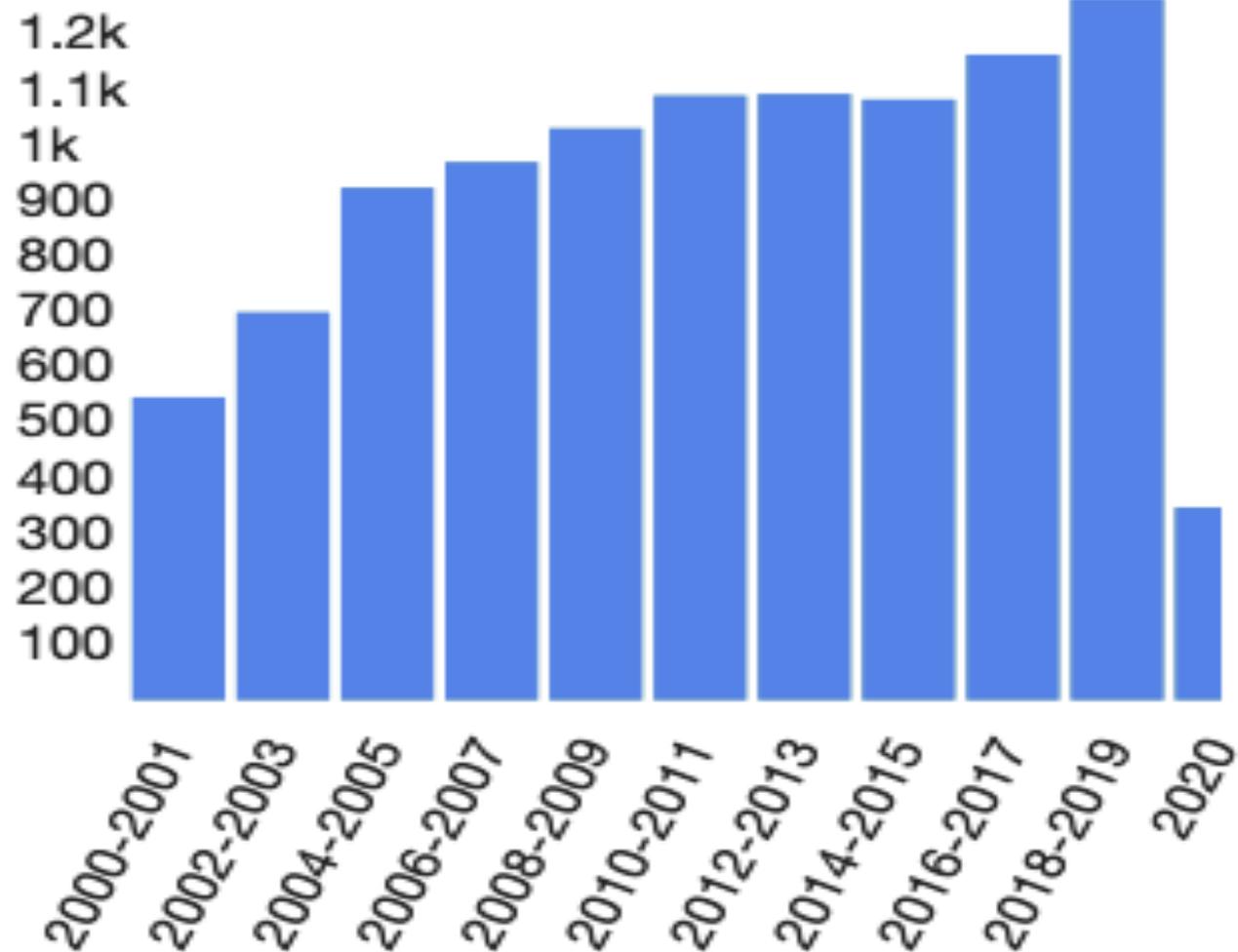# Statistical Issues in Combining High and Low Resolution Data

Keith Arnaud

Astronomy Dept., University of Maryland
HEASARC, NASA Goddard Space Flight Center
CRESST

Refereed papers including the word xspec (ADS)

Total papers :	11.052	
Citations of these papers :	411,186	
Collective h-index:	208	

XSPEC v12.11 and PyXspec v2.0.3 were released with HEAsoft 6.27 in March.

We try to do one new XSPEC release per year.

Additional models sent to us are made available and if requested are added as standard models in the next release (https://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/newmodels.html)

Note that HEAsoft now includes gsl so any models can use these routines.

Please do not use Numerical Recipes routines in models since these are copyrighted and are not supposed to be distributed.

HEAsoft 6.28 will be released in a few weeks. This will include the current patch level versions of XSPEC and PyXspec.

(A word of warning for Mac users. The new Xcode 12 which supports Apple Silicon looks like it is going to require a lot of changes to our code)

If you are not already a member I recommend joining the xspec Facebook group which now has ~940 members (https://www.facebook.com/groups/320119452570/).

The latest release includes a new way of handling very large response arrays.

For instance, the full XRISM/Resolve matrix has 60,000 energy bins and 60,000 channels. Since calorimeters have response down to zero energy this matrix is triangular and has a size over 7 Gb.

However, most of the response (the long tail to low energies) is smooth so doesn't need so many energy bins.

Solution is to split the response in two, e.g.:
    60,000 energies by 60,000 channels – a sparse matrix with just high resolution features
   1,000 energies by 60,000 channels – a triangular matrix with the long tail

The model is calculated on 60,000 energy bins, a copy is binned up to 1,000 energies, they are multiplied by the appropriate matrices and the results summed.

This gives order of magnitude speed-ups with no loss of accuracy.

This is now all handled automatically by xspec – the response file has one EBOUNDS and two MATRIX extensions.

In principle this could be extended to bin on channels as well to make the triangular, low-resolution matrix even smaller.

Jeremy Sanders
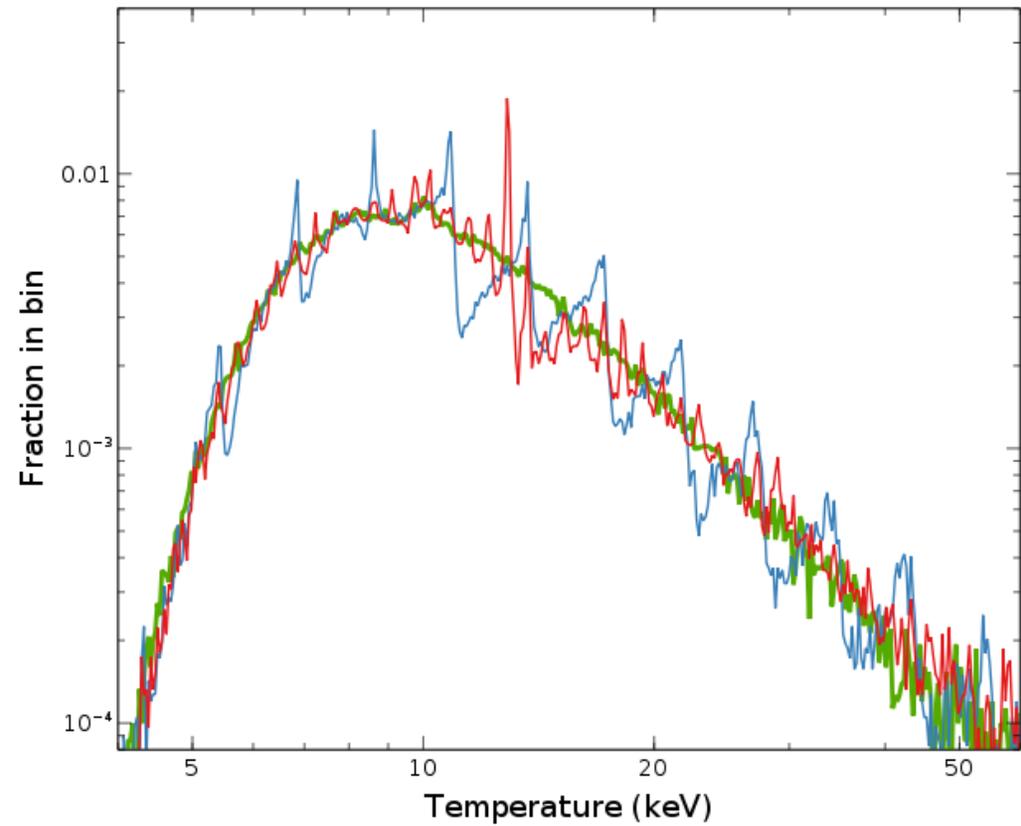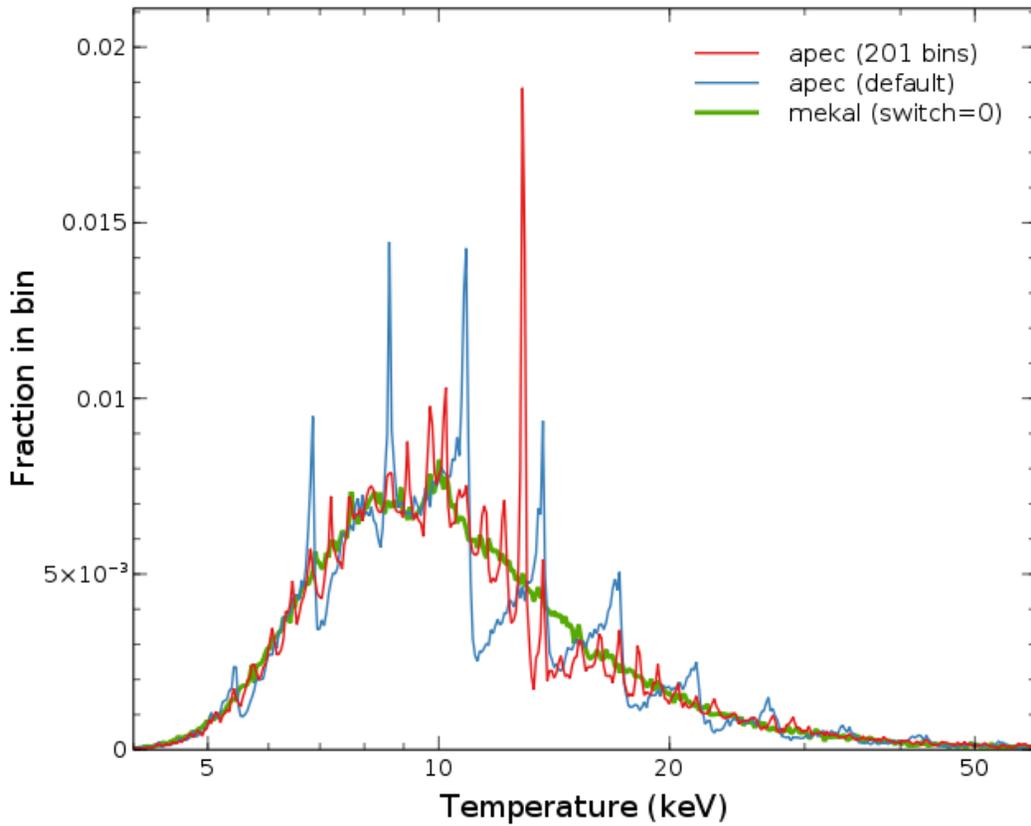
Fits to simulations of 10 keV apec model using eRosita response

Combining high res and low res data:
A Toy Case

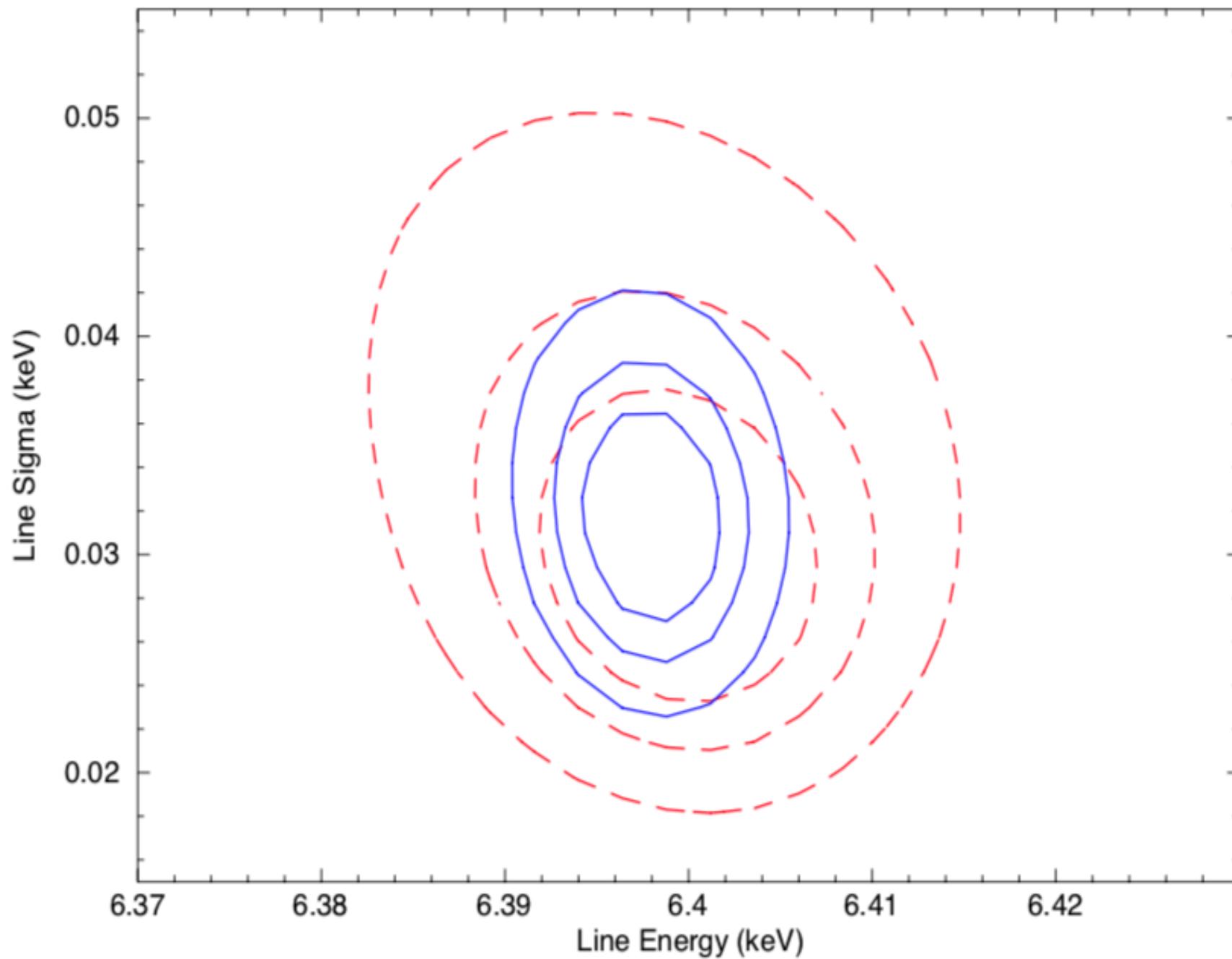High resolution detector with FWHM 5 eV

Low resolution detector with FWHM 100 eV and 10x the effective area of high resolution detector

Single gaussian line and power-law background.

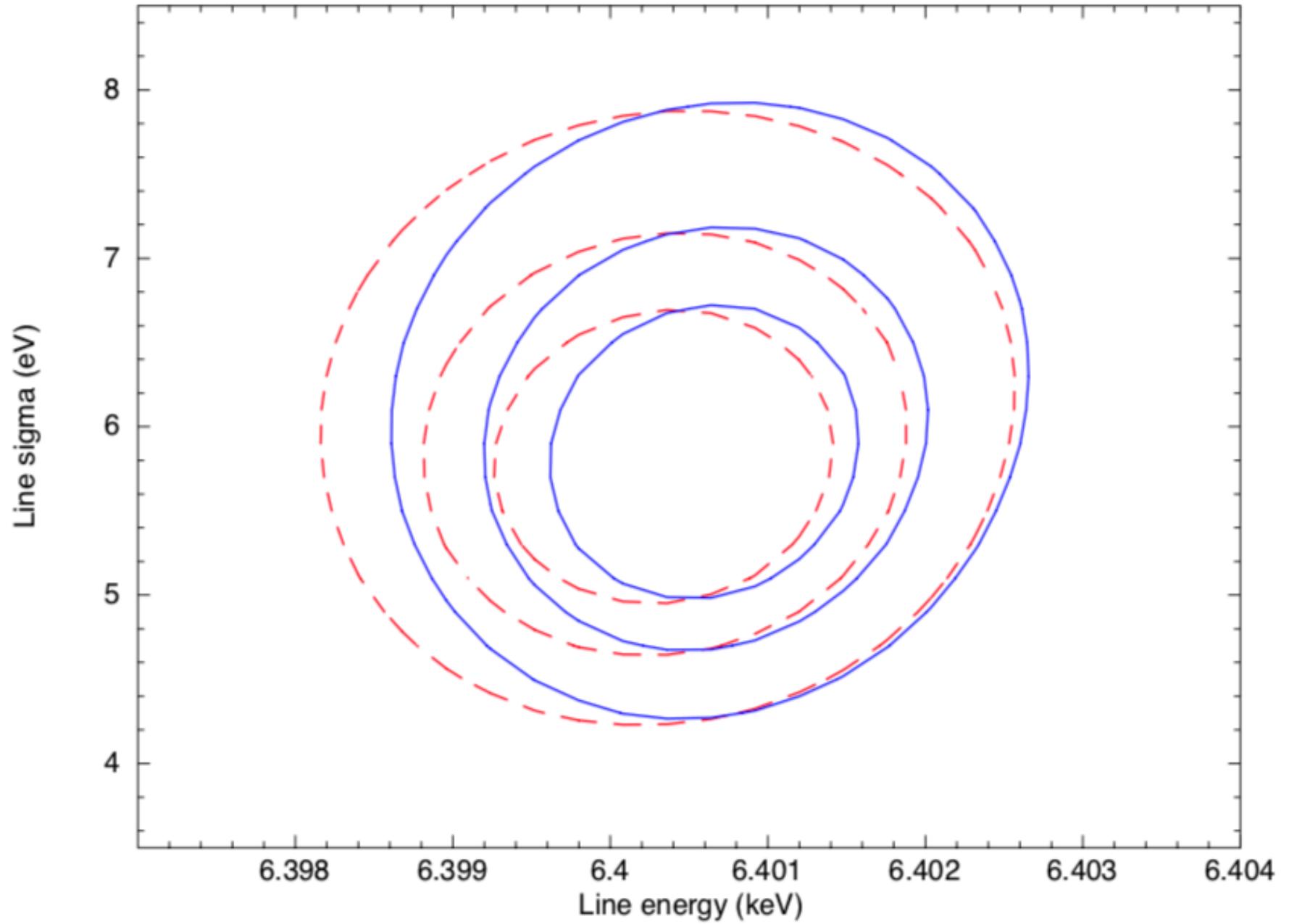Look at cases of 30 eV sigma line and 5 eV sigma line.

30 eV sigma line

Confidence contours
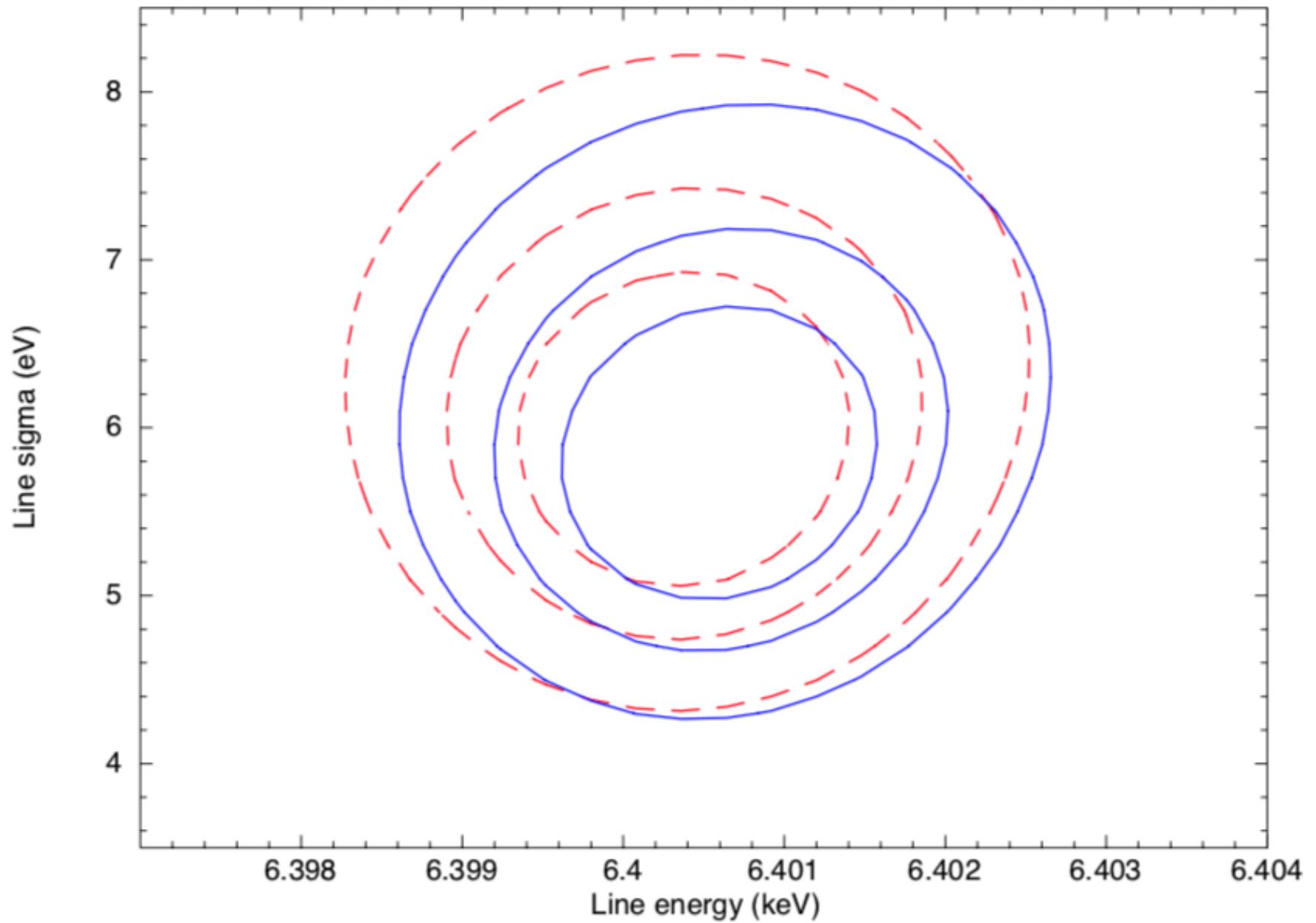High resolution only and both

Line Sigma (keV)

Line Energy (keV)

5 eV sigma line

Confidence contours
High resolution only and both

Line sigma (eV)

Line energy (keV)

5 eV sigma line

Confidence contours
With and without optimal binning

Line sigma (eV)

Line energy (keV)

5 eV sigma line

High resolution only and both

Δ C-statistic

Line normalization
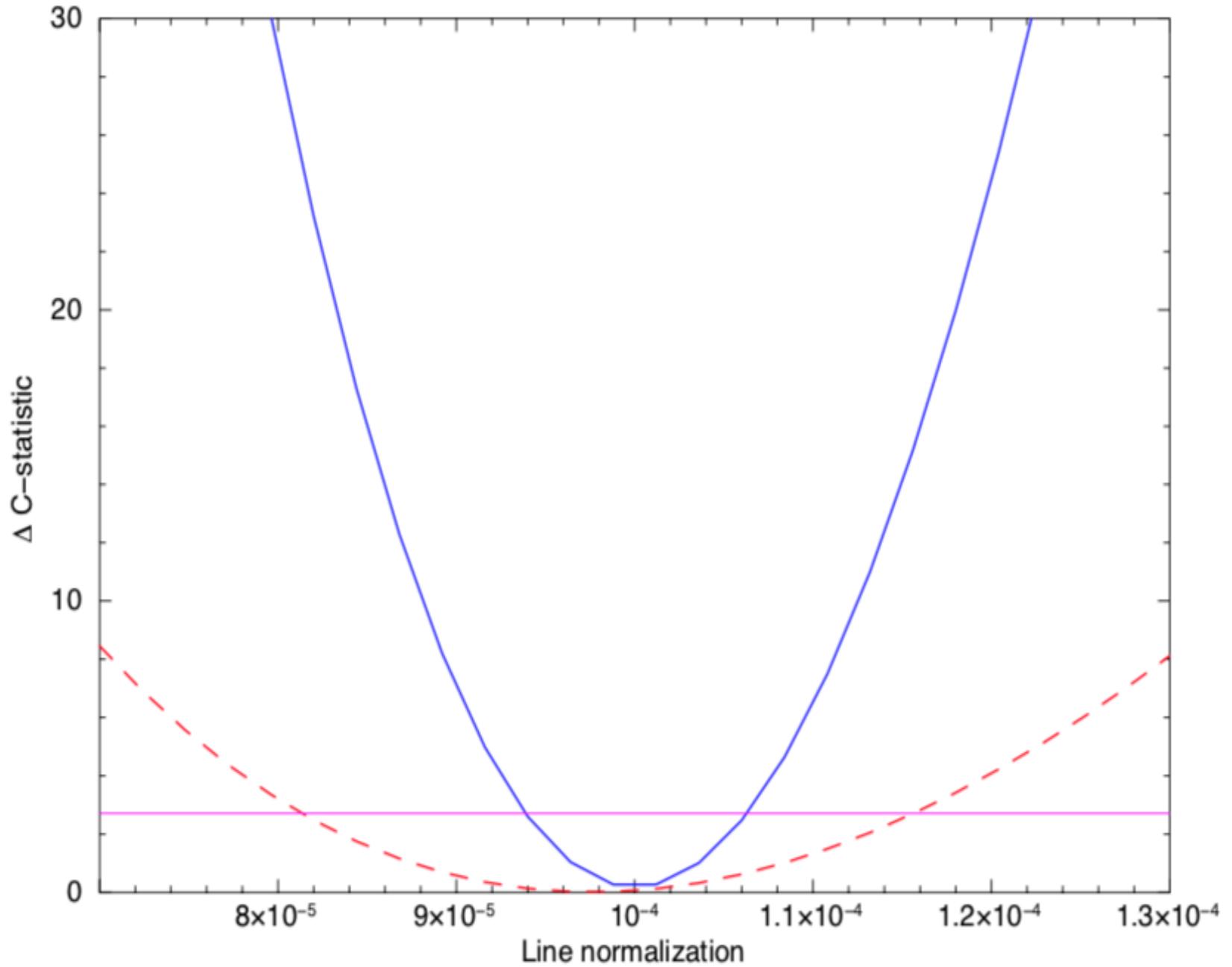
Combining high res and low res data:
# What could possibly go wrong?

If the model is correct the greater number of counts in the low res data could drive the fit to a local minimum which doesn't work for the high res data.   -   Probably best to fit the high res data first.

If the model is incorrect then the fit may end up in a global minimum which doesn't work for the high res data. Find a better model.

If the low res data has calibration problems then the fit may not work for the high res data and the parameters may be wrong.
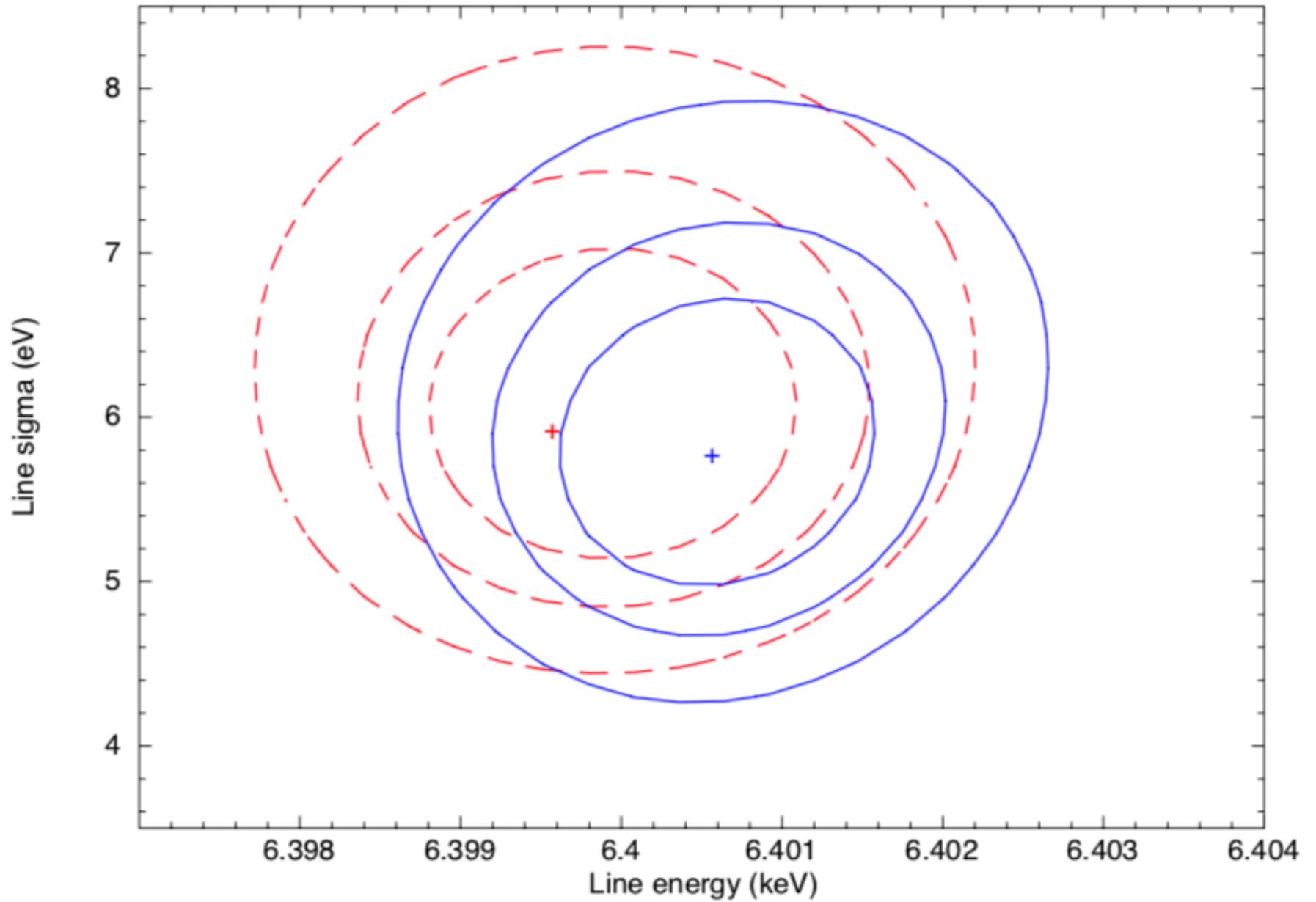
30 eV sigma line

Confidence contours
Original and with 0.1% gain shift for low resolution
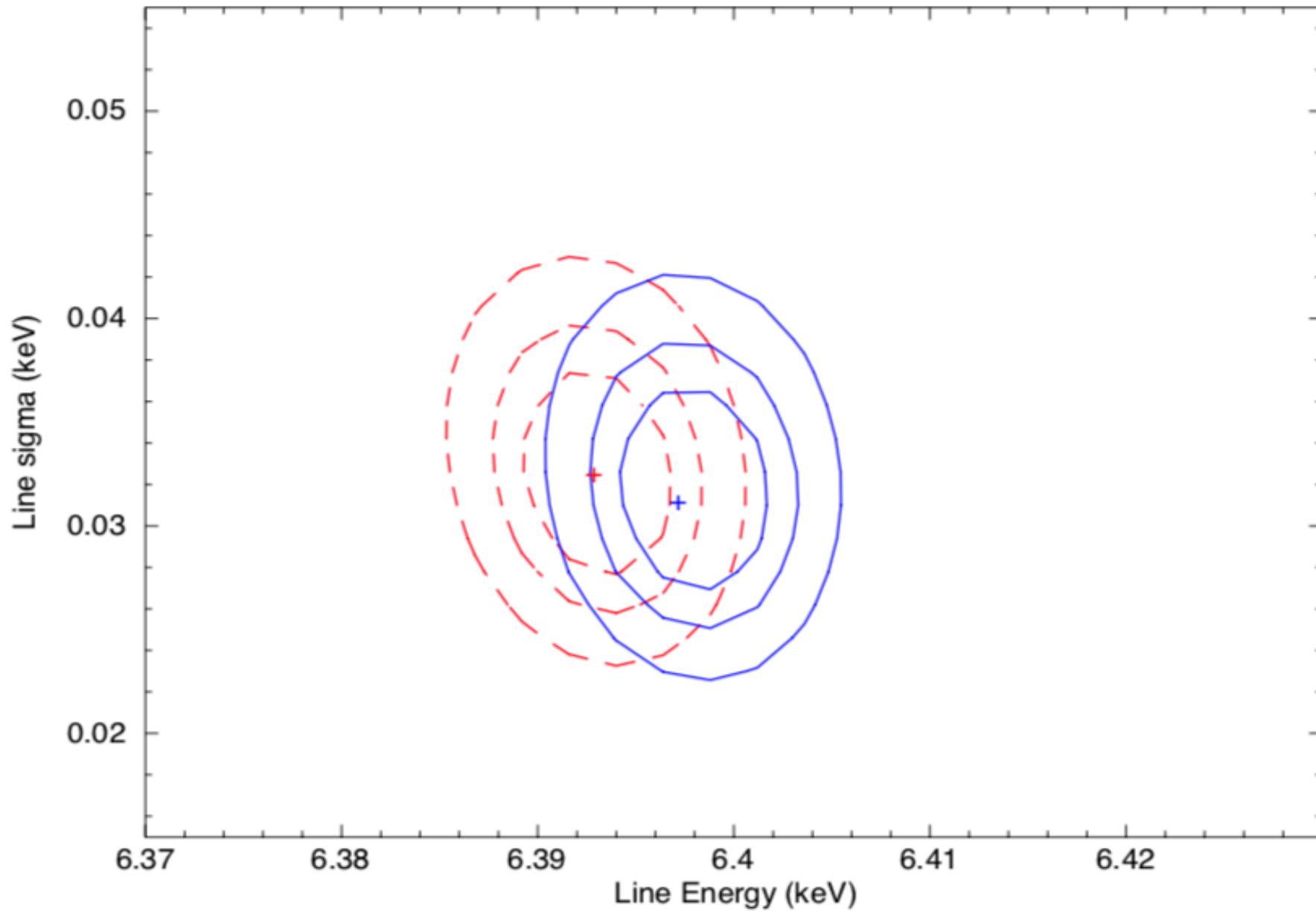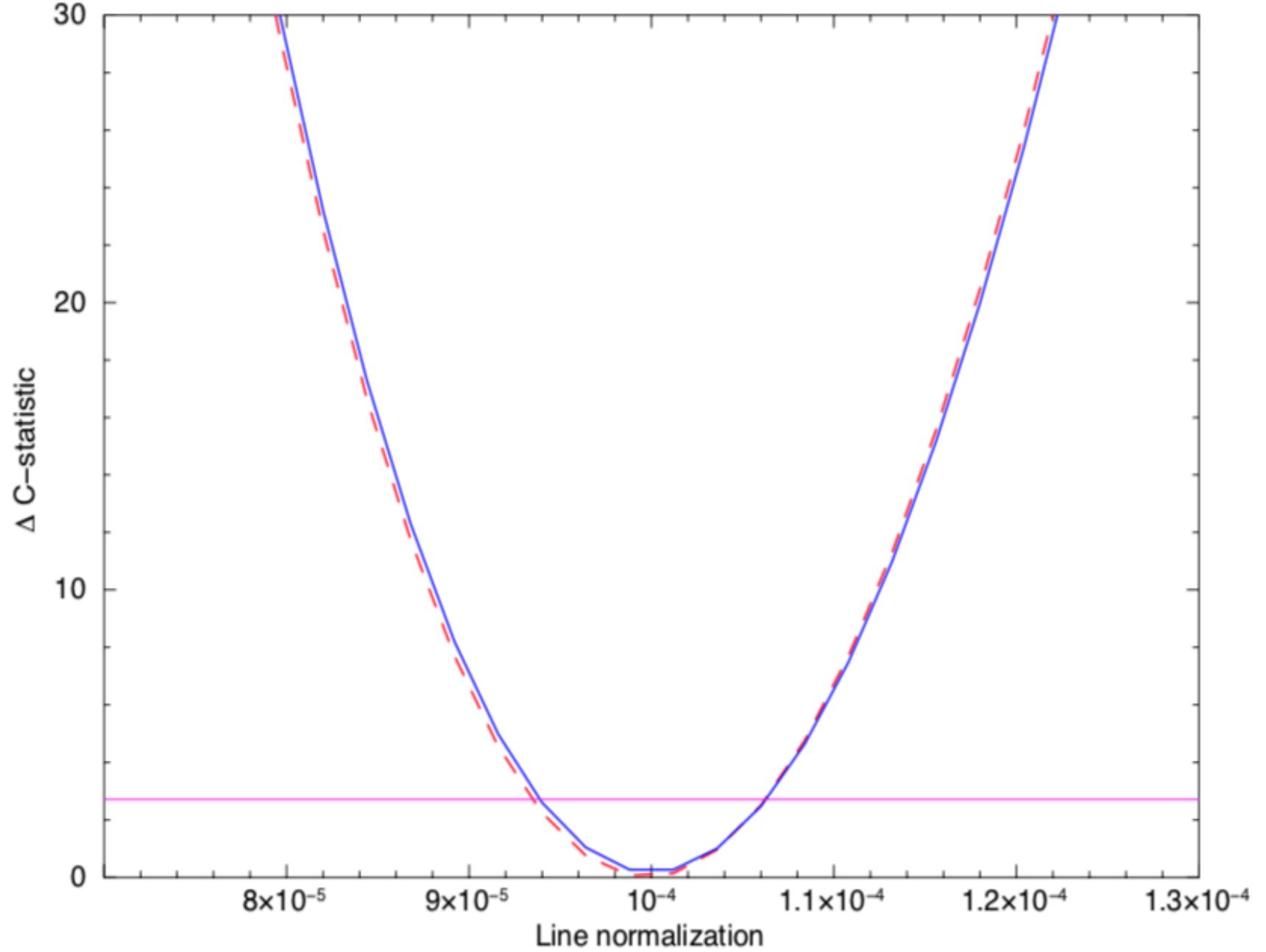
Line sigma (keV)

Line Energy (keV)

5 eV sigma line

Confidence contours
Original and with 0.1% gain shift for low resolution

Line sigma (eV)

Line energy (keV)

5 eV sigma line

Original and with 0.1% gain shift for low resolution

HEASP and HEASPTOOLS

heasp is a C++ library of routines to manipulate spectra, responses, table models.

A python module called heasp provides an interface to these C++ routines.

(https://heasarc.gsfc.nasa.gov/docs/software/lheasoft/headas/heasp/heasp_guide.html)

heasptools are a set of ftools built on top of heasp and will replace many old tools in caltools and heasarc.

For instance, ftrbnpha replaces rbnpha. ftgrouppha is a replacement for grppha with many more grouping options including optimal binning.

We are slowly replacing the old tools by scripts running the new tools. Please use the new versions and let us know of any problems.

(https://heasarc.gsfc.nasa.gov/lheasoft/ftools/headas/heasptools.html)

Some Changes in XSPEC v12.11

There is an issue with the various CEI models dependent on AtomDB due to the temperature binning. To mitigate this the AtomDB team have made new files using 201 temperatures instead of 51. The Xspec.init file has been changed to use these new files as the default. There are no significant speed impacts.

See http://atomdb.org/interpolation/index.php for an explanation.

The internal handling of statistics has been improved. You can define a different fit statistic for each spectrum but there is only one total test statistic. The output to the screen has changed slightly so if you have scripts which read xspec log files they will break.

There are some new models including:

Corrales' dust extinction models (ismdust and olivineabs)
Becker & Wolff's self-consistent cyclotron model (bwcycl)
A new Polish thermal comptonization convolution model (thcomp)
Nulsen's cluster mass models (clmass, nfwmass, monomass)