Getting big results from big data: Statistics meets mass spectrometry

Simon Spencer and Mark Barrow

As analytical techniques have advanced, new types of experiment become available, results are obtained at higher performance and incorporating more variables, and there is great potential for much more detailed scientific insights. There is the on-going challenge, however, of how to address much larger and more complex datasets and statistical methods in the analytical sciences have not always kept pace with the advances in experimental capabilities.

Collaboration between Simon Spencer and Mark Barrow has sought to address some of the issues, including development of three data processing approaches, coded in R: Themis, KairosMS, and Rhapso. Themis is used as part of a workflow for replicates of ultrahigh resolution data sets. KairosMS is used for hyphenated ultrahigh resolution experiments, such as the combination of Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) with chromatography or other separation methods, and has enabled new experiments to be performed. Rhapso is part of an advanced data stitching approach, named OCULAR (operation at constant ultrahigh resolution). The development of OCULAR has dramatically increased the performance of instruments and has led to a new world record in the number of assignments of molecular formulae in a single complex mixture, with 244,779 assignments.

In this project, you will develop the statistical methods required to extract scientific information from data results acquired using some of the world’s most advanced mass spectrometers. The Ion Cyclotron Resonance Laboratory (ICR Lab) at Warwick currently houses a modified 12 T solariX FTICR, a 15 T solariX XR FTICR (with new electronics and cell design), and a timsTOF Pro for ion mobility experiments. We have access to a range of ionization sources, separation methods, and dissociation techniques. A key aspect of the previous success with a joint PhD student, including the aforementioned world record, was close collaboration between experimentalists and statisticians.

The project will provide opportunities for both experimental work and computational work, such as writing code to fit statistical models, remove noise, and develop data analysis pipelines. Only by developing a fuller understanding of the measurement process through sample preparation, data acquisition, and data analysis and visualization can the full benefit of advanced analytical techniques be fully realized. There is scope for flexibility in project design to ensure the right balance for the student. If you have any questions, we would be very happy to talk to you (S.E.F.Spencer@warwick.ac.uk and M.P.Barrow@warwick.ac.uk).