Software Developments

James Allan
CAS/SEAES, University of Manchester
DIAC

A quick history…
Let’s see how old people really are…
2000: Quad MS software

- Carried over from previous work by Paul Williams & Phil Silva
  - Corresponding TOF analysis by Jose
  - Sucked

2001: ‘Unified’ version

- Incorporated both MS and TOF modes
  - Could actually produce proper data
  - Still handled each mass spectrum individually
  - Hellishly slow for anything but the smallest datasets
2002: ‘Deluxe’ version

Used monolithic, indexed data repositories
- No new functionality initially, but considerably faster (i.e. usable)
- Later down the line, incorporated plugins by Alice, Tim & Jonny
- Still used for quad analysis to this day
- But by no means perfect…

Core features

Deliver mass concentrations
- Mass distributions
- Mass spectra
- All quantitative and separated by chemical species
2005: Started work on Squirrel

A fresh start

The quad software works, but it has many major limitations:

- Can only deal with so much data
  - 10 Igor experiments for Chebogue point, averaging 620Mb each…
- Very limited selective processing
  - Certain operations using a mask waves would still require an entire dataset to be looked at
- Not brilliant at providing data usable for higher analysis
  - Remapping functions worked, to a point…
- All of these problems stood to get much worse with the advent of the ToF-AMS
Things to keep

Frag table system
- It works, it’s flexible and it’s proven

• Graphing system
  - Doug would get culture shock otherwise

• Igor itself
  - People already used to it
  - It’s powerful
  - Makes publication-quality graphs
  - It’s flexible
  - Fantastic support and user community
  - It’s cheap (compared to equivalent software)
  - Works on a Mac (if you’re into that sort of thing)
  - James would get culture shock otherwise

Squirrel design

Access raw data directly from a hard drive
- Made possible by adoption of HDF5
- Save any intermediate data back to the hard drive, saving memory

• Be able to process any combination of data as required
  - Individual runs, run intervals, ionisation methods, inlet status, meteorological conditions, etc., etc.,

• Be scalable
  - Only process as much data is needed for an individual calculation

• Be flexible
  - Allow the processing of quantitative data in whatever time, diameter or m/z space as required, be it a prescribed time base for a database, impactor sampling periods, flight legs or whatever

• Be expandable
  - Allow many individuals to work on the code independently
Key design features

Index data rather than loading it into memory
- Use a common interface for all functions to access the data
- Use ‘todo’ lists rather than specifying run intervals or mask waves
- Only load as much data as is needed, break into chunks if required
- Do all of this without any intervention by the operation being performed

Protocol

Each save identified by a unique run number
- Same as the quad
- All data types identified by name
  - MSOpen, MSClosed, MSSDiff, MSSClosed, pTOF, pTOF_sticks
- All operations retrieve data via Squirrel
- Every data operation requires a todo list
  - Even if it is only one run long
- Each operation is completely ignorant of how it gets its data and shall allow for multiple chunks of data
How it works

Front end (panel)

Data management

Index

Operations

Pretty graphs!!

Implementation

Started during SOAR
  – Coded by James, Jonny and Ed
  – Partly based on the quad software and Silke’s existing analysis software
• Worked on sporadically since
• Went up a gear after Donna joined the team
• Tim has recently added some bits
Does it work?

Hell yeah!!
- Ron Brown leg 2 analysis:
  - 3 Weeks of data
  - Combination of 60 and 3.5 second saves
  - >50,000 runs
  - 2 ionisation schemes
  - 2 MS modes
  - 1 file
  - <300 Mb
    - Same file that I started on day 1 of voyage, used for all analysis
- Try doing that in the so-called ‘deluxe’ version…

Can it do everything the quad software can?
- No error analysis yet
- Doesn’t normalise size-resolved data to mass spec
- A few features still a little rough around the edges
- Basically it still needs some work, but we’re nearly there
Can I at least get QA data?

Again, we’re nearly there but not quite…
- m/z calibration and stick recalculation
  - Donna has done a fantastic job on this, but we’re still collectively learning things about the instruments
- The size-resolved normalisation thing
- Applying NO\textsubscript{3} calibrations a little smarter
- These will all get dealt with ASAP
- But hey, not bad for a little over a year’s work…

What it can do now

Airbeam corrections
- Apply calibrations
- Mass concentrations
- Speciated stick and raw mass spectra
- Size distributions
- Image plots
- Stick spectra over diameter intervals
- Fill in missing ‘closed data’
Control freak!!

- Specify time, diameter and m/z (for raw spectra) for all data products
  - Can specify ‘as saved’, a regular interval or user-defined bin edges
  - Changes to size and m/z calibrations are taken into account on a run-by-run basis

What to expect soon

- Size distribution normalisation
- Raw mass spectra as a function of diameter
- Allowance for Joel’s new operating modes and 2GHz data
- Error analysis
  - Partly written, but not completed
- Multiple calibrations within a campaign
- Journaling
  - We already record all the parameters used in data generation, we just can’t access it…
- High resolution analysis
  - Pete?
- Near real time analysis?
  - Maybe not needed – logging software has some of this
- ‘Write paper’ button
  - Maybe after the honeymoon…
Sources of information

ToF-AMS Website
- Hit the ‘FAQs, support’ button

- tofamsusers email list
  - More than for Marc’s opinions about charge symbols and duplicates of all the job adverts Jose sends to ams-users

- Donna and myself
  - Don’t be shy

Acknowledgements

Doug & Jose (as always)
- Donna, Ed, Jonny and Joel
- Wavemetrics, in particular Howard Rodstein
- Anyone who dared try it in the early days
  - Shane, Peter and Tim in particular
Any questions?