Expanding the chemical space of fragmentation spectra-based risk prediction models

Helen Sepman[1,2], Ida Rahu[1], Drew Szabo[1], Gordian Sandberg[1], Emma Apelgren[1], Anneli Kruve[1,2]

1) Department of Materials and Environmental Chemistry, Stockholm University, Svante Arrhenius väg 16, 106 91 Stockholm, Sweden;
2) Department of Environmental Science, Stockholm University, Svante Arrhenius väg 8, 106 91 Stockholm, Sweden

1. Background

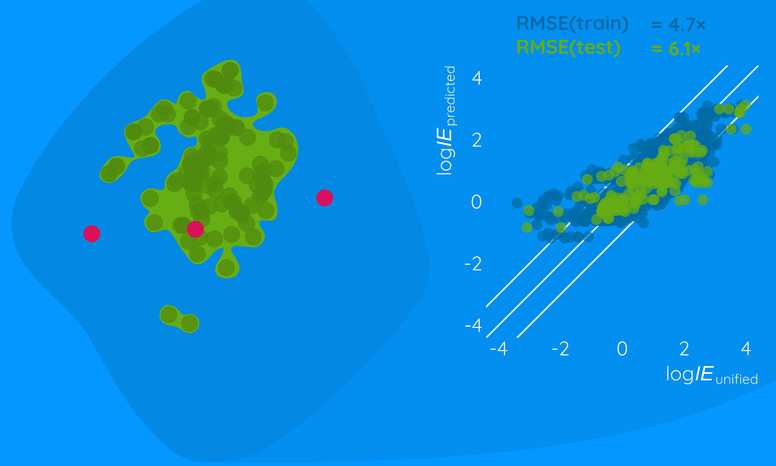
Prediction models provide the opportunity to rapidly estimate the risk posed by LC/HRMS features detected in non-target screening (NTS). Such models are developed on datasets that differ from heterogeneous real-life samples. Validation on a wide range of chemicals can reveal uncovered regions in chemical space.

2. MS2Quant

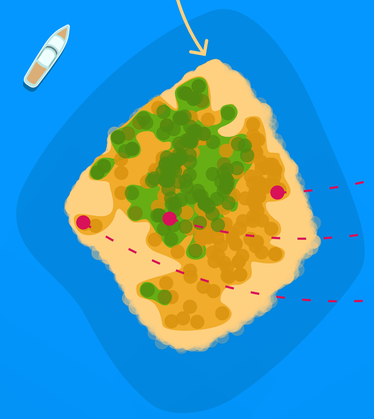
MS2Quant is a concentration prediction model for tentatively identified or unidentified LC/HRMS features with MS^2 spectra.[1] The model is trained on structural fingerprints using ionization efficiency (*IE*) values of 1191 chemicals measured in positive electrospray ionization mode.

2.1. Latest model for ESI-

Recently, Lauria et al. developed a molecular descriptors-based *IE* prediction model with a focus on PFASs (a total of 132 chemicals).[2] The performance of the structural fingerprints-based model developed on the same dataset was comparable to the models developed by Lauria et al.



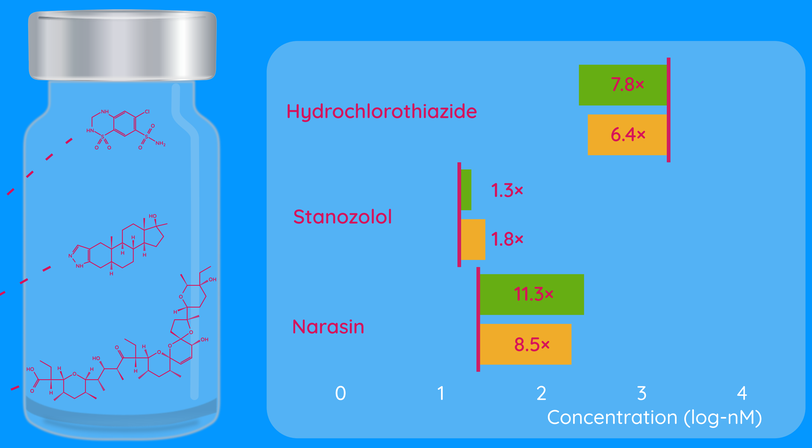
Principal component analysis; to visualise the chemical space of environmental contaminants, **PubChemLite** database was used as basis.



Chemical space was expanded to predict concentrations in ESI- mode

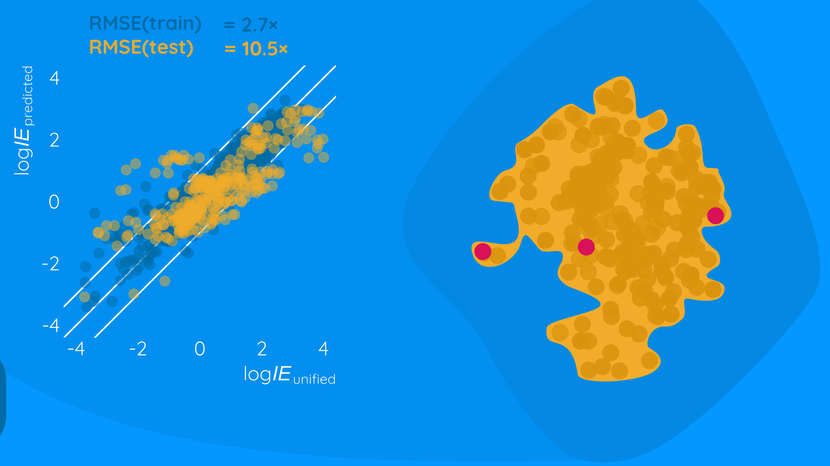
3. Interlaboratory comparison

During an interlaboratory comparison in autumn 2023, three high-risk chemicals were identified in spiked blind samples. The concentrations predicted for these chemicals in ESI+ mode varied from 2× (stanozolol) to 30,000× (hydrochlorothiazide) indicating a low representation of similar chemicals in the training data.



4. Impact of added chemicals

The concentration predictions in negative mode were improved slightly compared to the model developed on data from Lauria et al. However, negative mode model predictions improved significantly compared to positive mode predictions. Evaluation of the positive mode MS2Quant model with chemicals similar to suspects added to the training data is ongoing.



1) H. Sepman, L. Malm, P. Peets, M. MacLeod, J. Martin, M. Breitholtz and A. Kruve, Anal. Chem., 2023, 95, 12329–12338.
2) M. Z. Lauria, H, Sepman, T. Ledbetter, M. Plassmann, A. M. Roos, M. Simon, J. P. Benskin and A. Kruve, Environ. Sci. Technol., 2024, 58, 2458–2467

The funding has been generously provided by Swedish Research Council for Sustainable Development grant 2020-01511. The authors declare no competing financial interest.

Helen.Sepman@aces.su.se

Kruvelab.com