Introduction

Due to a lack of analytical standards non-target analysis is mostly qualitative. Current semi-quantification relies on tentatively known structures. However, quantification is important for risk assessments and peak prioritisation even at lover levels of identification.

Method

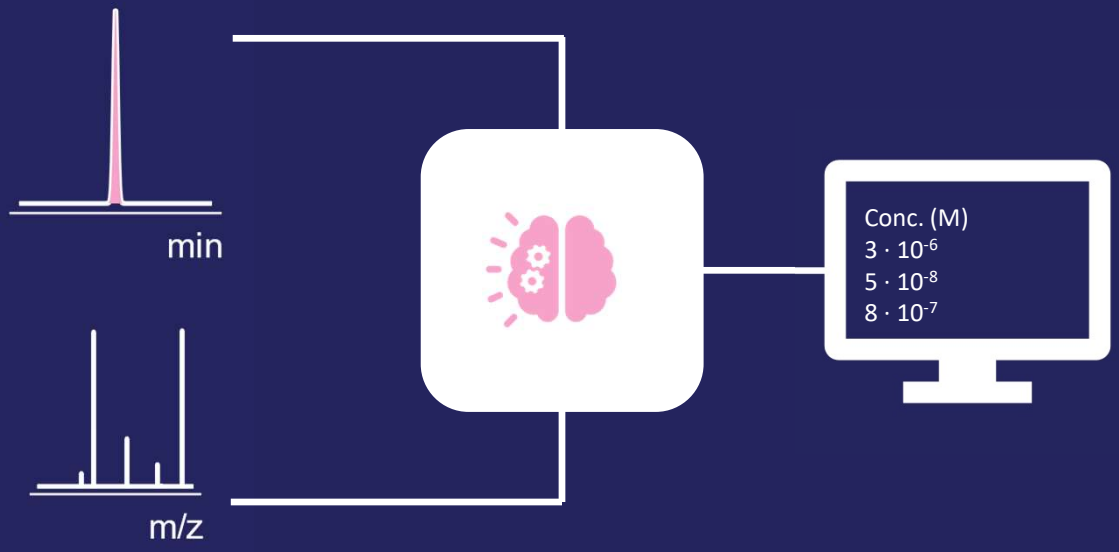
- Training test and validation set (92, 29 and 48 compounds respectively) were measured in reversed phase at pH 2.7, 8.0 and 10.0 in positive and negative ionisation mode.
- Random forest models were trained from 12 LC/MS features obtained from the training set.
- The models were applied to the test set and the validation set (2 blind spiked water samples).
- The models were compared to two baseline approaches and to a previously developed model that uses 2D structural features.

Results

- The model using LC/MS features performed better than the two baseline models and had a mean error factor of 10.4 for the test set. See Figure 1.
- The most important feature was the relative intensity between positive and negative mode measurements. See Figure 2
- The spiked blind validation samples had a mean error factor of 6.0.
- The model had comparable accuracy to models using 2D structural features. See Figure 3
- The model could be simplified to a linear model using only the logarithm of the intensity ratio in positive and negative mode at pH 8. The mean error factor of the test set in this case was 1.7. See Figure 4

Article: Non-target quantification of water contaminants based on chromatographic and mass spectrometric properties using LC/ESI/HRMS
Emma Palm, Anneli Kruve
Submitted for publication

Non-target quantification can be done directly from LC/MS features



Emma Palm, Miklos Mohai, Louise Malm, Anneli Kruve
Stockholm University, Department of Materials and Environmental Science

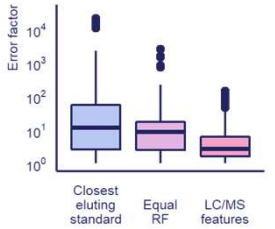


Figure 1 Error factors for the random forest model as well as the baseline models

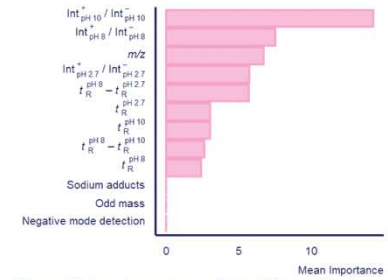


Figure 2 Mean importance of the different features used in the random forest models

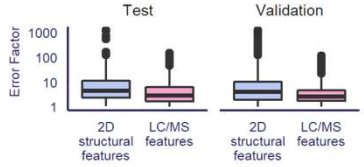


Figure 3 Error factors for the test and validation set for the developed LC/MS features models and a model using 2D structural features.

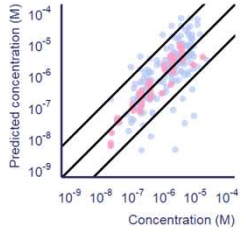


Figure 4 Predicted and true concentrations for the training (blue) and test (pink) datasets using the simplified model.