

## Mass spectrometry quality control

Results of a mass spectrometry experiment = subject to a large variability  
→ **Quality control metrics**

Quality control data = **standardized quality control samples** (e.g. BSA samples):

- Extensive frequency: Periodically ran before, during, and after experimental samples → Detect problems as soon as possible
- Low complexity → Limited variability

Two sources of quality control metrics providing complementary insights:

- **Derived from experimental results**<sup>1,2</sup>: Capture high-level operational characteristics of a mass spectrometer
- **Instrument settings**<sup>3</sup>: Indicate which part of the instrument is malfunctioning

## Instrument monitoring

Advantages of **monitoring instrument settings**:

- Can be directly related to the part of the instrument that is malfunctioning
- Highly sensitive → Small differences in instrument behavior will be rapidly detected

Logging of **external events** that occurred to the mass spectrometer instrument:

- E.g. calibrations, (periodic) maintenance, unexpected incidents, ...
- Vital information when interpreting the evolution of the instrument settings over the course of multiple experiments

**iMonDB: Instrument MONitoring DataBase**

- Tools to extract instrument settings from experimental raw files and store these in a database for longitudinal analysis
- Tools to visualize instrument settings and to record structured information about external events (Figure 1)
- API for use by developers
- Tools, documentation and source code released as open source and freely available from: <https://bitbucket.org/proteinspector/jmondb/>

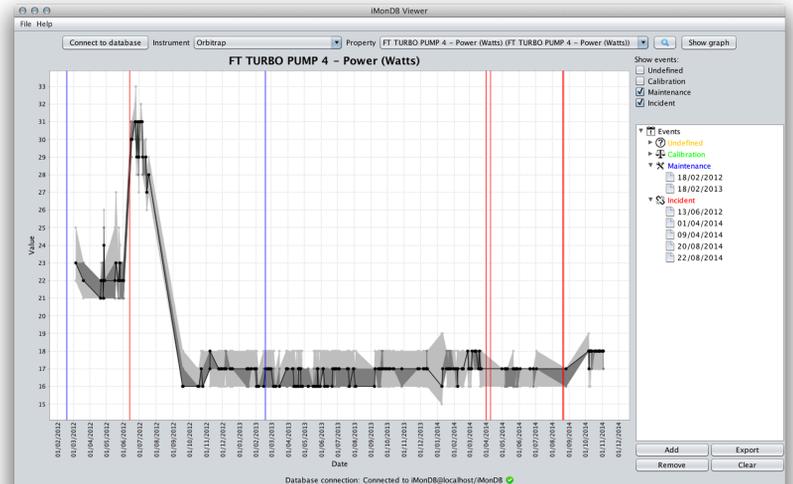


Figure 1: Instrument behavior can be monitored over time and can be related to external events. For example, the turbo pump of the mass spectrometer broke down, resulting in an increased power consumption, and had to be replaced.

## Pattern mining

Due to the curse of dimensionality, specialized data mining techniques are required to detect patterns in the high-dimensional quality control data.

**Subspace mining:**

- Find a suitable subset of the original feature space by disregarding irrelevant dimensions
- Within each subspace: clustering, outlier detection, ...

**Subspace clustering:**

1. Subspace detection: Cartification<sup>4</sup> & frequent itemset mining<sup>5</sup>
2. Clustering in the detected (low-dimensional) subspaces

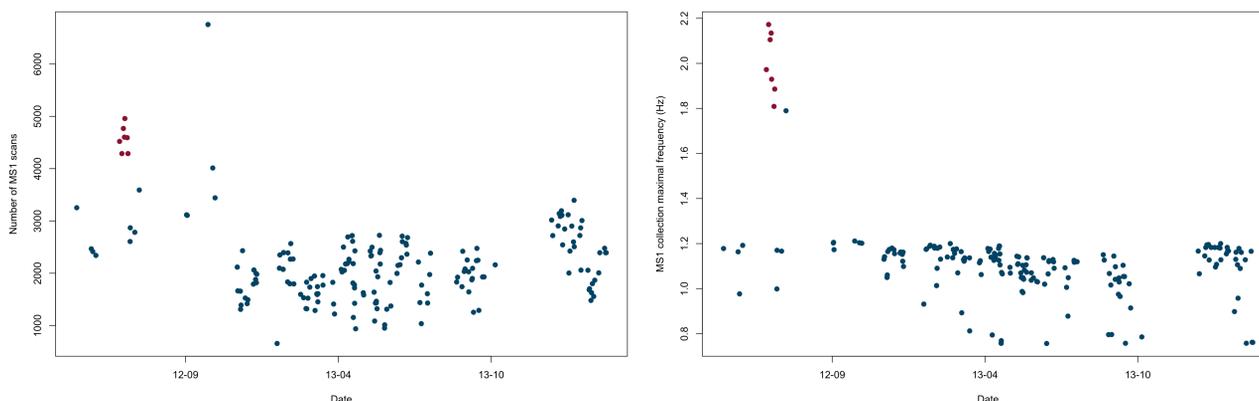


Figure 2: Subspace clustering is able to detect relationships between metrics. For example, experiments that have a higher number of MS1 scans due to a higher MS1 collection frequency are clustered together.

**Detected subspaces:**

- Related metrics: significant overlap with previous manually defined groups of co-occurring metrics
- New relationships between metrics to be validated using expert knowledge

**Detected clusters:**

- Highly dependent on projected subspaces
- Capture valid relationships between experiments (Figure 2)
- Indicate faulty experimental results (Figure 3)

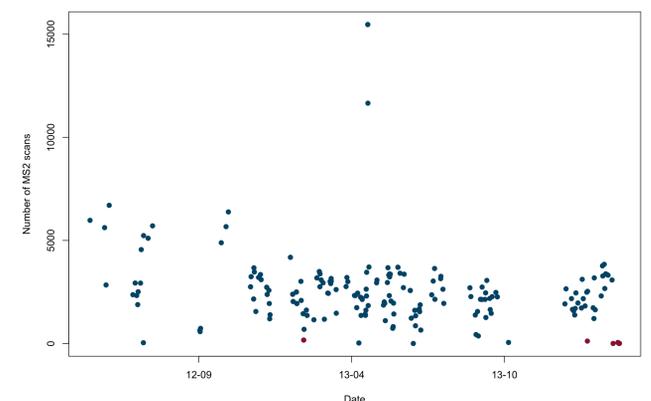


Figure 3: Clusters can be linked to instrument performance. For example, the highlighted cluster indicates experiments where no MS2 scans were collected due to a broken S-lens.

## Conclusion

The awareness has risen that suitable quality control information is mandatory to assess the validity of a mass spectrometry experiment. A complementary source of qualitative information is available in the form of the mass spectrometer instrument settings, which provide additional insights into the operational characteristics of a mass spectrometer. Furthermore, we have developed specialized pattern mining algorithms to interpret this high-dimensional data. The derived patterns could subsequently be used to optimize experimental design and mass spectrometry instrument settings.

## References

1. Rudnick et al. (2010) *Molecular & Cellular Proteomics* **9**:225–241.
2. Ma et al. (2012) *Analytical Chemistry* **84**:5845–5850.
3. Bittremieux et al. *In preparation*.
4. Aksehirli et al. (2013) *ICDM '13* 937–942.
5. Naulaerts et al. (2013) *Briefings in Bioinformatics*.