A Cheminformatics Workflow for Higher-Throughput Modeling of Chemical Exposures From Biosolids

Paul M. Kruse[1,2], Caroline L. Ring[2]

1. Oak Ridge Institute for Science and Education; Oak Ridge, TN
2. Center for Computational Toxicology and Exposure, United States Environmental Protection Agency; Research Triangle Park, NC

Abstract number: 4397
Poster number: P184

63RD ANNUAL MEETING & ToxExpo • March 10–14, 2024 SALT LAKE CITY, UTAH

Paul Kruse | kruse.paul@epa.gov | 0000-0001-5516-9717

Introduction

- Biosolids are treated sewage sludge produced as a byproduct of the wastewater treatment process
- Under the Clean Water Act, the US EPA Office of Water (OW) has the responsibility to protect human health and the environment from adverse effects of pollutants that may be present in biosolids
- EPA has identified over 700 chemicals found in previous National Sewage Sludge Surveys (NSSS) and Biennial Reviews (BR) [1, 2]
- OW has developed a Biosolids Screening Tool (BST), a software tool implementing a model for risk screening for potential human and ecological exposures to biosolids chemicals
- The BST models two disposal pathways: land application of biosolids and landfilling of sewage sludge

Methods

- The BST requires many chemical-specific input parameters, including physico-chemical and fate and transport property data
- We created the R package ccdR to access and retrieve data from the Center for Computational Toxicology and Exposure (CCTE) APIs
- We developed an automated workflow to collect and process the input parameters the BST requires to run for each chemical. This new R-based workflow:
  - Integrates chemical information from publicly available cheminformatics databases and tools
    - EPA CompTox Chemicals Dashboard (CCD) [3]
    - OPEn (quantitative) structure-activity/property Relationship App (OPERA) [4]
    - ClassyFire chemical-classification tool [5]
    - *Httk* R package [6]
  - Interfaces with the existing Microsoft Access implementation of the BST via Microsoft Excel input/output

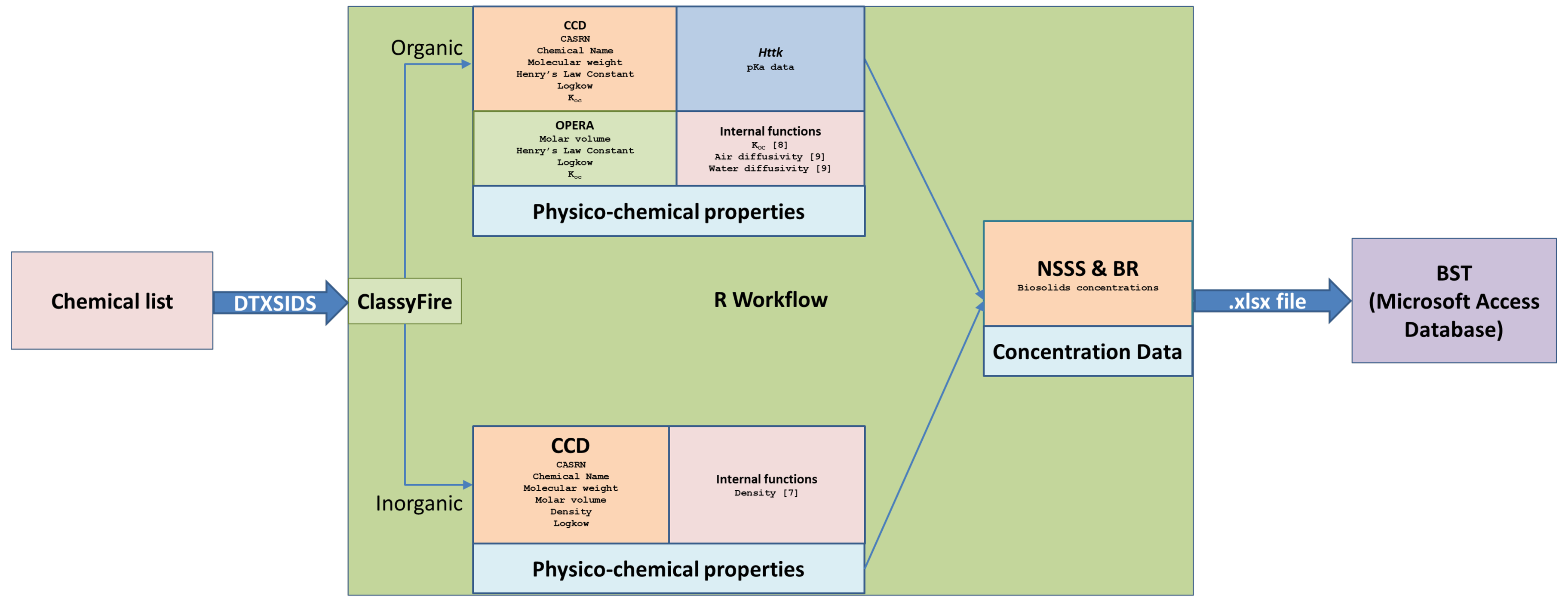

Figure 3: From a list of chemical identifiers, the R workflow interfaces with publicly available cheminformatics resources to create the .xlsx input file for the BST. The ClassyFire API is called to stratify chemicals and physico-chemical property data is obtained via the newly developed ccdR R package (OPERA predictions and CCD data), httk functions, and internal functions. Lastly, chemical concentration data is added via biosolids monitoring databases.

Workflows

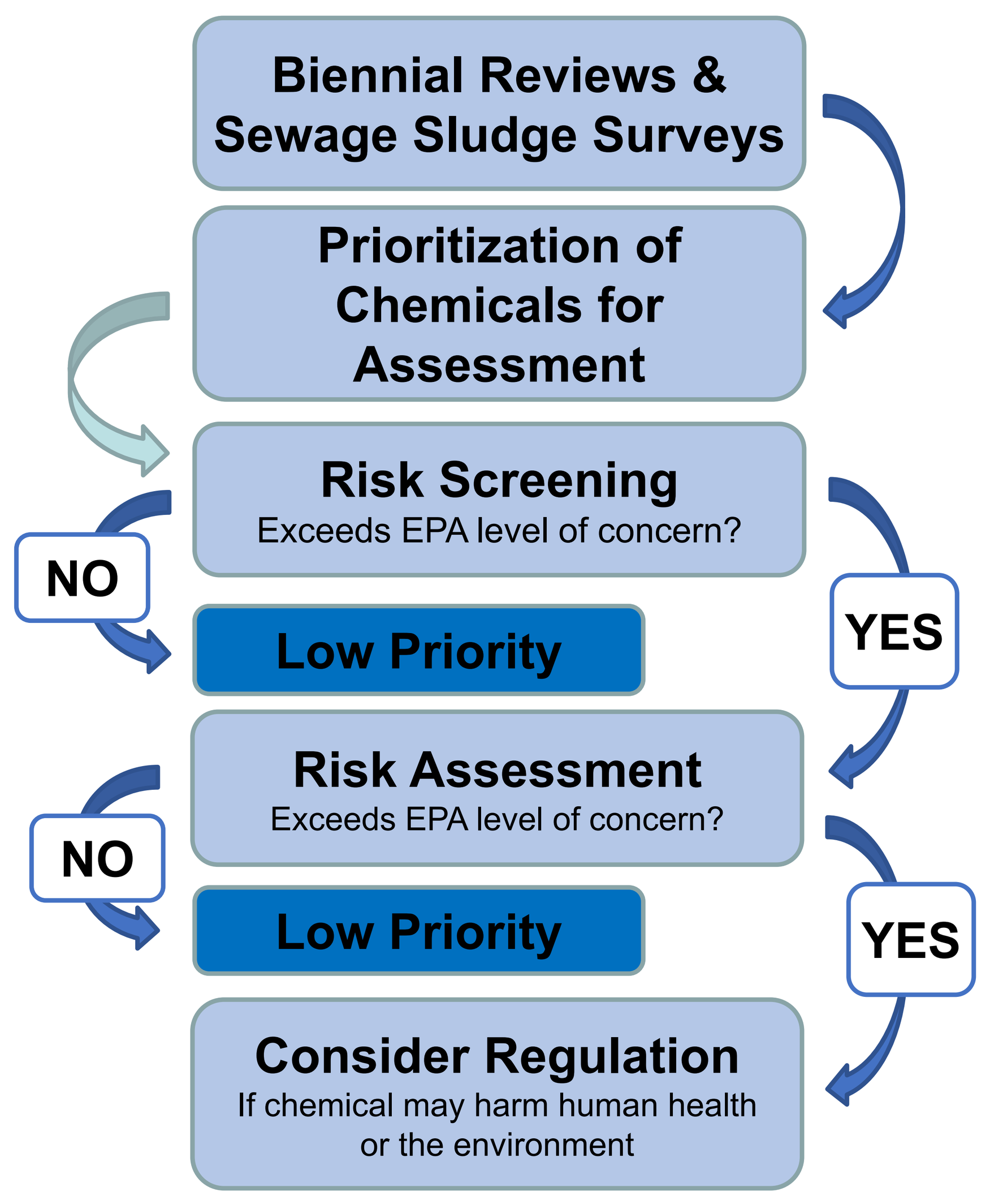
Biennial Reviews & Sewage Sludge Surveys
Prioritization of Chemicals for Assessment
Risk Screening
Exceeds EPA level of concern?
NO
YES
Low Priority
Risk Assessment
Exceeds EPA level of concern?
NO
YES
Low Priority
Consider Regulation
If chemical may harm human health or the environment

Figure 1: EPA's proposed prioritization and risk assessment framework for chemicals in sewage sludge and biosolids

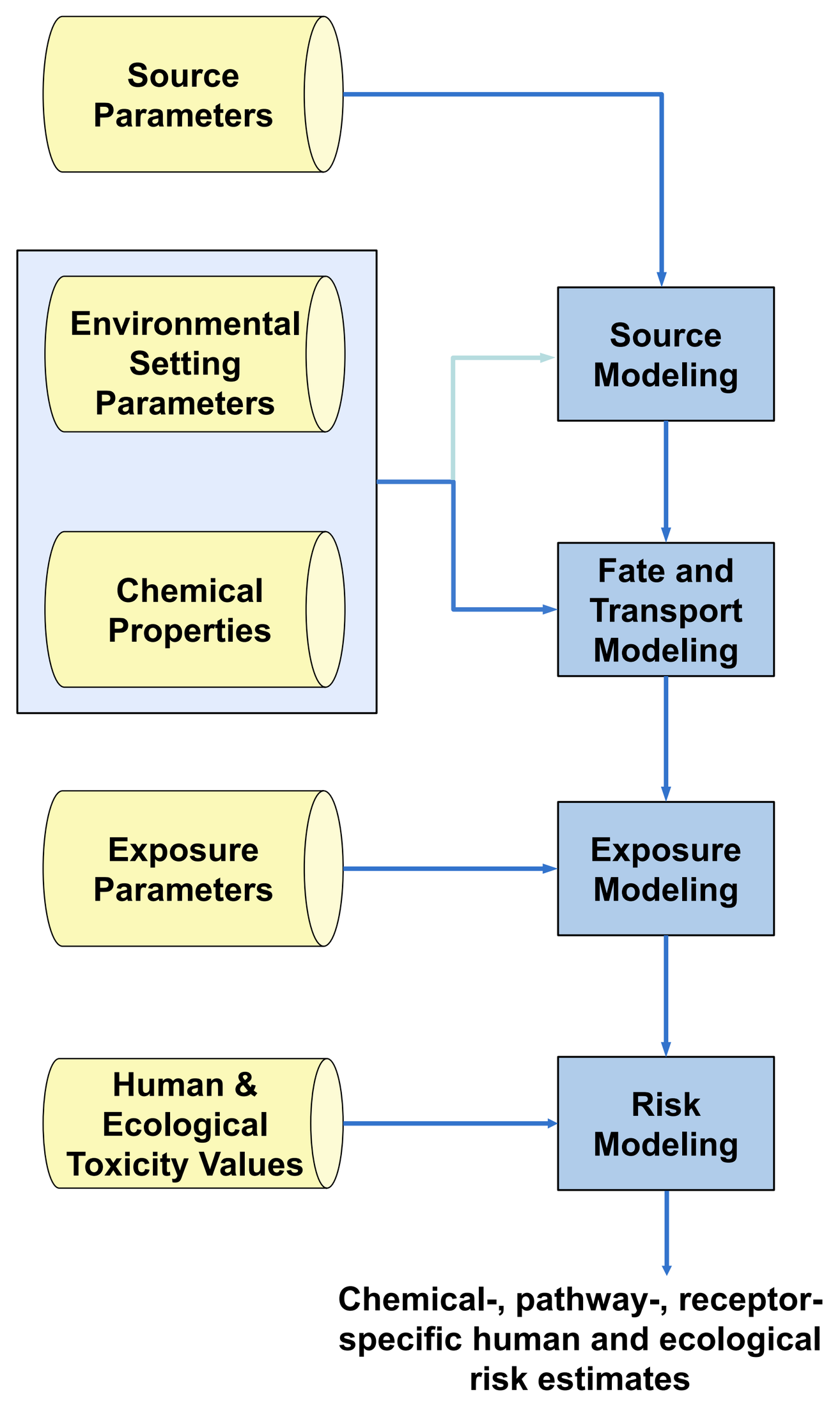
Source Parameters
Environmental Setting Parameters
Chemical Properties
Exposure Parameters
Human & Ecological Toxicity Values
Source Modeling
Fate and Transport Modeling
Exposure Modeling
Risk Modeling
Chemical-, pathway-, receptor-specific human and ecological risk estimates

Figure 2: Overview of BST conceptual framework

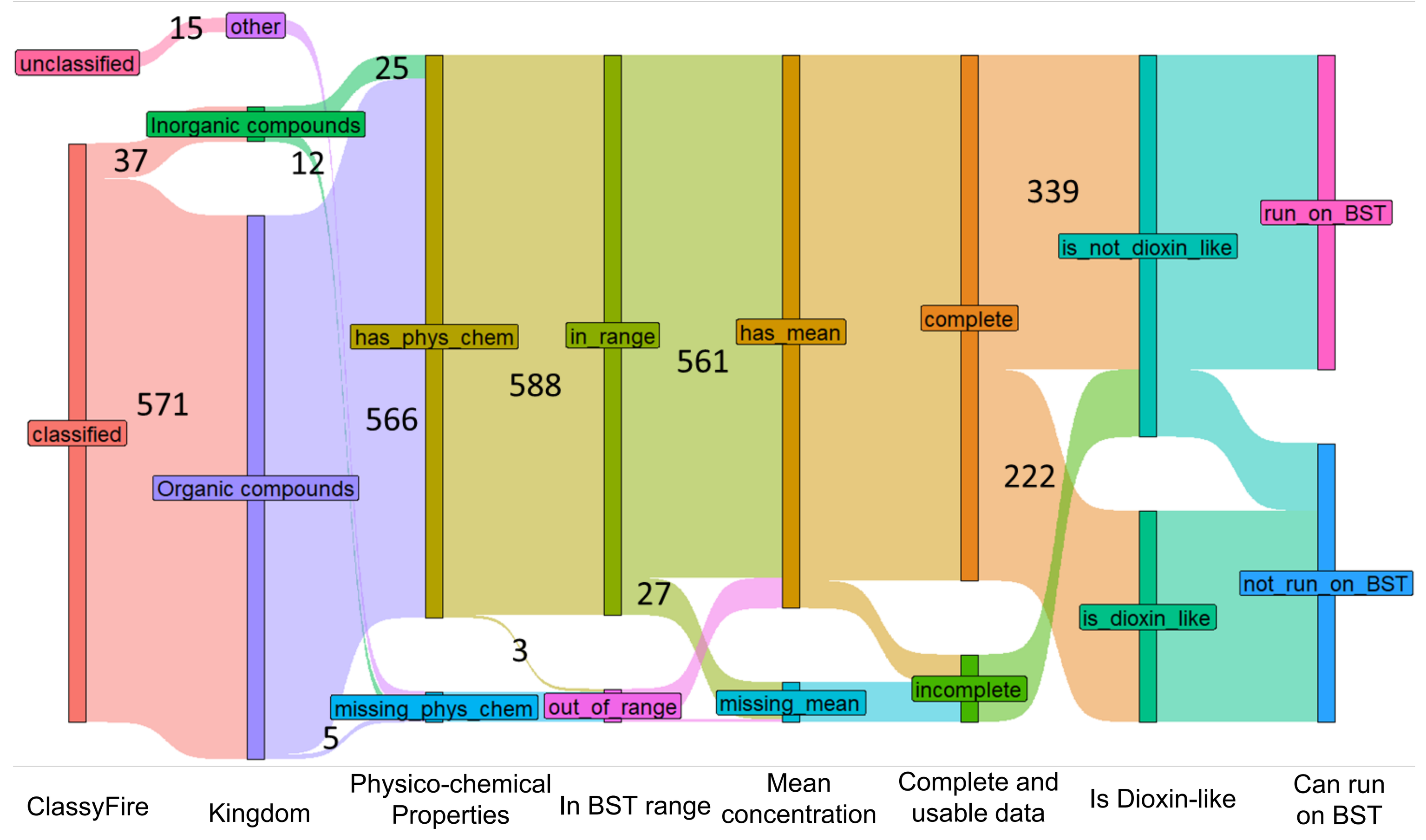
unclassified
15
other
25
Inorganic compounds
37
12
classified
571
Organic compounds
566
has_phys_chem
588
in_range
561
has_mean
complete
339
is_not_dioxin_like
run_on_BST
222
is_dioxin_like
not_run_on_BST
27
3
5
missing_phys_chem
out_of_range
missing_mean
incomplete
ClassyFire
Kingdom
Physico-chemical Properties
In BST range
Mean concentration
Complete and usable data
Is Dioxin-like
Can run on BST

Figure 4: The availability of physico-chemical data, concentration data, quality of the data, and restrictions of the domain of applicability of the BST model filter out chemicals in a variety of ways. A selection of filter points illustrates this process, starting with the input list of chemicals and ending with the chemicals prepared for upload to the BST for mean concentration simulations.

Results and Discussion

- 591 of the 623 Biosolids chemicals with ClassyFire classifications and concentration data had available physico-chemical property data. An additional three chemicals were removed due to incomplete physico-chemical property data or property data values outside the range of applicability for the BST, leaving 588
- Removed 222 Dioxin-like compounds from this list since they require cumulative exposure assessment and cannot be modeled appropriately using the BST, leaving 366 chemicals
- Removed chemicals missing sufficient biosolids concentration data, leaving 339 remaining chemicals for mean concentration and 345 for 95th percentile concentration simulations
- The workflow for collecting physico-chemical property data, filtering out chemicals missing necessary values for running the BST, and correctly formatting data for use within the BST takes under 15 minutes to execute for roughly 700 chemicals
- The BST takes about 30 hours to run all possible scenarios for roughly 340 chemicals (average/wet/dry climate types, surface disposal/land applied use, mean/95th percentile biosolids concentration)
- The workflow is easily adjusted to accommodate high-throughput screening efforts, and the BST runs on the output of the R-based workflow

Conclusion

- This workflow combines several publicly available cheminformatics tools and databases to prepare chemicals for screening using the BST
- This high-throughput workflow is easily adaptable for facilitating rapid chemical prioritization under other chemical fate scenarios
- The workflow leverages the ccdR R package we developed, which can be deployed rapidly and easily in a variety of data pipeline workflows

References

Conflict of interest: The authors declare that they have no conflict of interest. ***The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.***

Funding: This poster was supported in part by an appointment to the Research Participation Program at the Office of Research and Development, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA.