TOPCAT and STILTS: Software for Tables

TOPCAT http://www.starlink.ac.uk/topcat/

STILTS http://www.starlink.ac.uk/stilts/

Mark Taylor (AstroGrid/Bristol)

EuroVO AIDA VO School,
ESO Garching,

30 March 2009

$Id: tsoft.tex,v 1.11 2009/03/30 11:39:05 mbt Exp $

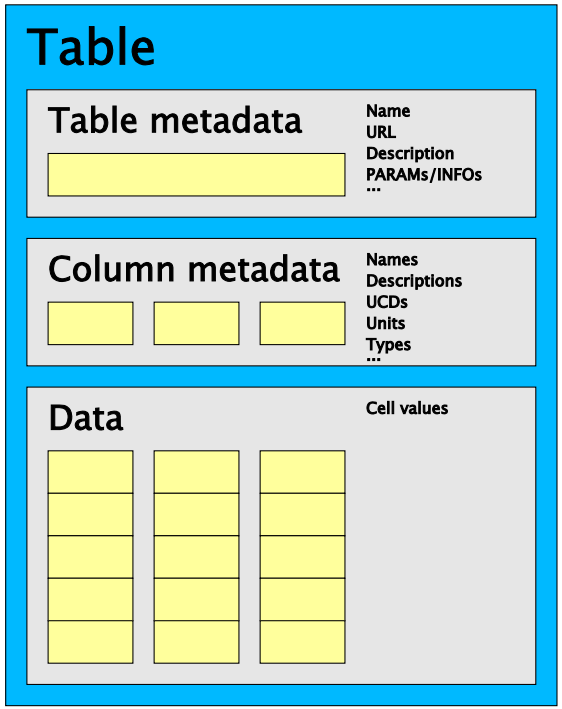
Outline

- What's a Table?
- Table processing: requirements and options
- TOPCAT/STILTS comparison
- TOPCAT capabilities: overview and demo
- STILTS capabilities: overview and examples

What is a Table?

What does a table consist of?

- Data in rows and columns
  - One or more row
  - Each column has single data type
- Column metadata
  - e.g. name, type, units, UCD, . . .
- Table metadata
  - e.g. title, author, history, . . .



Where do tables come from?

- Typically source catalogues
  - e.g. SExtractor output, survey data, Vizier, Cone search, ADQL results, . . .
- but may be other things
  - e.g. time series, event list, SIAP/SSAP query response, . . .

What form do they come in?

- FITS, CSV, TST, "plain" ASCII, RDBMS, VOTable, HDF5, AIPS++, . . .

```
# Name       RA      Dec
M31        10.50   41.266
M83       204.25  -29.866
NGC4414   186.61   31.224
```

Working with Tables

What do you want to do with tables?

- Read/Write in different data formats
- Look at and modify the data
- Look at and modify the metadata
- Add/remove/rearrange columns
- Make calculations based on columns
- Make row selections on various criteria
- Calculate statistics
- Visualise data graphically (plots, histograms)
- Join tables (especially perform cross-matches)

Available tools

- Spreadsheets
- Relational databases
- Programming languages (Python, Perl, IDL, C, FORTRAN, . . .)
- Purpose-built tools, e.g. TOPCAT & STILTS

TOPCAT/STILTS

Different interfaces:

TOPCAT 	**STILTS** 
GUI	Command line
Interactive	Scriptable
Easy to use	Reproducible
Good for data exploration	Good for batch/programmed use
Exploratory phase	Production phase
$\lesssim 10^6$ rows	Unlimited size (for most things)

Standard usage:

- start off with TOPCAT
- maybe move on to STILTS for more specialised requirements

TOPCAT/STILTS

Shared characteristics:

- General purpose table manipulation software
  - plus astronomical features (sky coordinates, fluxes, cross-matching, . . .)
  - plus VO features (PLASTIC, SAMP, SSAP, SIAP, Cone Search, VOTable, . . .)
- Reads/writes many, though not all, important astro data formats
  - FITS, VOTable, CSV, SQL, some ASCII, . . . — *extensible*
- Powerful algebraic expression language — *extensible*
- Can cope with large datasets
- Fast
- No programming required by the user
- Fully documented, both reference and tutorial
- Pure Java (→ multi-platform, easy to install)
- Open source
- Good support (for now)

"Does what you want with tables"

Capabilities

It can do:

- View/edit data
- View/edit metadata
- Make, combine, display, save row selections in various ways
- Plotting: 2D, 3D and histograms
- Crossmatching: sky or other coordinates, very flexible
- Load tables from VO services: Cone Search, SIA, SSA
- Talk to other tools using PLASTIC/SAMP: exchange tables, row selections, positions
- Do calculations on data
- . . . and more

Expression Language

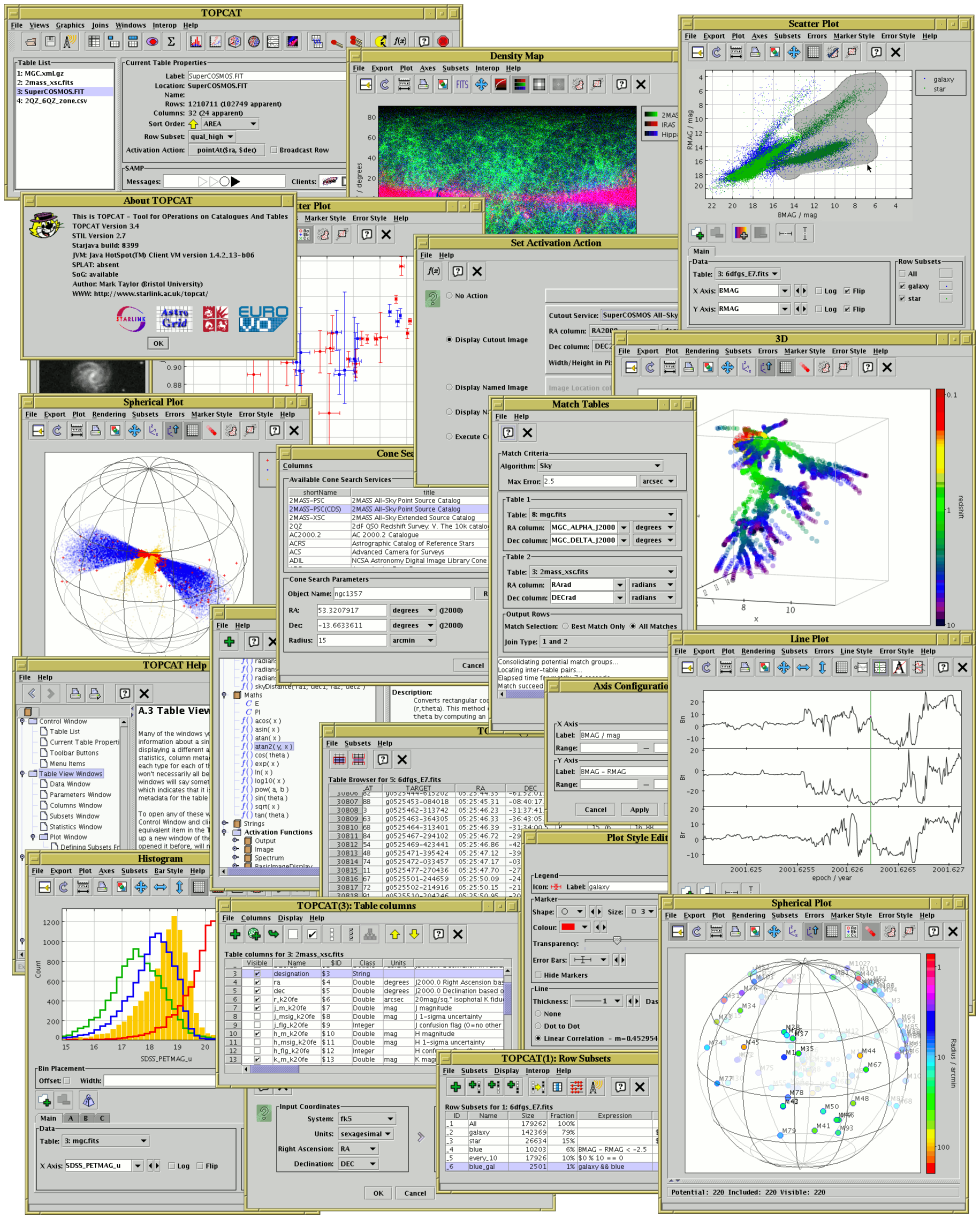
Algebraic language for adding new columns, defining selections etc

- Refer to values in row using column name (or index `$1`, `$2`, . . .)
- Standard arithmetic operators (`+`, `-`, `/`, `*`)
- Standard mathematical functions (`abs`, `max`, `round`, `sin`, `cos`, `pow`, . . .)
- Time conversions (ISO8601, MJD, Julian, Besselian)
- Sky coordinates (degrees, sexagesimal, sky distances)
- Cosmological distances (redshift, luminosity dist, lookback time, . . .)
- Fluxes (Johnson AB Magnitudes, Jansky)
- . . . and more
- It's extensible

Sky coordinate conversions also available

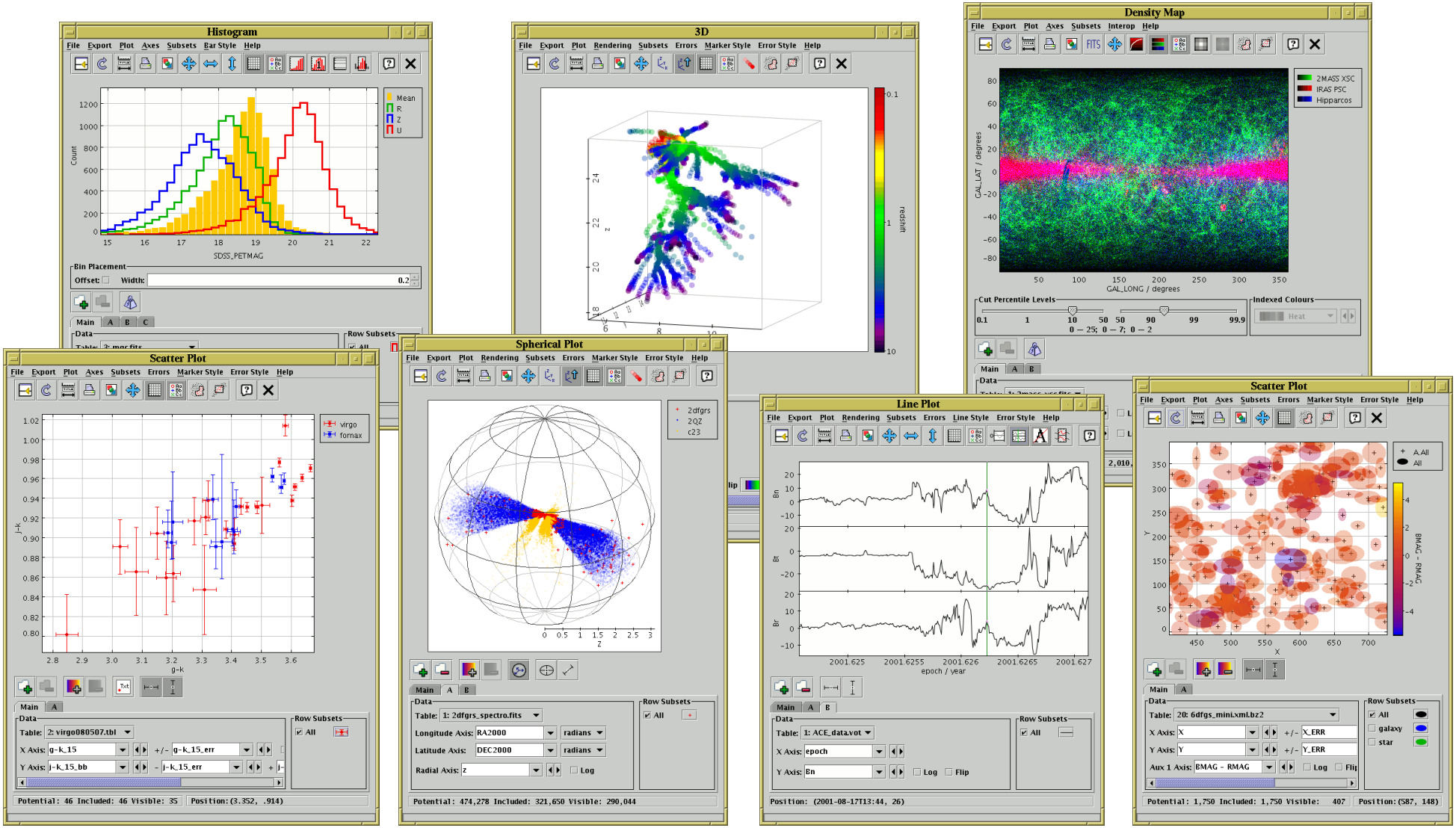
- ICRS, FK4, FK5, Galactic, Supergalactic, Ecliptic
- Degrees, radians, sexagesimal

TOPCAT Demo


TOPCAT
About TOPCAT
Scatter Plot
Density Map
Set Activation Action
3D
Spherical Plot
Match Tables
Cone Search
TOPCAT Help
Line Plot
Histogram
TOPCAT(3): Table columns
TOPCAT(1): Row Subsets
Plot Style Edit
Axis Configuration

Plotting

Many options. . .



STILTS Invocation

- Syntax:

```
stilts <task-name> <param1>=<value1> <param2>=<value2> ...
```

- Simple example:

```
stilts tcopy in=cat.ascii ifmt=ascii out=cat.fits ofmt=fits
```

- Complicated example:

```
stilts plot2d
    in1=iras_psc.fits cmd1='addskycoords fk5 galactic RA DEC GLON GLAT'
    xdata1=GLON ydata1=GLAT auxdata1=FNU_100 auxlog=true auxflip=true size1=0 transparency1=3
    in2=messier.xml cmd2='addskycoords fk5 galactic RA DEC GLON GLAT'
    xdata2=GLON ydata2=GLAT txtlabel2=RADIUS>16?("M"+ID):"") cmd2='addcol SIZE sqrt(RADIUS/2)'
    xerror2=SIZE yerror2=SIZE subset2a=true hide2a=true colour2a=black errstyle2a=ellipse
    subset2b=true hide2b=true colour2b=black errstyle2b=filled_ellipse transparency2b=6
    xlabel='Galactic Longitude' ylabel='Galactic Latitude' title='The Sky'
    legend=false grid=false fontsize=16 fontstyle=bold-italic
    xlo=0 xhi=360 ylo=-90 yhi=+90 xpix=800 ypix=400
    out=skyplot.eps
```

STILTS: Help!

- The manual (HTML or PDF) contains:
  - Tutorial chapters on syntax and invocation
  - Expression language reference
  - Reference section for each command:
    - Description of what the command does
    - Detailed description of each parameter and the values it can take
    - Several *examples* of the command in use
- Help available on command line from STILTS itself:
  - List of available tasks:

    ```
    stilts -help
    ```

  - Usage (list of parameters) for each task:

    ```
    stilts tpipe -help
    ```

  - Detailed description of each parameter:

    ```
    stilts tpipe help=ofmt
    ```

  - Interactive mode:

    ```
    % stilts tmatch2
    in1 - Location of first input table:
    ```

  - Hopefully useful error messages

STILTS Pipelines

Manipulate tables by stacking up "filters" in a "pipeline"

- Like Unix pipelines (`grep | awk | sed | sort | ...`)
- "Filters" available for almost any manipulation:

  `addcol`: add column
  `select`: select only rows satisfying given condition
  `sort`: sort on values in one or more columns
  `head`: retain only first *N* rows
  `badval`: replace bad values, e.g. -9999→NULL
  . . . and many more

- Use them in STILTS commands as values of `cmd` (or `icmd`, `ocmd`, . . .) parameters

```
stilts tpipe in=in.fits out=out.fits cmd=... cmd=...
```

- Most make use of TOPCAT/STILTS algebraic expression language
- Examples:
  - Add a new calculated column:
    `cmd='addcol B_R BMAG-RMAG'`
  - Select only rows in a given region:
    `cmd='select skyDistanceDegrees(RA,DEC,78.63,-8.20)<0.001'`
  - Select only ten reddest objects:
    `cmd='sort RMAG-BMAG' cmd='head 10'`

Useful STILTS Commands

- The most useful commands include:
  - General purpose pipeline: `tpipe`
  - Format conversion: `tcopy` (or `tpipe`)
  - Crossmatching: `tskymatch2`, `tmatch2`, `tmatch1`, `tmatchn`
  - Plotting: `plot2d`, `plot3d`, `plothist`
  - VOTable verification: `votlint`
  - Multiple cone search: `coneskymatch`

Tips

- Large datasets
  - TOPCAT: <= million rows, hundred columns
  - STILTS: unlimited size for some things
  - Try: `-Xmx256M`, `-disk`, use FITS, . . .
- STILTS quoting
  - Protect spaces from shell using single ' or double " quotes
  - can get a bit tricky :-(
- [How] Can TOPCAT do *XXX*?
  - Is it the sort of thing a table manipulation program *ought* to be able to do?
  - Look at menus, toolbars, online help button 
  - Try the manual or FAQ
  - Ask/email me