Using EPA's ChemExpo Tool to Build Updated Quantitative Structure-Use Relationship (QSUR) Models

Katherine Phillips EPA/ORD/CCTE
Victoria Hull ORAU SSC
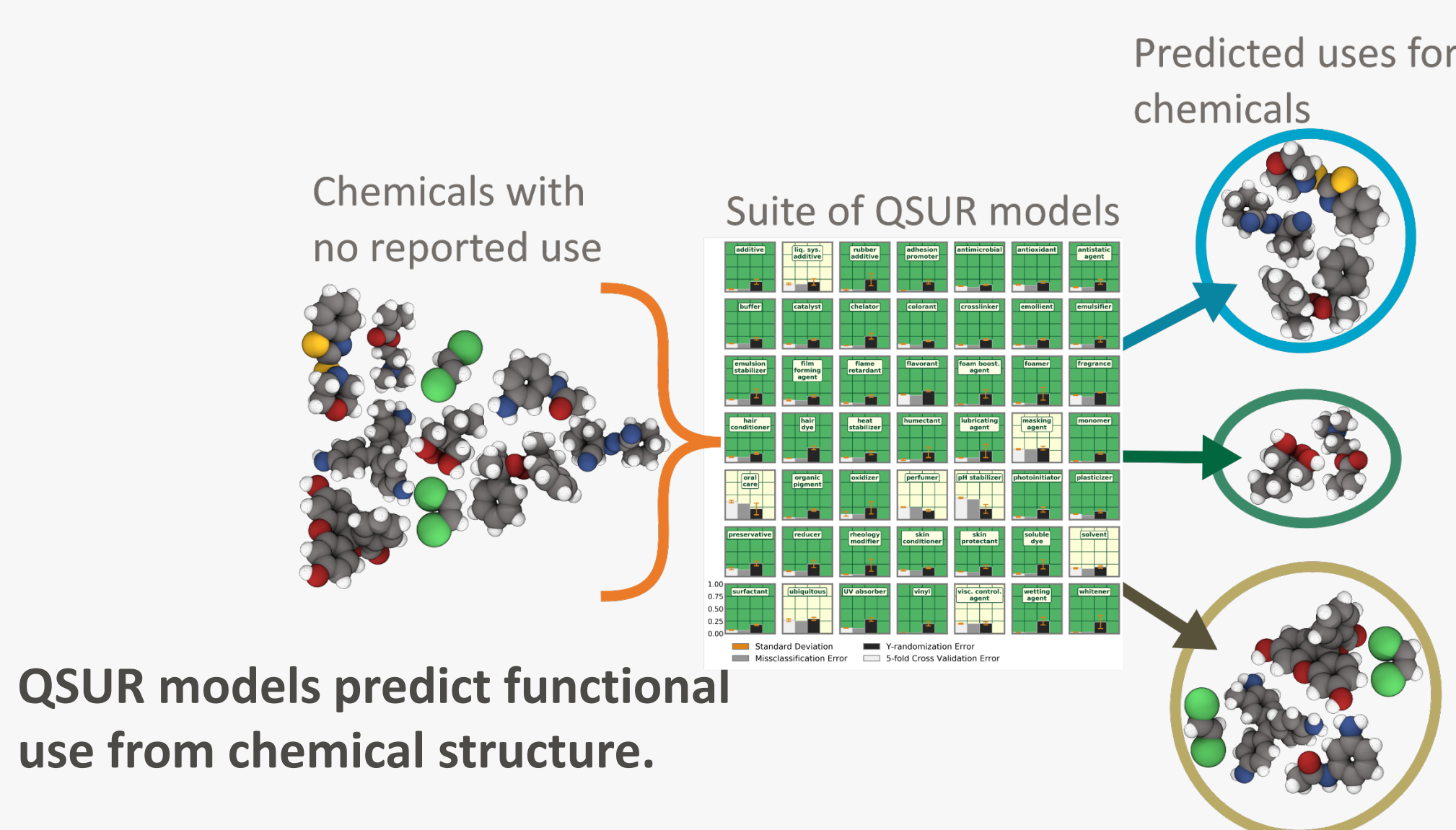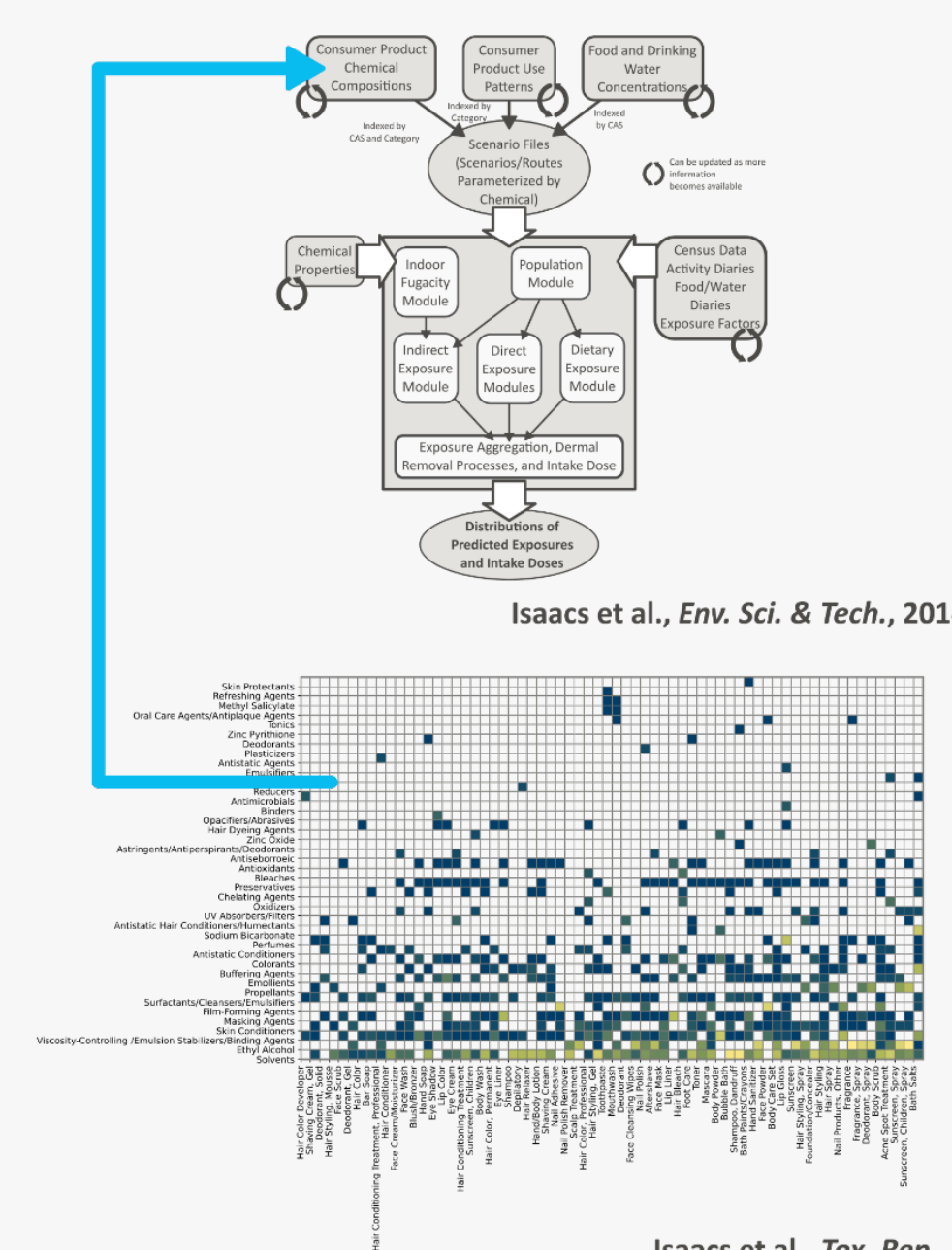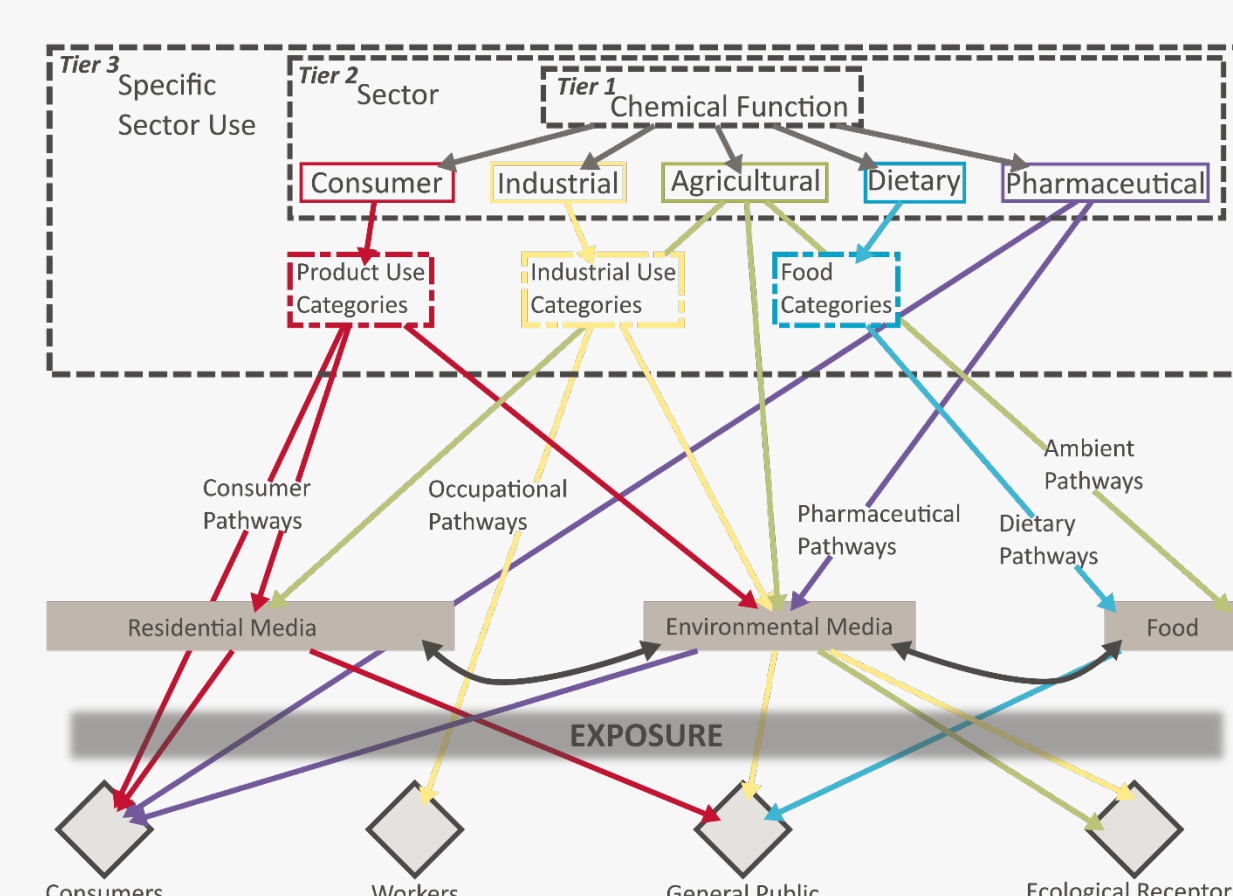Sakshi Handa EPA/ORD/CCTE
Kristin Isaacs EPA/ORD/CCTE

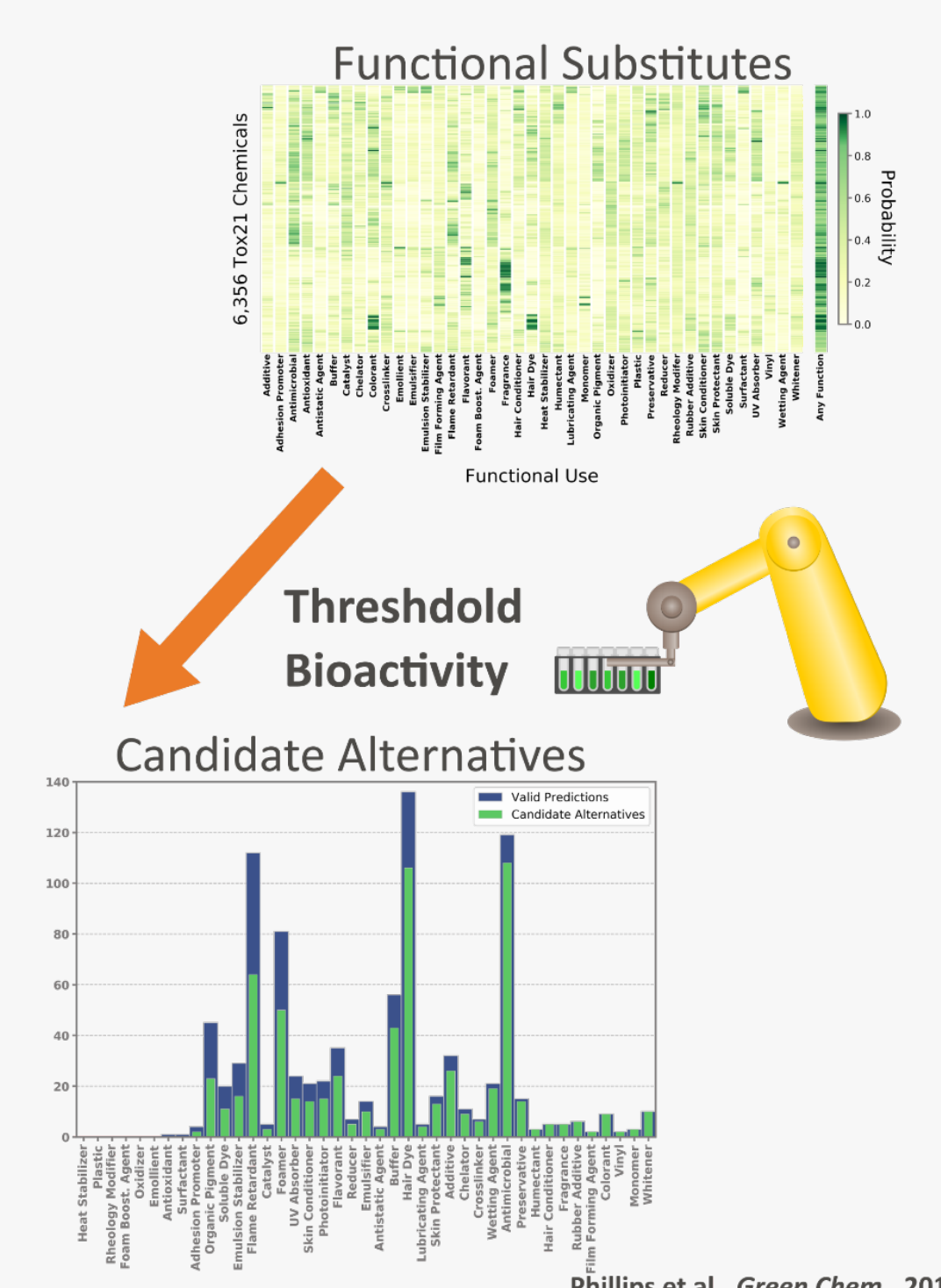Contact: phillips.katherine@epa.gov

Abstract ID: 445

QSURs: A History

Quantitative Structure Use-Relationship (QSUR) models were developed to predict a chemical's role in a consumer product and have been applied to many research areas at EPA. The original QSUR models were constructed more than 5 years ago, but data collection and curation has since expanded the possible training set for these models. Using new tools and data, these models have been rebuilt and their coverage of chemical space has been expanded.

Functional use provides critical information regarding potential exposure pathways.

Chemicals with no reported use

Suite of QSUR models

Predicted uses for chemicals

QSUR models predict functional use from chemical structure.

Isaacs et al., *Env. Sci. & Tech.*, 2014

Isaacs et al., *Tox. Rep.*, 2016

Input parameters to high-throughput exposure models

Functional Substitutes

Threshdold Bioactivity

Candidate Alternatives

Phillips et al., *Green Chem.*, 2016

Method for screening libraries of chemicals for candidate alternatives

Unique Chemicals

Chemical Function

Phillips et al., *Env. Sci. & Tech.*, 2017
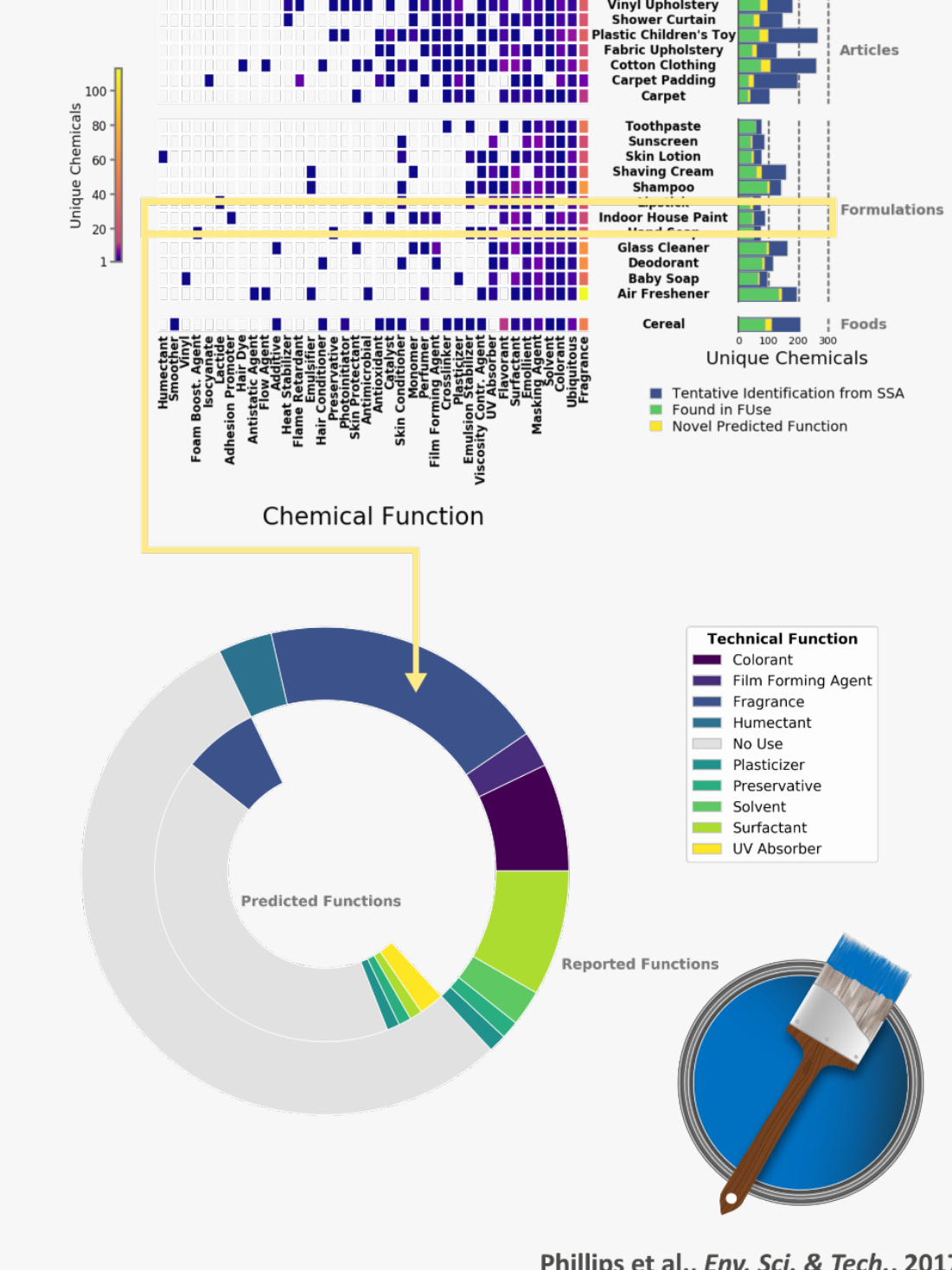
Bolstering non-targeted/suspect screening analysis identifications

QSUR 2021
QSUR 2017
Data Source

Bevington et al., *Sci. Data*, 2022
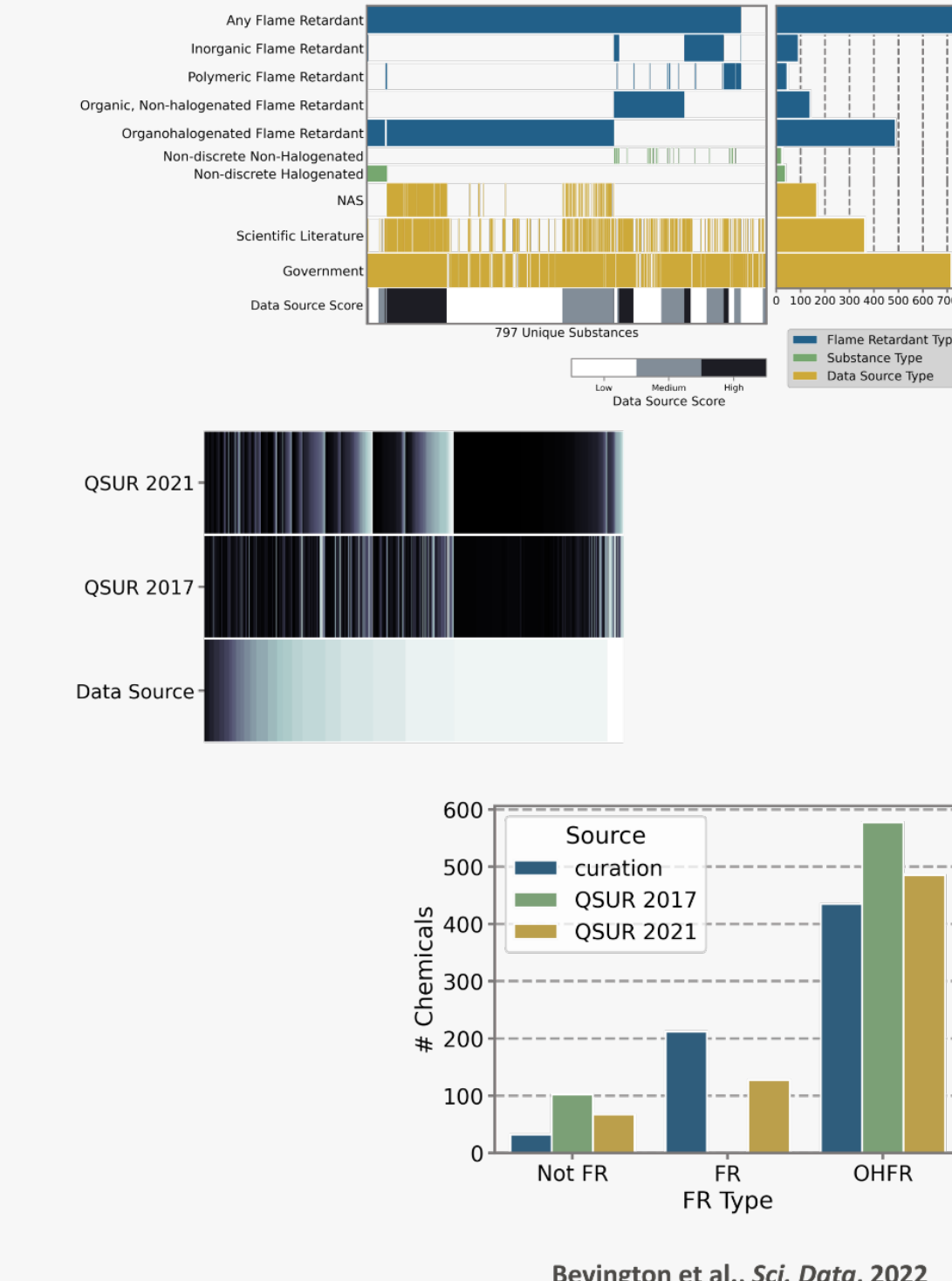
Screen potential flame retardants for risk assessment

Better ingredients = better ~~pizza~~ *chemical use models*

Developing New Models

EPA's <u>internal</u> Factotum curation tool allows exposure curators to collect, curate, annotate, and store exposure-relevant data from credible, publicly available sources. This tool keeps track of original data sources to aid curators in QA/QC processes of data collected.

EPA's <u>public</u> ChemExpo Knowledgebase is an online data discovery tool that surfaces all curated, QC'ed data from Factotum. This new tool allows users outside the EPA to search, explore, and download exposure-relevant data.

Bulk downloading function category (FC) information from ChemExpo provides a training set with chemical functional use mapped to the OECD's technical functions.

Train the Model

Predict with the Model
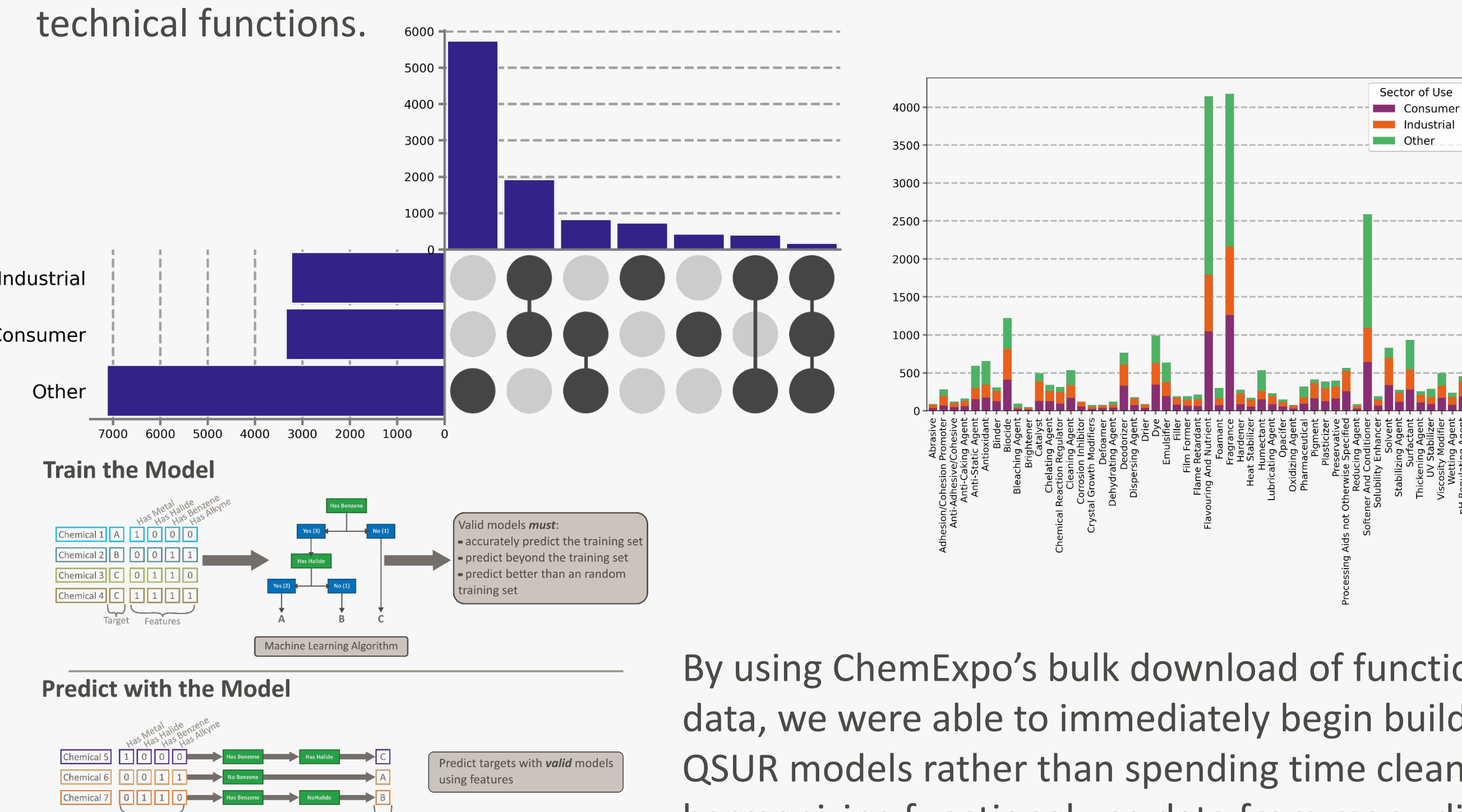
By using ChemExpo's bulk download of functional use data, we were able to immediately begin building new QSUR models rather than spending time cleaning and harmonizing functional use data from many different data sources. We further were able to include data from many different commercial sectors.
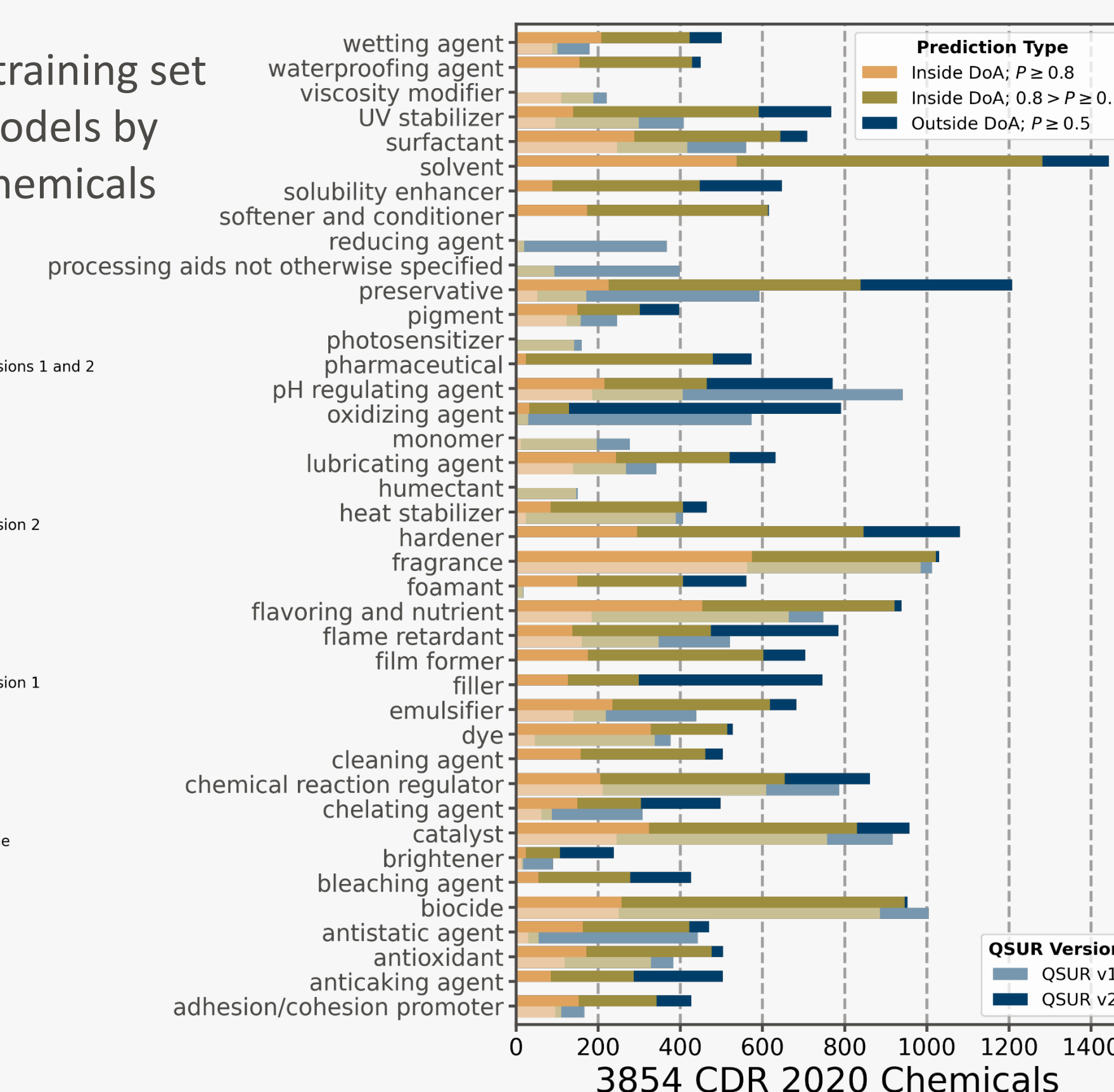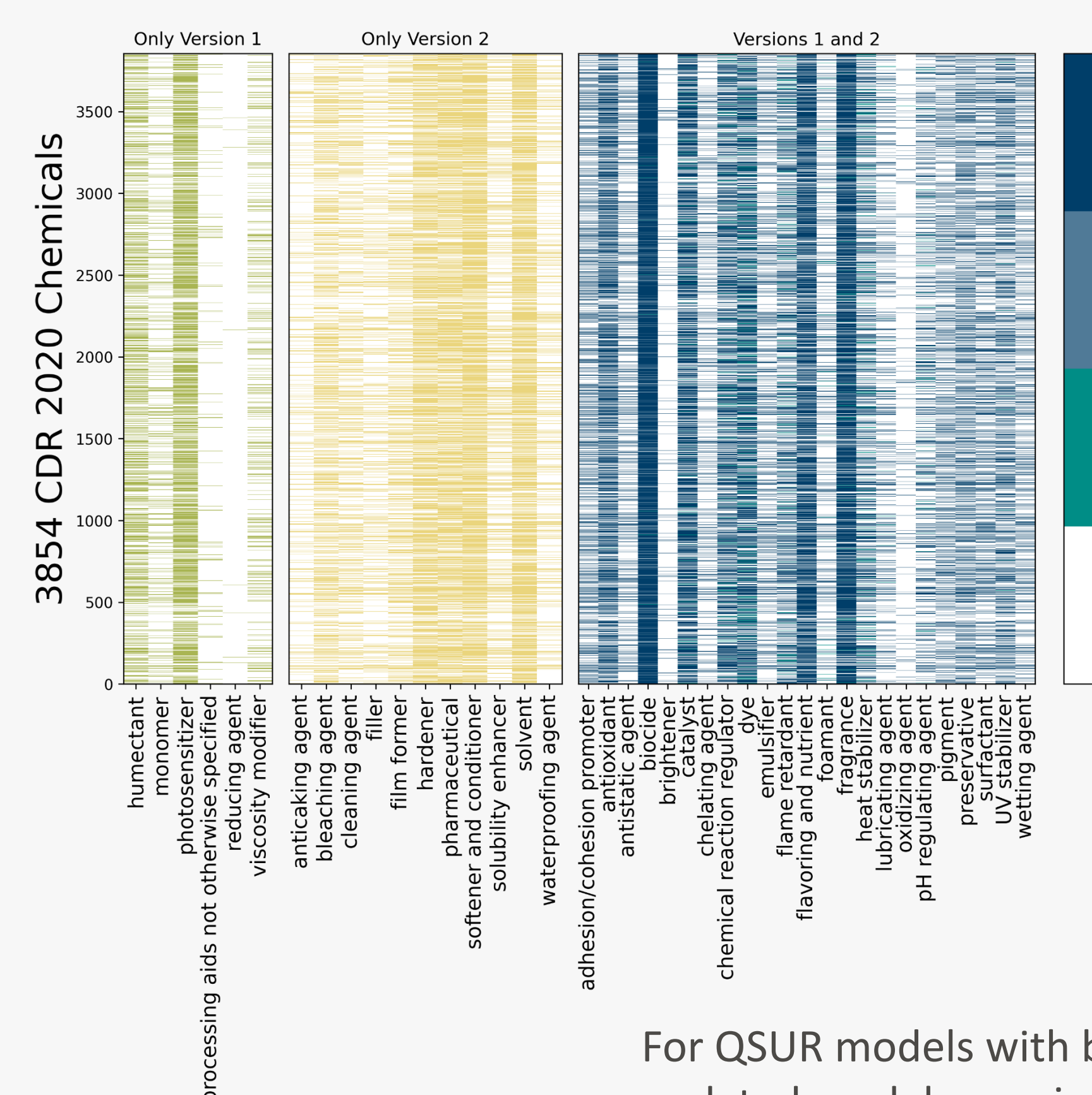
Evaluation of QSUR Models

Chemical Data Reporting under the Toxic Substances Control Act

The 2020 cycle of EPA's Chemical Data Reporting (CDR) was used as an external evaluation set as it has both industrial and consumer/commercial chemicals and functional uses of those chemicals have been reported with OECD technical functions.

QSUR models have been improved by:
1) Increasing number of positive predictions by increasing training set
2) Increasing the domain of applicability beyond original models by including industrial chemicals in addition to consumer chemicals

Only Version 1   Only Version 2   Versions 1 and 2

3854 CDR 2020 Chemicals

Prediction Type
Inside DoA, P ≥ 0.8
Inside DoA, 0.8 > P ≥ 0.5
Outside DoA, P ≥ 0.5

QSUR Version
QSUR v1
QSUR v2

3854 CDR 2020 Chemicals

For QSUR models with both existing and updated models, version 2 (new) models cover more chemicals in their modeling domains.