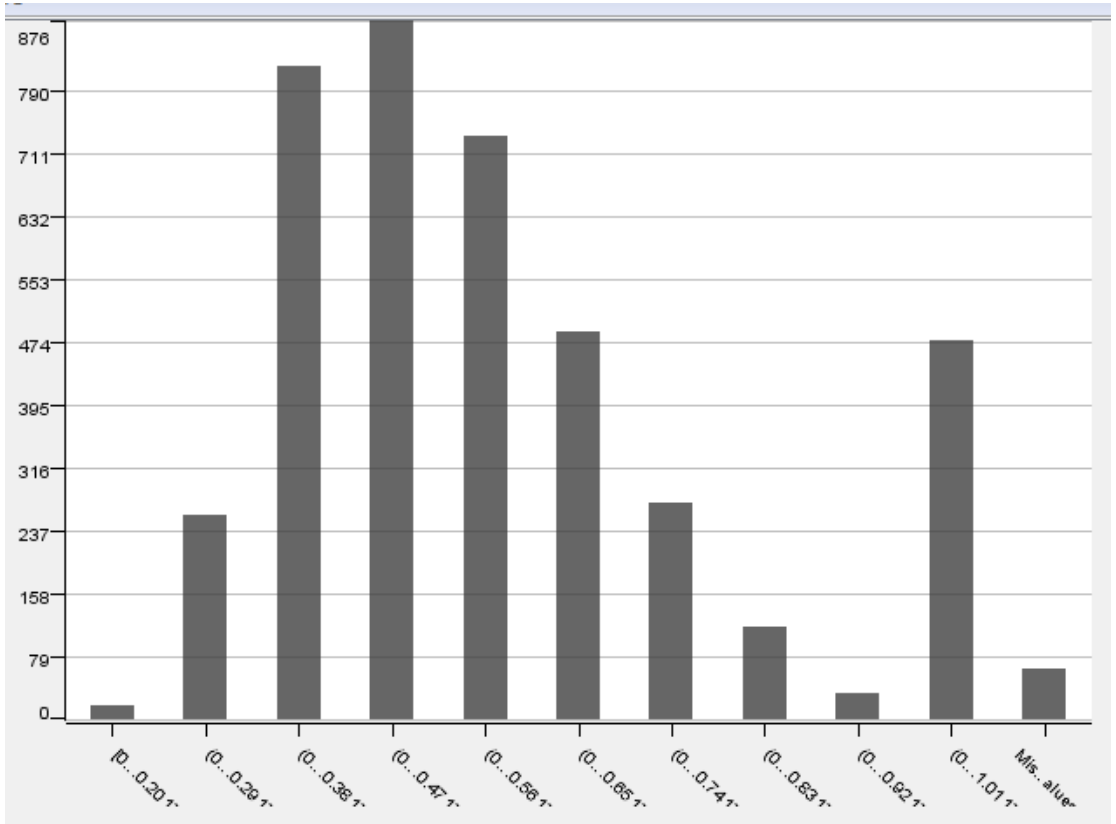
Discordance analysis

<70% concordance

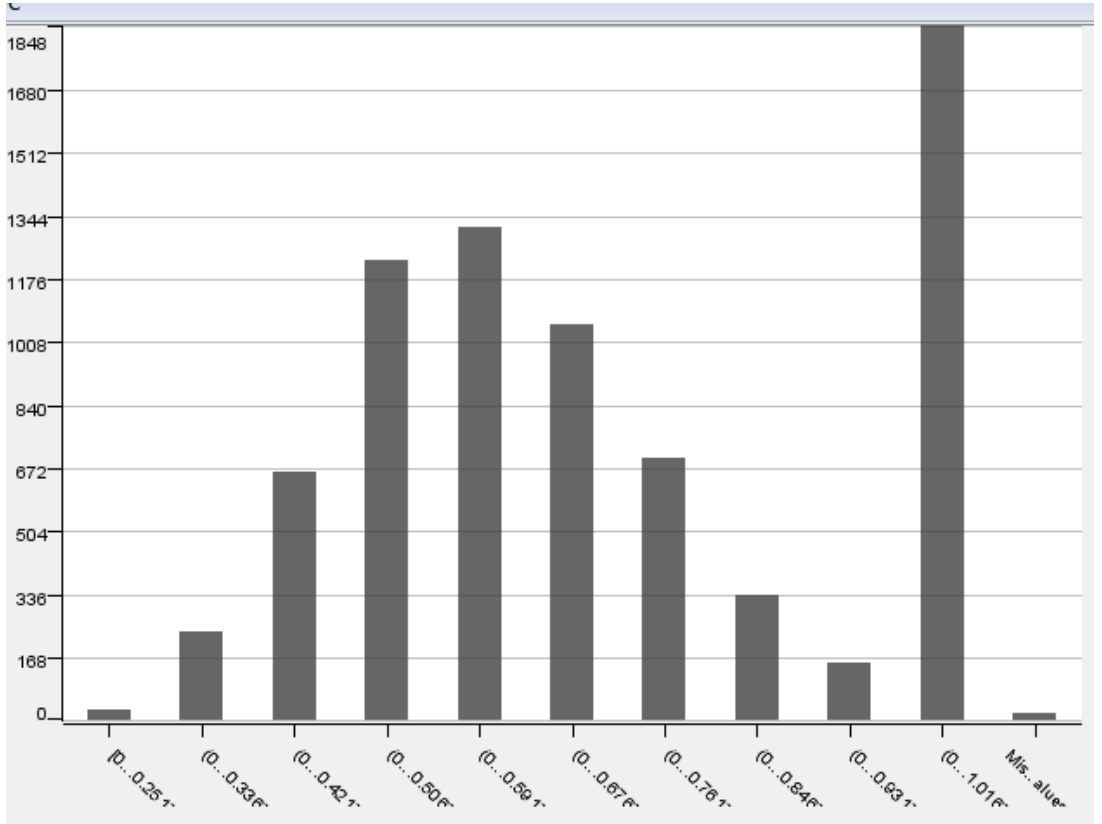
- VT: 1374
- NT: 12778
- EPA: 27364
- GHS: 24659
- LD50: 21043



Structural similarity to the training set

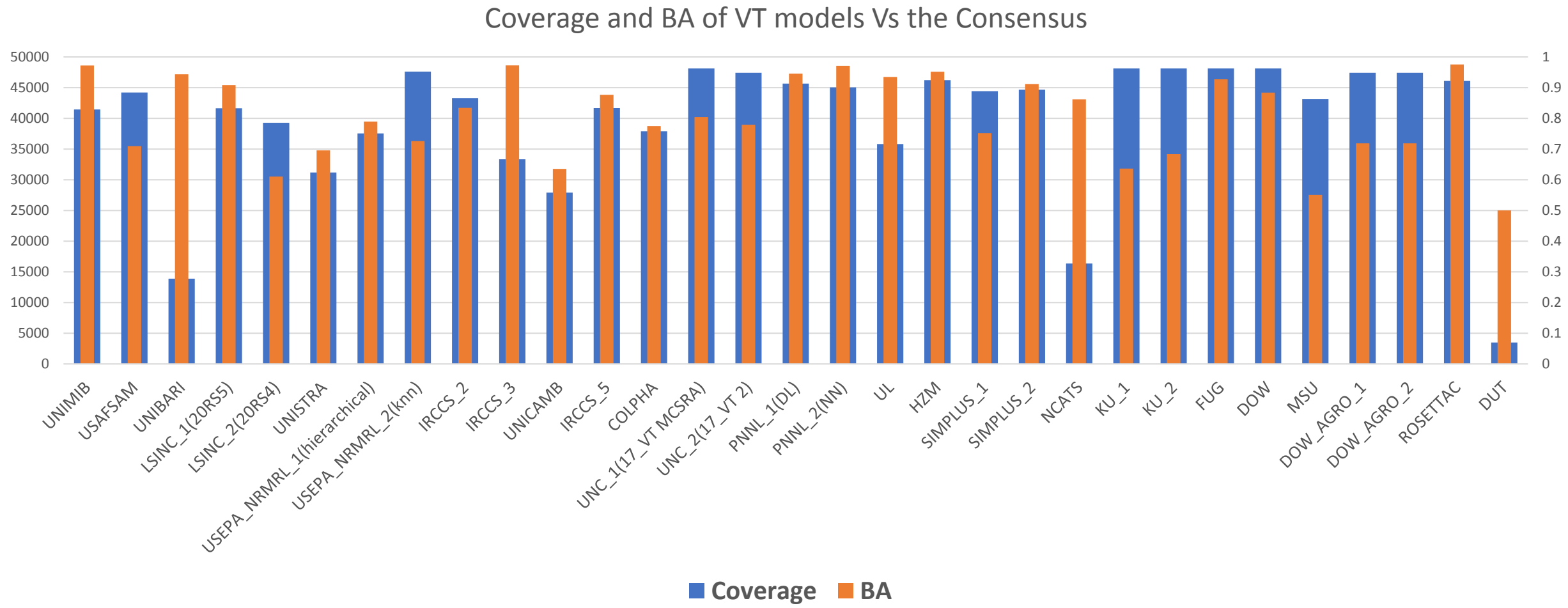


Most disconcordant (<0.7)
4135 chemicals



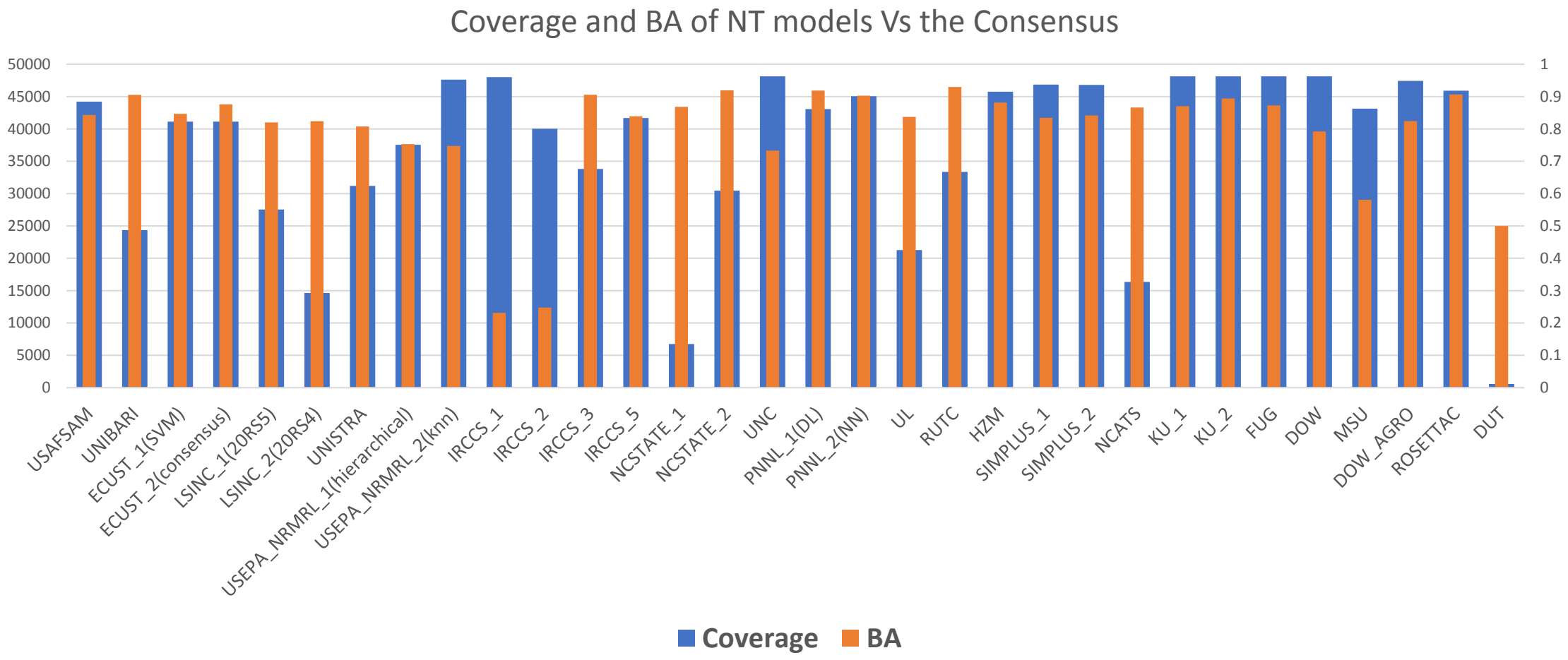
Most concordant (>0.7)
7525 chemicals

Models to consensus evaluation



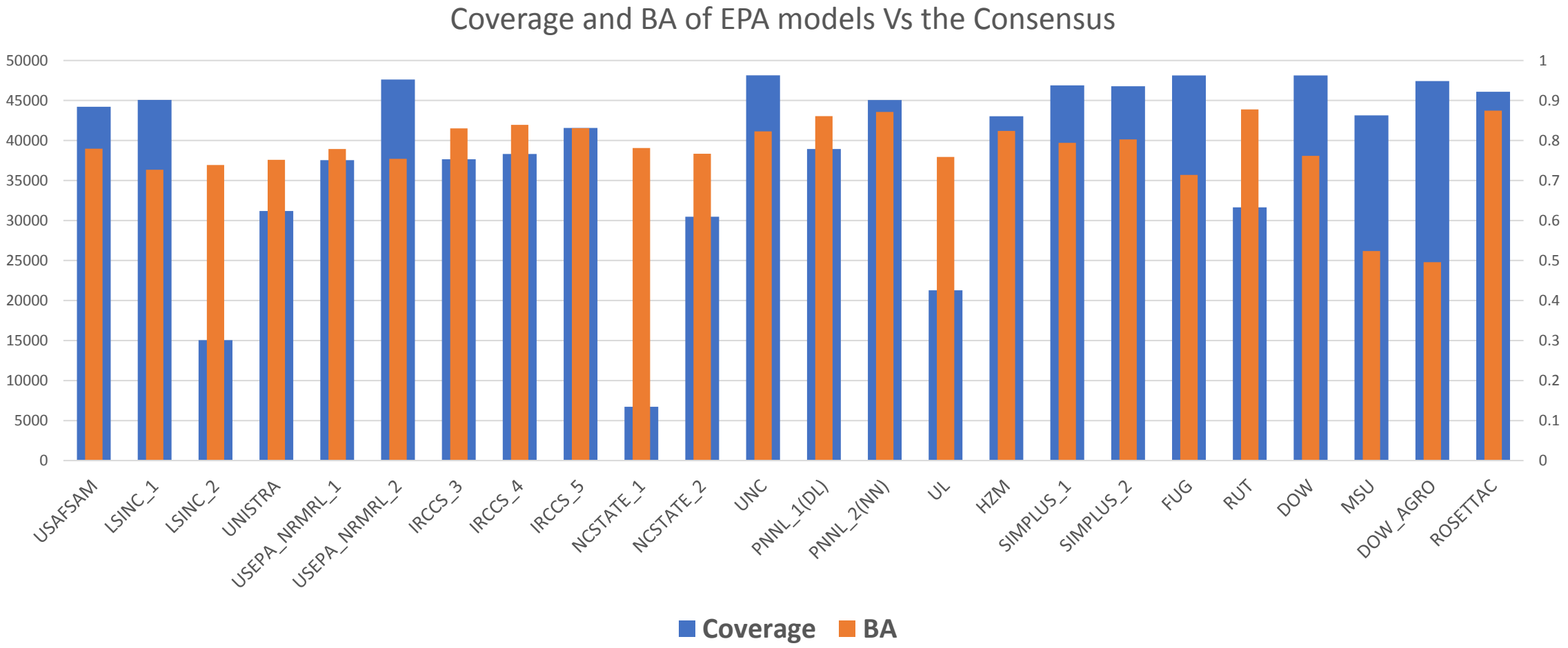
Coverage: number of predicted chemicals within the AD

Models to consensus evaluation



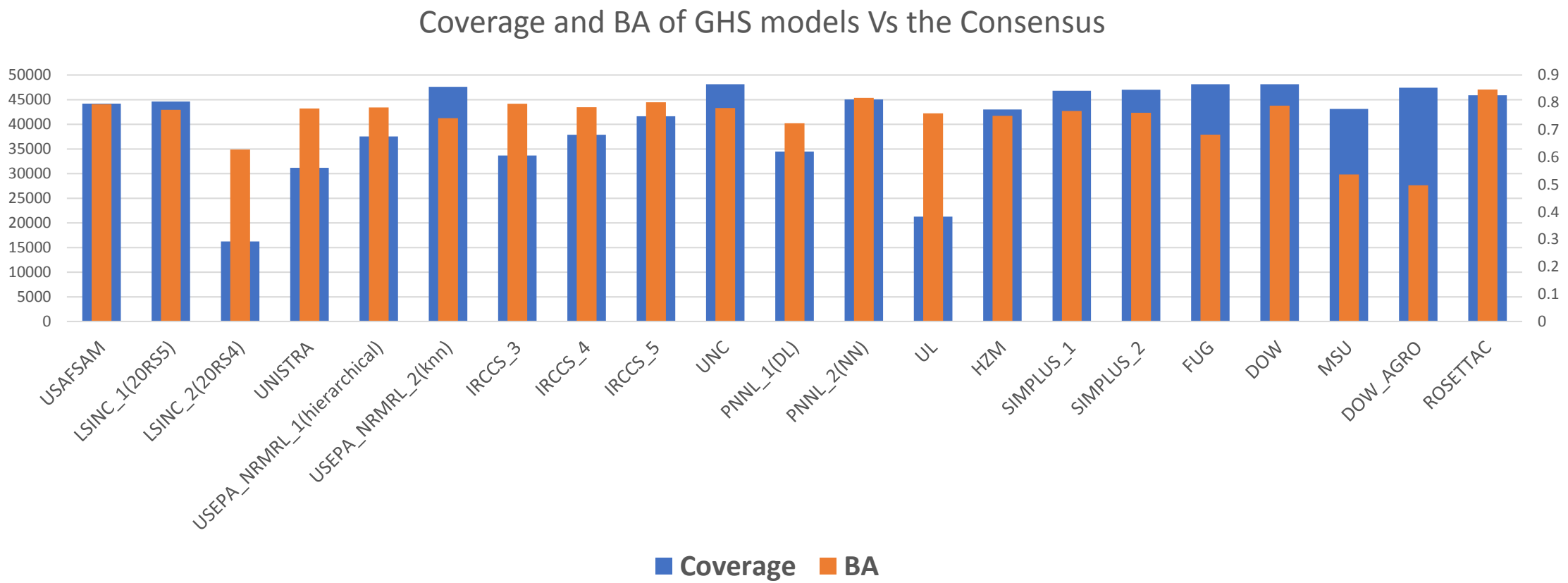
Coverage: number of predicted chemicals within the AD

Models to consensus evaluation



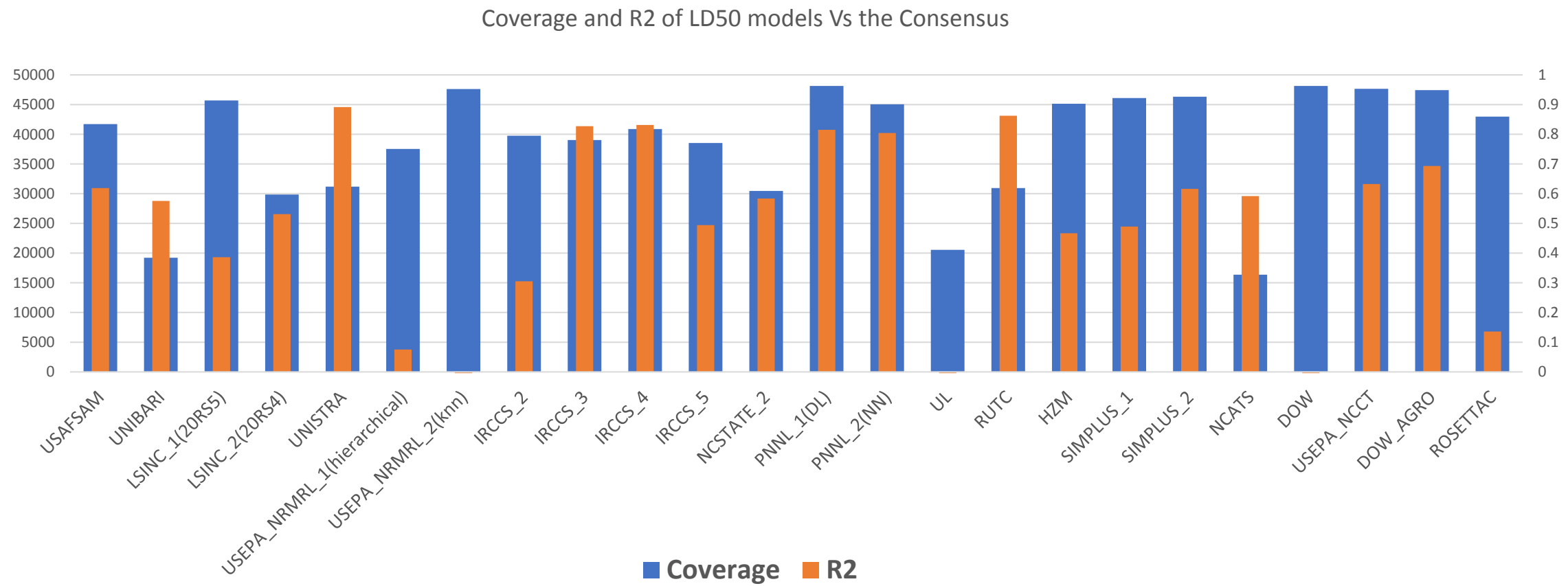
Coverage: number of predicted chemicals within the AD

Models to consensus evaluation



Coverage: number of predicted chemicals within the AD

Models to consensus evaluation



Coverage: number of predicted chemicals within the AD

Consensus implementation

>=85% concordance

- VT: 44523
- NT: 21659

>=75% concordance

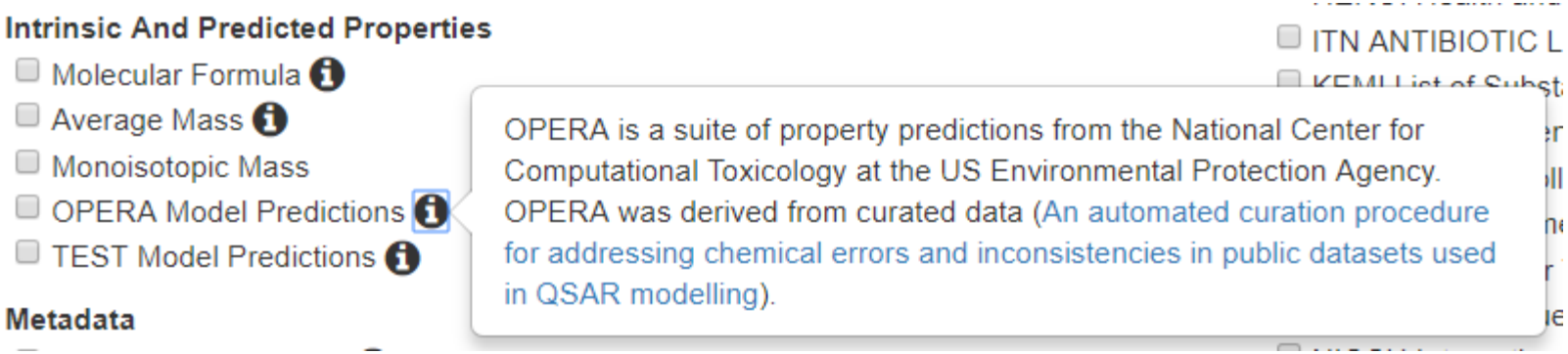
- EPA: 16959
- GHS: 20215
- LD50: 22738

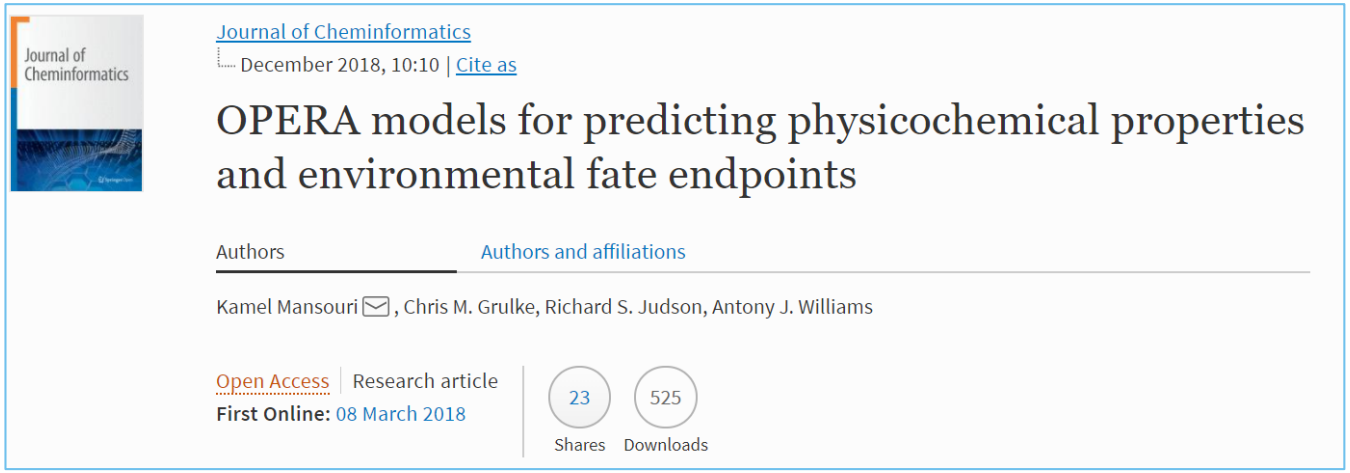
Implementation for regulatory use:

- A defined endpoint
- An unambiguous algorithm
- A defined domain of applicability
- Appropriate measures of goodness-of-fit, robustness and predictivity
- Mechanistic interpretation, if possible

OPERA and the EPA CompTox dashboard

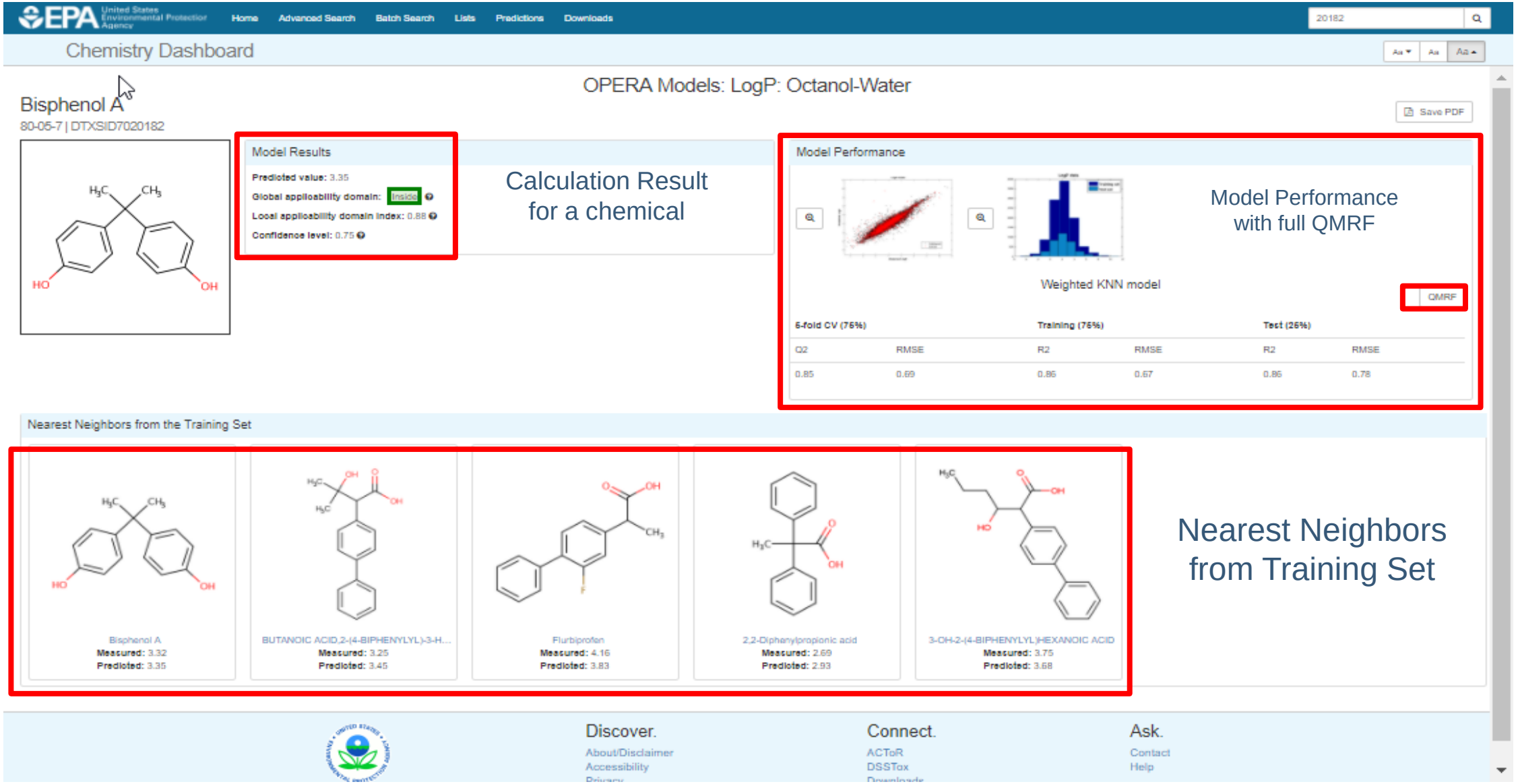
OPERA
OPEn (q)saR App





Mansouri et al. OPERA models (https://link.springer.com/article/10.1186/s13321-018-0263-1)

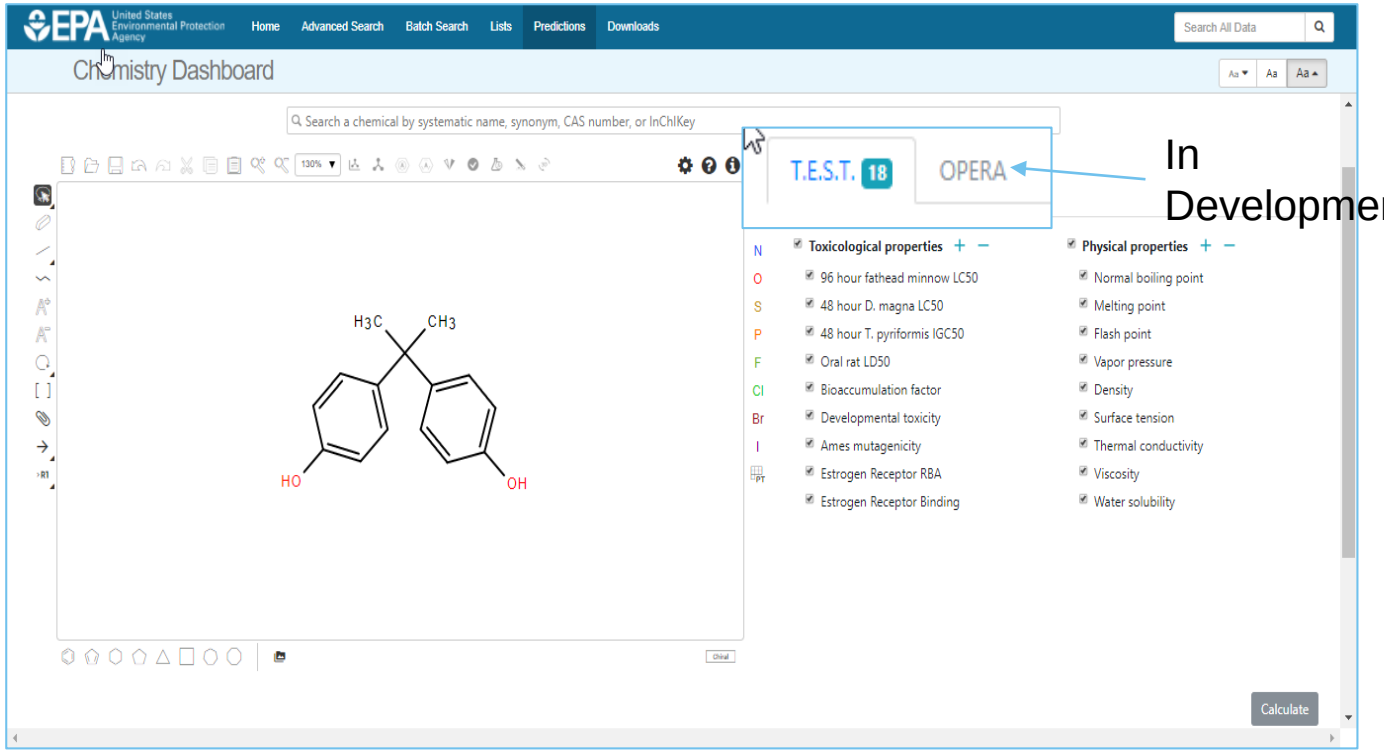
OPERA prediction report



Desktop and online Predictions:

https://github.com/kmansouri/OPERA

Standalone app:
batch mode for new chemicals



EPA Comptox dashboard:
batch mode download or drawing

Summary

- Generated high quality data and models that can be used prospectively to screen the chemical universe
- Screened tens of thousands of chemicals in a fast accurate and economic way.
- Free & open-source code and workflows
- Consensus models being implemented for future use to help with regulatory process
- Data and predictions will be soon available via the EPA's CompTox dashboard and the NICEATM Integrated Chemical Environment

Acknowledgments

- All collaborating groups

- EPA/NCCT
  - Grace Patlewicz
  - Jeremy Fitzpatrick
  - Prachi Pradeep

- ILS/NICEATM
  - Dave Allen
  - Shannon Bell
  - Agnes Karmaus
  - Patricia Ceger
  - Judy Strickland

- NTP/NICEATM
  - Nicole Kleinstreuer
  - Warren Casey