

## STILTS Plotting Tools

Mark B. Taylor

*H H Wills Physics Laboratory, University of Bristol, UK*

**Abstract.** The new plotting functionality in version 2 of STILTS is described. STILTS is a mature and powerful package for all kinds of table manipulation, and this version adds facilities for generating plots from one or more tables to its existing wide range of non-graphical capabilities. 2- and 3-dimensional scatter plots and 1-dimensional histograms may be generated using highly configurable style parameters. Features include multiple dataset overplotting, variable transparency, 1-, 2- or 3-dimensional symmetric or asymmetric error bars, higher dimensional visualization using color, and textual point labeling. Vector and bitmapped output formats are supported. The plotting options provide enough flexibility to perform meaningful visualization on datasets from a few points up to tens of millions. Arbitrarily large datasets can be plotted without heavy memory usage.

### 1. Introduction

Tabular datasets, especially in the form of object catalogs, are common in astronomy, and much of the data and processing in the virtual observatory in particular is dedicated to tables in various forms. STILTS<sup>1</sup> (Taylor 2006) is a mature package of command-line utilities which provides many kinds of table manipulation, including crossmatching, format conversion, statistics, row selection, column calculations and so on. It can read and write many important astronomical data formats (FITS, VOTable, CSV, SQL and others) and is written with efficiency and scalability in mind, so that tables of a few million rows, and for many purposes much larger, may be handled with modest resource usage. STILTS is in many ways the command-line counterpart of the GUI application TOPCAT<sup>2</sup>, with which it shares much overlapping functionality and implementation, but until now has lacked TOPCAT's graphical capabilities. This paper describes the newly added graphical features which have been introduced in STILTS version 2.0, released October 2008.

### 2. New Capabilities

At this release, three new STILTS commands for plotting have been introduced:

---

<sup>1</sup><http://www.starlink.ac.uk/stilts/>

<sup>2</sup><http://www.starlink.ac.uk/topcat/>

- `plthist`: 1-dimensional histogram
- `plot2d`: 2-dimensional scatter plot
- `plot3d`: 3-dimensional scatter plot

These commands share the structure and generic capabilities of existing STILTS commands, which means amongst other things that a wide range of input formats is accepted and extremely flexible pre-processing may be applied to the input data before the plotting itself takes place. It is easy for instance by supplying filtering commands upstream of the plot operation to transform coordinates, generate derived quantities, avoid or replace “bad” values, or restrict the plotted data based on arbitrarily complex selection criteria.

The plotted output is generated by the same implementation code as that used for TOPCAT’s plotting (considerable refactoring effort was required to achieve this), and so the appearance of the output and the kinds of options available will be familiar to users of that tool. The options and features include:

- Highly configurable styles for all kinds of plot
- Overplotting of multiple datasets
- Symmetric or asymmetric  $N$ -dimensional error bars/ellipses/cuboids...
- Higher dimensional visualization using colors
- Variable transparency of markers
- Text labeling of points with labels taken from table data
- Choice of output formats (EPS, PNG, GIF, JPEG or to screen)

STILTS also offers better configurability than TOPCAT for certain more cosmetic plot attributes such as font size and style, marker colors, plot labeling and so on. Along with its scriptability this may make it more suitable for generating plots for publication or external presentation.

Furthermore, the basic model of data access in STILTS is streaming, which means that an arbitrarily large number of input rows and/or output points can be processed without heavy memory usage. For such large plots the ability to assign partial transparency to points so that overplotting density is visible in the output is invaluable. Density maps like TOPCAT’s and contour drawing may be introduced in a future release as alternative ways to render such overdense plots meaningful.

Some example plots are shown in Figure 1.

### 3. Invocation

As for other STILTS commands, invocation is using a series of `name=value` parameter specifications. A color-magnitude diagram for a catalog with columns RMAG and BMAG may be generated with a command as simple as:

```
stilts plot2d in=cat.fits xdata=RMAG-BMAG ydata=BMAG out=cm.eps
```

The value of the `xdata` parameter here is given by an expression in STILTS’s powerful and extensible expression language, which uses column names as variables and contains many astronomical and general functions as well as arithmetic operators. The output in this example is to a named Encapsulated PostScript file. If the `out` parameter is omitted, the plot is displayed in a window on screen for convenience.

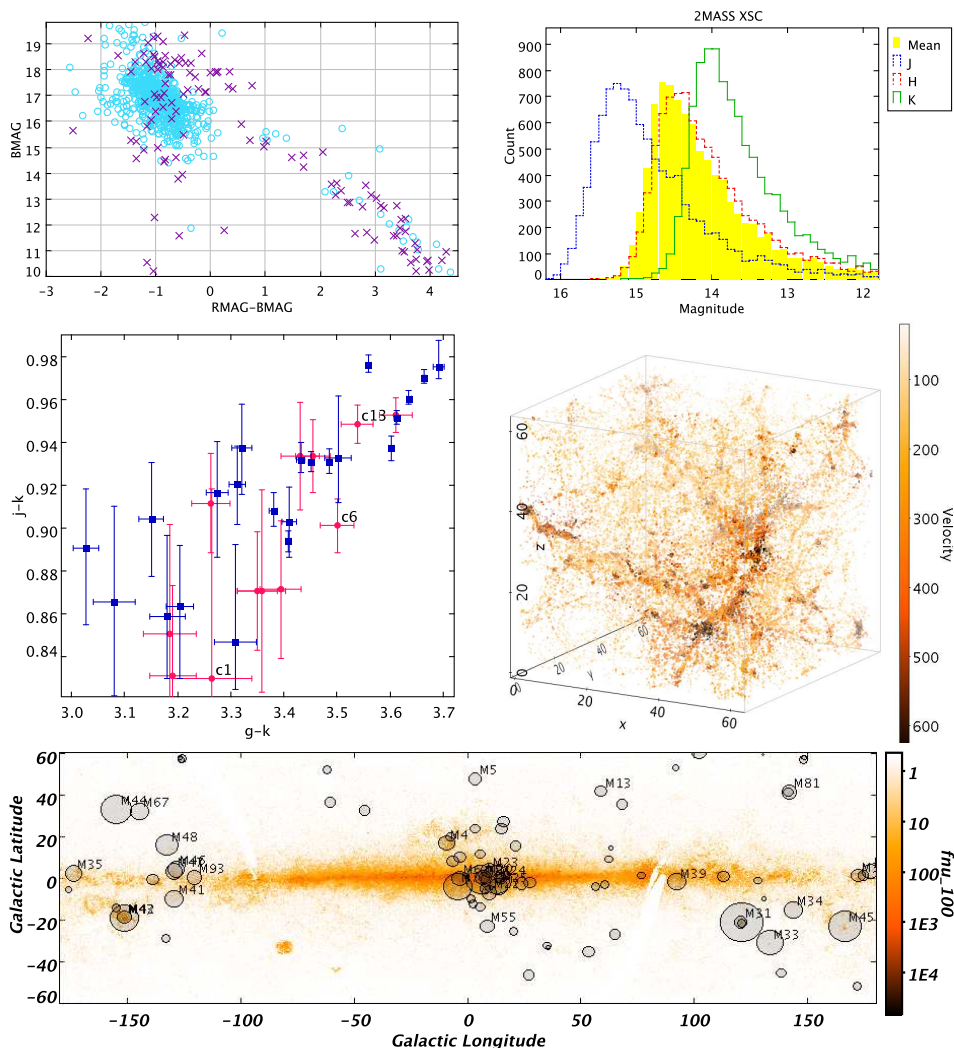


Figure 1. Some example plots generated by STILTS.

`plot2d` has more than 40 parameters, so invocations can get more complicated than the example above. Although the wealth of available options may seem daunting at first, care has been taken to provide full documentation for each one in the manual (available in HTML and PDF format) and also from the command line. “`stilts plot2d help`” lists all the parameters for the command, while “`stilts plot2d help=xdata`” prints help on the named parameter. Error messages try to be as helpful as possible about what went wrong. The manual also provides tutorial material and examples of invoking each command.

#### 4. Use Cases

There are many other packages which can generate plots from the command line; some which have been popular with astronomers are IDL, gnuplot and

PONGO. STILTS, like the others, has its own particular strengths. It can work with, and provide meaningful visualization of, very large datasets (though small ones are equally well catered for). It can plot data directly from astronomy-friendly file formats such as FITS and VOTable as well as from less domain-specific sources such as CSV files or direct interface with relational databases. Since it is written in pure Java and licensed under the Gnu GPL, there are no problematic installation or licensing issues. Finally, users of TOPCAT will find STILTS input, output and way of doing things familiar.

For plotting as well as for the other capabilities common to STILTS and TOPCAT, it can be very productive to use these two tools in a complementary fashion. In an initial exploratory phase the data can be investigated using the highly interactive and responsive TOPCAT to understand the kind of analysis which needs to be done, and in a subsequent production phase a repeatable STILTS command or script can be prepared to apply that analysis to multiple similar datasets.

Another environment in which this software may find use is for server-side plot generation. By co-locating STILTS with server-held data or computation, plots can be generated on the fly and served to HTTP clients as an alternative to, or as well as, raw table data. A new server/servlet mode has been introduced alongside the plotting capabilities to facilitate this option. Trial deployments along these lines have already been made by Thomas Boch (CDS) and Marco Molinaro (INAF/OATs).

## 5. Future Work

The initial release of STILTS plotting is fully functional but some further work on it is desirable. One item is to include the remaining TOPCAT plot types not yet provided, namely the spherical 3D plot, density map and stacked line plot. Another is to improve STILTS/TOPCAT interoperability, probably by introducing a TOPCAT feature to report the STILTS command line corresponding to a currently visible plot; this would make it easier for users to get started with command-line plotting, giving them a template which could be fine-tuned by hand. There may also be enhancements to the underlying plotting library which would be available from both STILTS and TOPCAT, for instance contour drawing.

Decisions about the priority for these and other possible upgrades to the software will be taken in the light of user take-up and feedback.

## References

Taylor, M. B. 2006, in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 666