TOPCAT and Gaia

Mark Taylor
University of BRISTOL

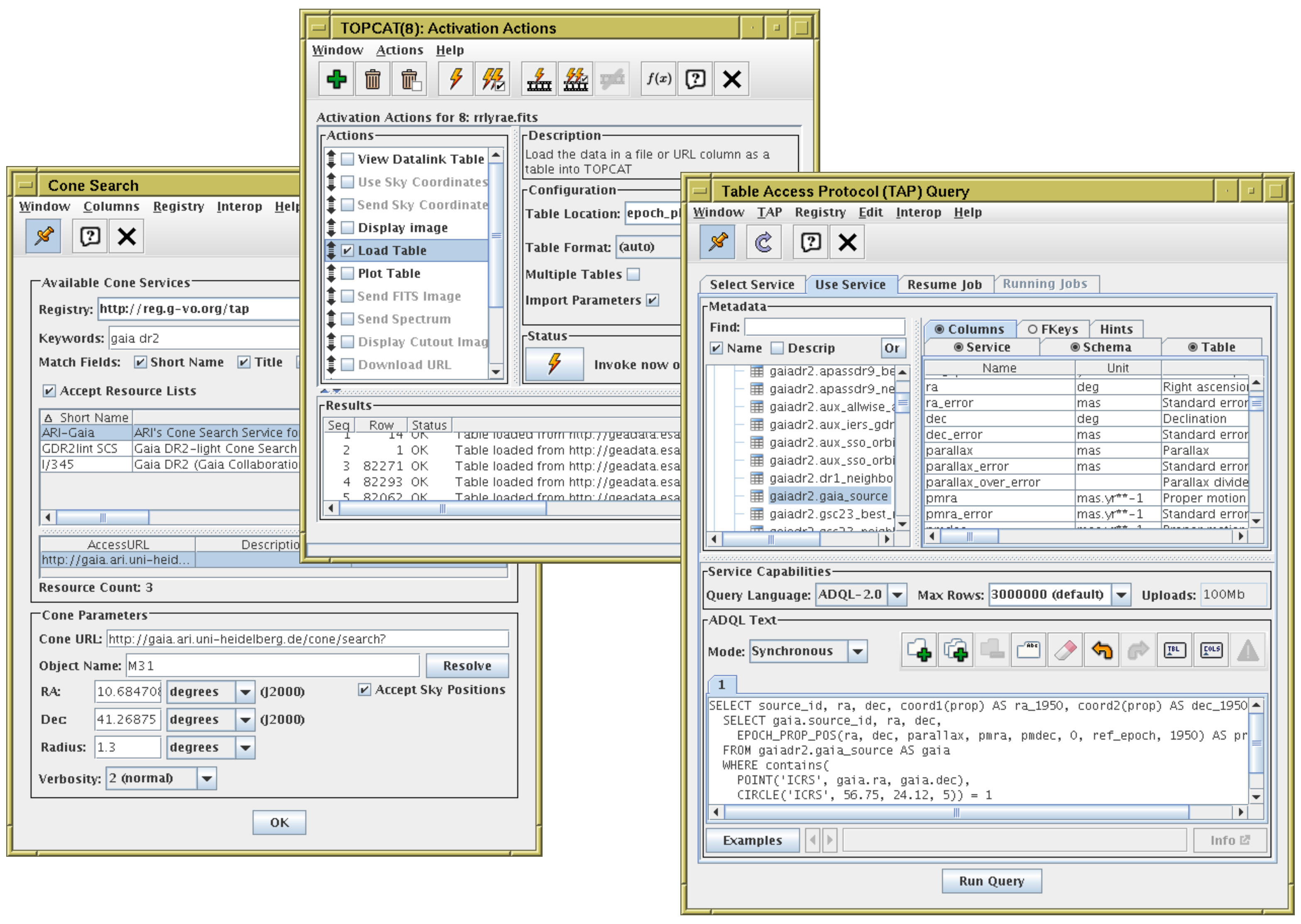
Abstract

TOPCAT, and its command line counterpart STILTS, are powerful tools for working with large source catalogues. ESA's *Gaia* mission, most recently with its second data release, is producing source catalogues of unprecedented quality for more than a billion sources. We present here some examples of how TOPCAT and STILTS can be used for analysis of Gaia data.

Data Access

The primary access to the Gaia catalogue is via Virtual Observatory protocols, provided by the GACS system at ESAC, and other data centers elsewhere. TOPCAT provides user-friendly GUI interfaces to the relevant protocols, including

- TAP (submit SQL-like queries to the database),
- Cone Search (retrieve all sources round a given sky position)
- DataLink (request e.g. time series for one or more sources)



TOPCAT(8): Activation Actions — Window Actions Help

Activation Actions for 8: rrlyrae.fits

Actions: □ View Datalink Table, □ Use Sky Coordinates, □ Send Sky Coordinate, □ Display image, ☑ Load Table, □ Plot Table, □ Send FITS Image, □ Send Spectrum, □ Display Cutout Imag, □ Download URL

Description: Load the data in a file or URL column as a table into TOPCAT

Configuration: Table Location: epoch_pl; Table Format: (auto); Multiple Tables □; Import Parameters ☑

Status: Invoke now o

Results:

Seq	Row	Status	
1	14	OK	Table loaded from http://geadata.esa
2	1	OK	Table loaded from http://geadata.esa
3	82271	OK	Table loaded from http://geadata.esa
4	82293	OK	Table loaded from http://geadata.esa
5	82062	OK	Table loaded from http://geadata.esa

Cone Search — Window Columns Registry Interop Help

Available Cone Services: Registry: http://reg.g-vo.org/tap; Keywords: gaia dr2; Match Fields: ☑ Short Name ☑ Title; ☑ Accept Resource Lists

Short Name	
ARI-Gaia	ARI's Cone Search Service fo
GDR2lint SCS	Gaia DR2-light Cone Search
I/345	Gaia DR2 (Gaia Collaboratio

AccessURL	Descriptio
http://gaia.ari.uni-heid...	

Resource Count: 3

Cone Parameters: Cone URL: http://gaia.ari.uni-heidelberg.de/cone/search?; Object Name: M31 (Resolve); RA: 10.684708 degrees (J2000); Dec: 41.26875 degrees (J2000); ☑ Accept Sky Positions; Radius: 1.3 degrees; Verbosity: 2 (normal); OK

Table Access Protocol (TAP) Query — Window TAP Registry Edit Interop Help

Select Service | Use Service | Resume Job | Running Jobs

Metadata: Find: ; ☑ Name □ Descrip Or; ● Columns ○ FKeys Hints; ● Service ● Schema ● Table

gaiadr2.apassdr9_be, gaiadr2.apassdr9_ne, gaiadr2.aux_allwise_, gaiadr2.aux_iers_gdr, gaiadr2.aux_sso_orbi, gaiadr2.aux_sso_orbi, gaiadr2.dr1_neighbo, gaiadr2.gaia_source, gaiadr2.gsc23_best_

Name	Unit	
ra	deg	Right ascensio
ra_error	mas	Standard error
dec	deg	Declination
dec_error	mas	Standard error
parallax	mas	Parallax
parallax_error	mas	Standard error
parallax_over_error		Parallax divide
pmra	mas.yr**-1	Proper motion
pmra_error	mas.yr**-1	Standard error

Service Capabilities: Query Language: ADQL-2.0; Max Rows: 3000000 (default); Uploads: 100Mb

ADQL Text: Mode: Synchronous

```
SELECT source_id, ra, dec, coord1(prop) AS ra_1950, coord2(prop) AS dec_1950
  SELECT gaia.source_id, ra, dec,
    EPOCH_PROP_POS(ra, dec, parallax, pmra, pmdec, 0, ref_epoch, 1950) AS pr
  FROM gaiadr2.gaia_source AS gaia
  WHERE contains(
    POINT('ICRS', gaia.ra, gaia.dec),
    CIRCLE('ICRS', 56.75, 24.12, 5)) = 1
```

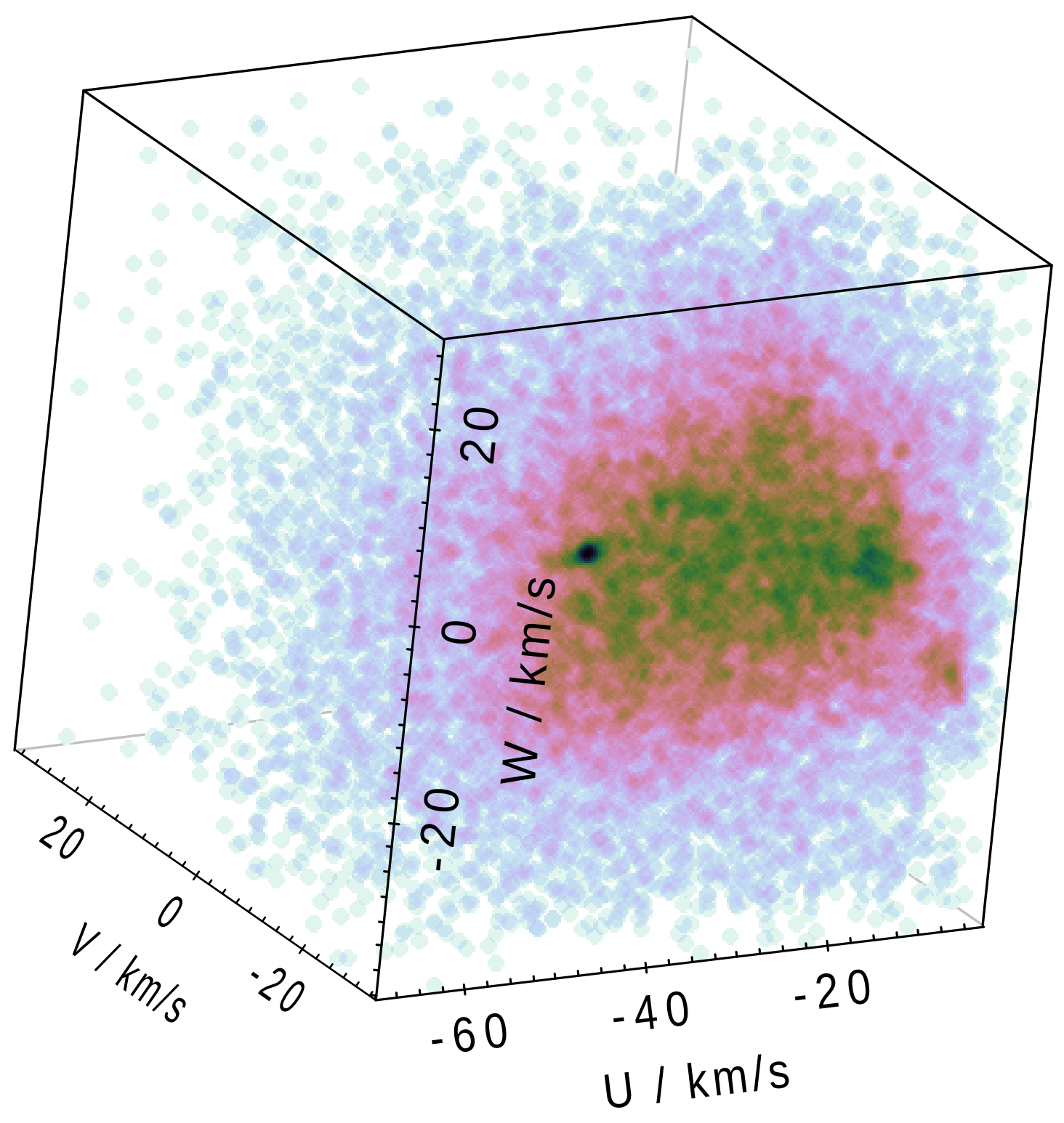
Examples | Info | Run Query

TOPCAT can also read files in the GBIN format used internally within the Gaia data processing consortium.

Expression Language

TOPCAT provides a powerful language for evaluating algebraic expressions to define new columns, row selections, and plot coordinates. As well as standard arithmetic, trigonometric and astronomical operations, the function library contains a number of Gaia-friendly functions:

- Astrometry propagation to earlier/later epoch
  (functions `epochProp`, `epochPropErr`)
- Astrometric to Cartesian position/velocity conversion
  (functions `polarXYZ`, `astromXYZ`, `astromUVW`, `galToIcrs`, `eclToIcrs`, `rvKmsToMasyr`, ...)
- Bayesian estimation of distance from parallax (following Bailer-Jones et al.)
  (functions `distanceEstimateEdsd`, `distanceBoundsEdsd`, `distanceQuantilesEdsd`)



This figure shows nearby stars for which DR2 contains radial velocity (`radial_velocity IS NOT NULL`, `parallax > 10`) in a region of galactic U, V, W (velocity) space. The small dense blob near the center is the Hyades. The coordinates were calculated with the expression:

`icrsToGal(astromUVW(ra, dec, pmra, pmdec, radial_velocity, 1000./parallax, false))`

Scalability

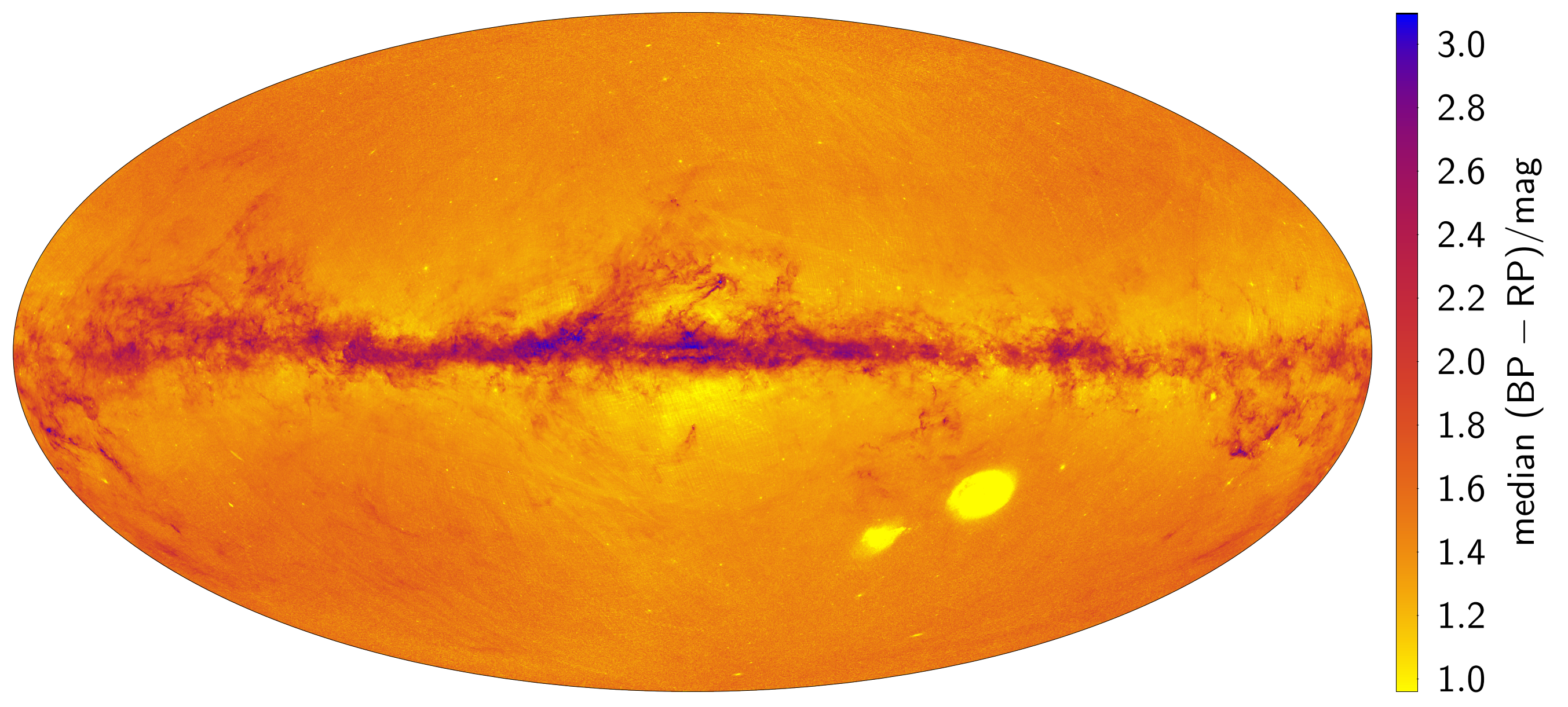
Gaia DR2 contains 1.7 billion sources, and investigating this dataset often requires working with large tables.

TOPCAT provides easy interactive analysis, including flexible exploratory visualisation, for order 10^6–10^7 rows, on modest hardware with no special data preparation (e.g. data downloaded from external services, FITS files, even CSV). For much larger tables, interactive performance will degrade and memory may be exhausted.

STILTS (TOPCAT's command-line counterpart) works in streaming mode, so can cope with arbitrarily large tables, most tasks scaling as $O(N^1)$ in processing time and $O(N^0)$ in memory for N rows. The image below was produced as follows:

1. download 61 000 gzipped CSV files (0.5 Tb) from ESAC (`wget`, 10 hours)
2. convert to 61 000 small FITS files (STILTS `tpipe`, 5 days)
3. convert to single 0.8 Tb column-oriented FITS file (STILTS `tpipe`, 12 hours)
4. aggregate into level-9 HEALPix map (STILTS `tskymap`, 45 minutes)
5. render plot to PDF or PNG (STILTS `plot2sky`, a few seconds)

In practice, steps 4 and 5 were iterated using TOPCAT visualisation interactively on smaller datasets to achieve the best results.

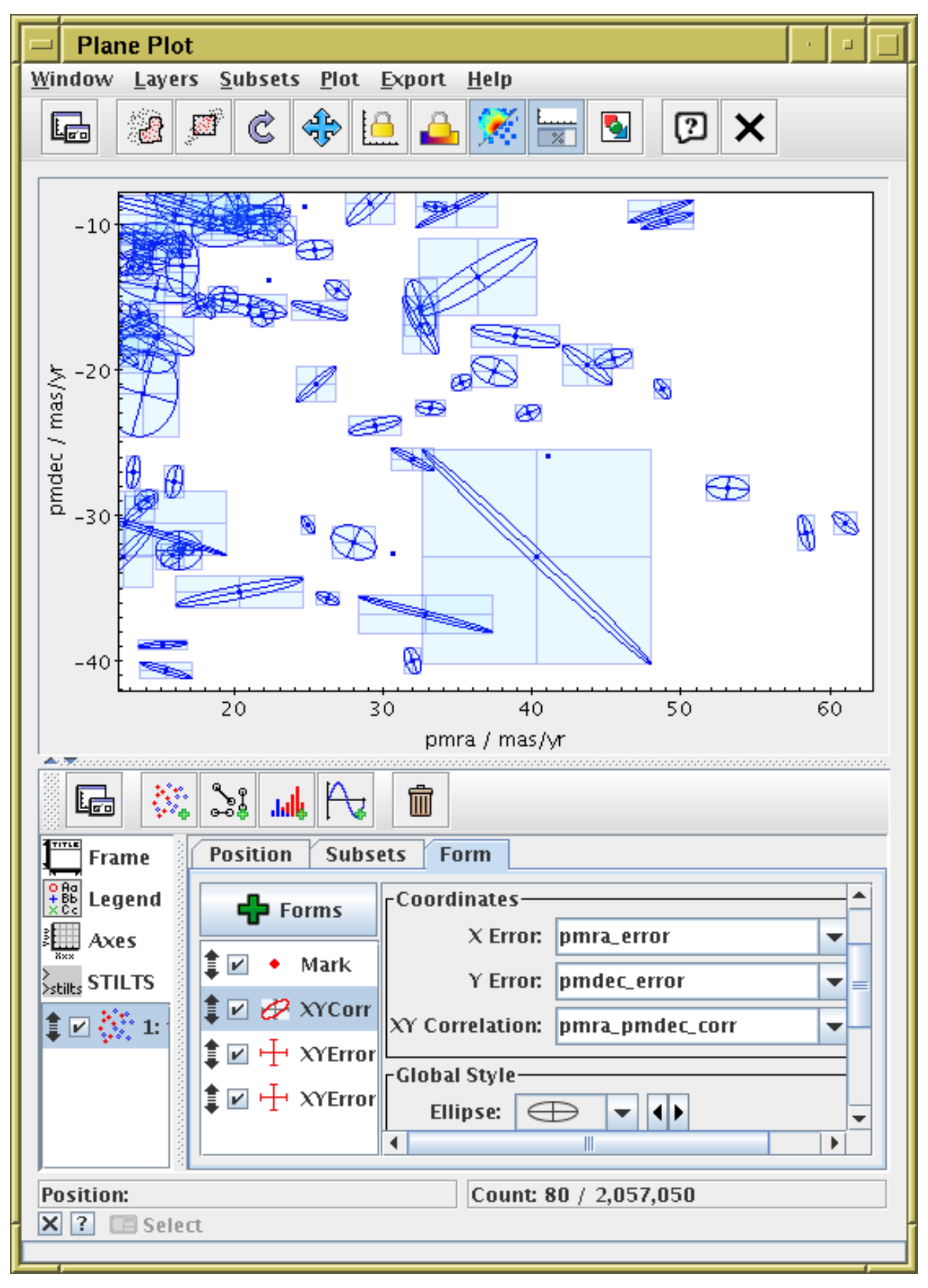


Computation near the data is usually a better way! But STILTS can handle the volume.

Visualisation

TOPCAT has many visualisation modes enabling highly configurable interactive exploration of high-dimensional data, suitable for both large and small data sets. Special attention is given to providing comprehensible representations of very large data sets — a simple scatter plot is not useful when there are many more points than pixels.

Plots specifically intended for Gaia data include the Sky/XY Correlation and Sky Vector layers:



Plane Plot — Window Layers Subsets Plot Export Help

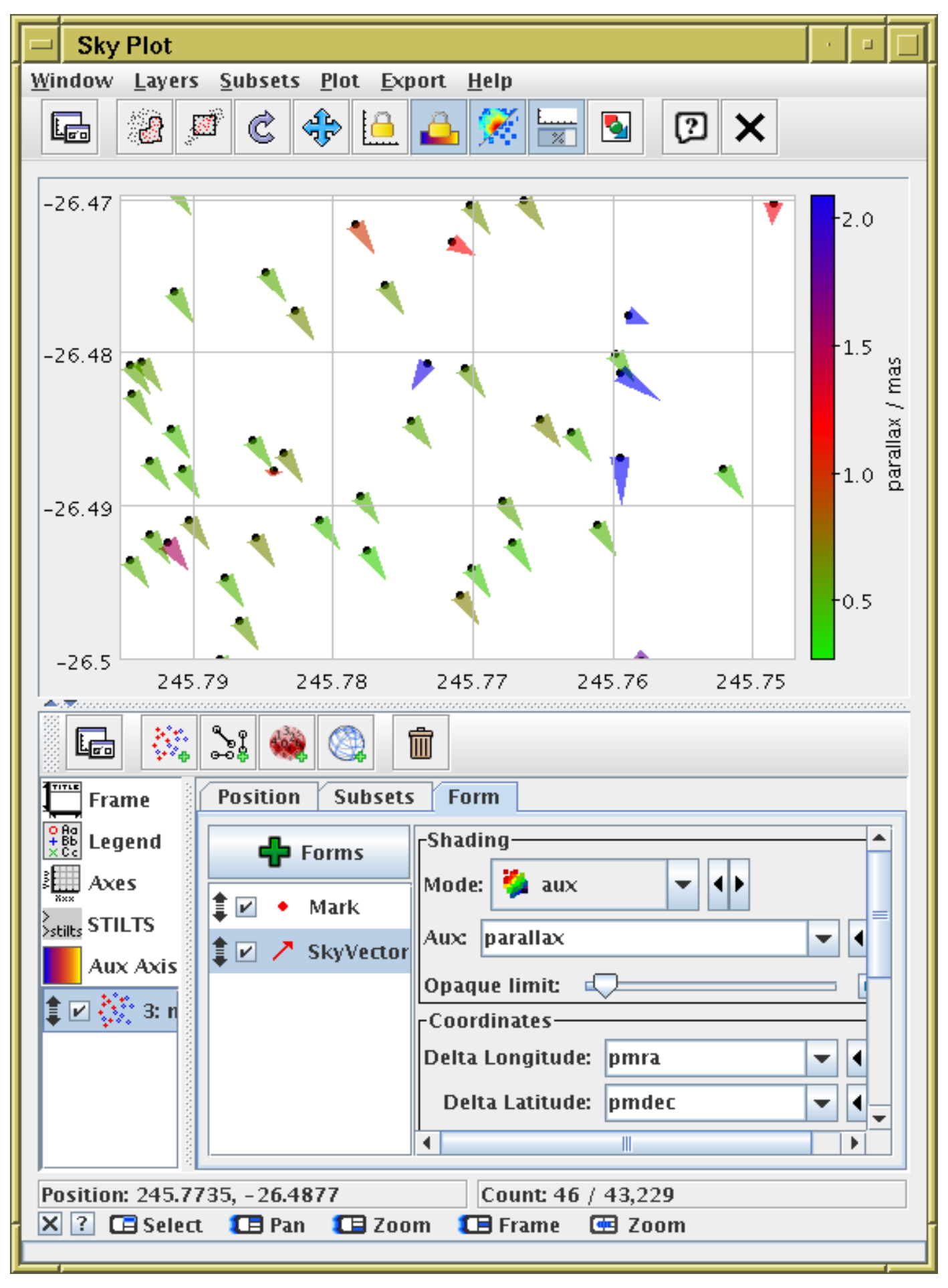
Frame, Legend, Axes, STILTS; Position | Subsets | Form; Forms: ☑ Mark, ☑ XYCorr, ☑ XYError, ☑ XYError

Coordinates: X Error: pmra_error; Y Error: pmdec_error; XY Correlation: pmra_pmdec_corr

Global Style: Ellipse:

Position: Count: 80 / 2,057,050

The XY Correlation plot shows proper motion error ellipses derived from the Gaia `pmra_pmdec_corr` column; the ellipses give much more information than the simple `pmra_error`/`pmdec_error` error boxes.



Sky Plot — Window Layers Subsets Plot Export Help

Frame, Legend, Axes, STILTS, Aux Axis; Position | Subsets | Form; Forms: ☑ Mark, ☑ SkyVector

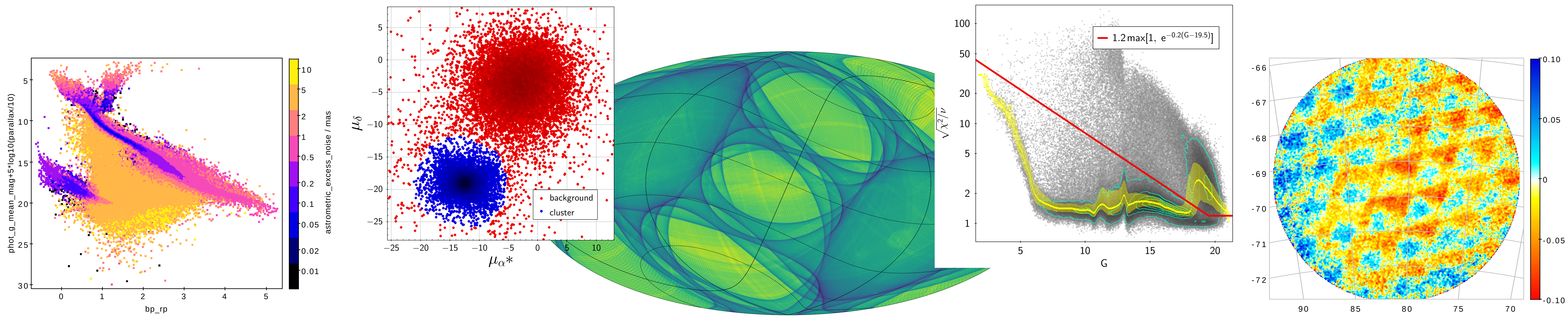
Shading: Mode: aux; Aux: parallax; Opaque limit:

Coordinates: Delta Longitude: pmra; Delta Latitude: pmdec

Position: 245.7735, -26.4877 Count: 46 / 43,229

This Sky Vector plot displays proper motion vectors by shape, and parallax by colour.

There are very many non-Gaia-specific visualisation features useful for Gaia data, including 1-, 2- and 3-dimensional scatter plots and histograms, colour coding by density or aggregated values, sky plots, contours, quantiles, line and function plotting, error bars/ellipses, and more. Some examples are shown below.



Acknowledgements

This work has been primarily funded by the UK's Science and Technology Facilities Council. It has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`).
Special thanks to the EU Horizon 2020 project ASTERICS for funding presentation of this work at ADASS 2018.

Do It Yourself!

Most of the figures on this poster were generated using STILTS with data acquired from Gaia VO services. A makefile and source files containing everything necessary to obtain the data and build the poster are available from `https://github.com/mbtaylor/adass2018-tcgaia`.