

# Breaking into the Black Box: Line Fitting and Interpretation for the Confounded\*

Michael A. Nowak (WUSTL)

— with help, but not blame, also to be attributed to  
Jörn Wilms & Victoria Grinberg —

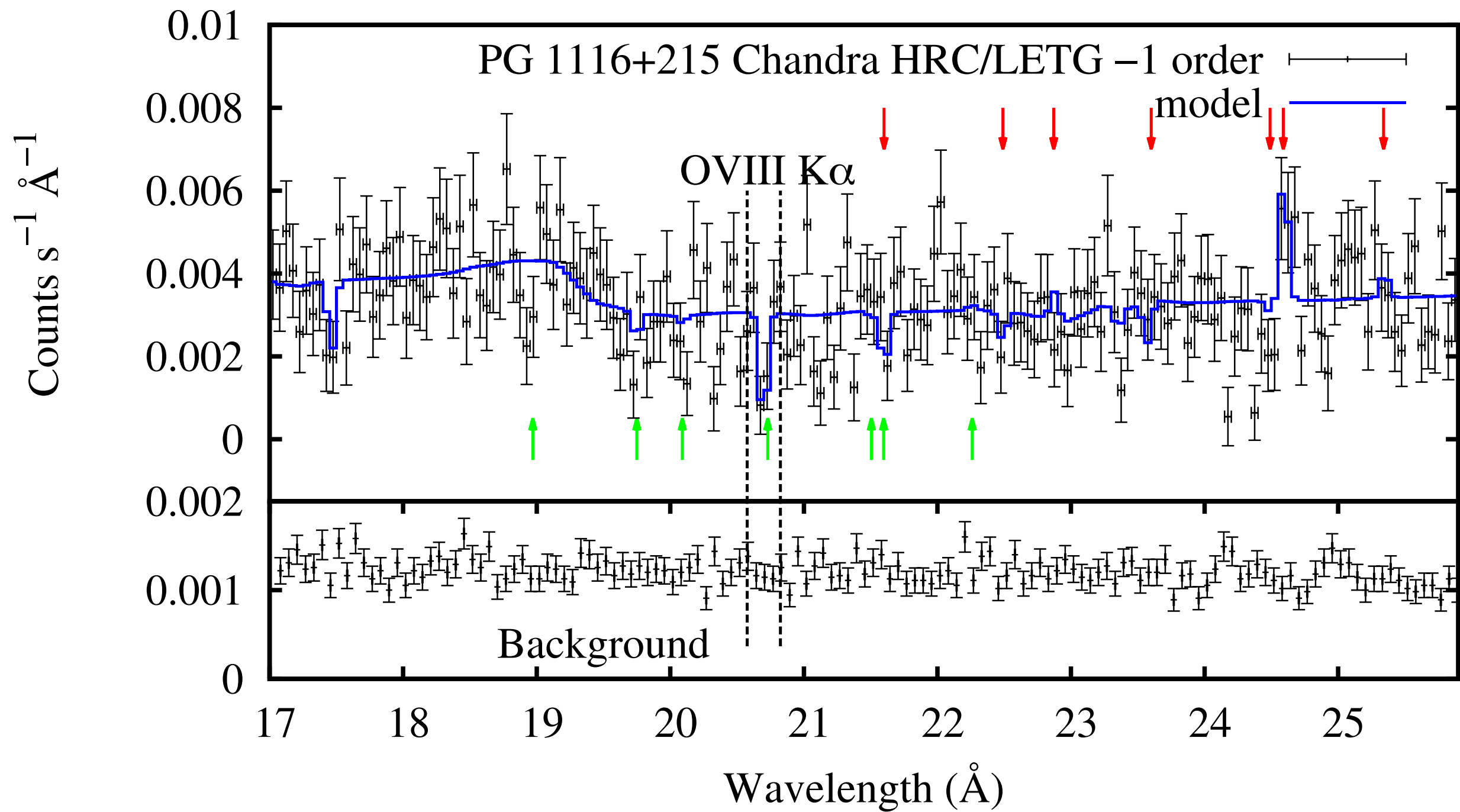
(\*I very much mean me, in this case.)

- High Resolution X-ray Spectroscopy is *hard*  
(optical folks don't know how easy they have it!)
- Expertise applied to community model development is rare
- Chris Reynolds's farewell address as Chair of HEAD —  
bar Randall Smith, Jelle Kaastra, and Tim Kallman from  
ever sharing the same transportation
- We make it harder on ourselves by pushing (not always  
properly) to the limits of S/N
  - Need to Fill Bins + Slow, Complex Models => Pain!
  - Need large effective area missions! (*XRISM* great for  
 $\gtrsim 1$  keV; really need *Arcus* for  $\lesssim 1$  keV.)

- Two Kinds of Examples — Search for Warm Hot Intergalactic Medium (WHIM); Ultra-fast Outflows (UFOs) in AGN
- WHIM: we need to be much more careful with statistics & *systematics*: simulate, simulate, simulate ...
  - Really is a project for the future (i.e., *Arcus*)
- UFOs: Going from “blind line searches” to physical models
  - We need to think more about the statistics of this
  - Parallelization very helpful (slow! how to make faster?)  
MCMC very helpful (but open models for ancillary quantities!)
  - Heard similar thoughts from Anna Ogorzałek for High S/N case of NGC 4051

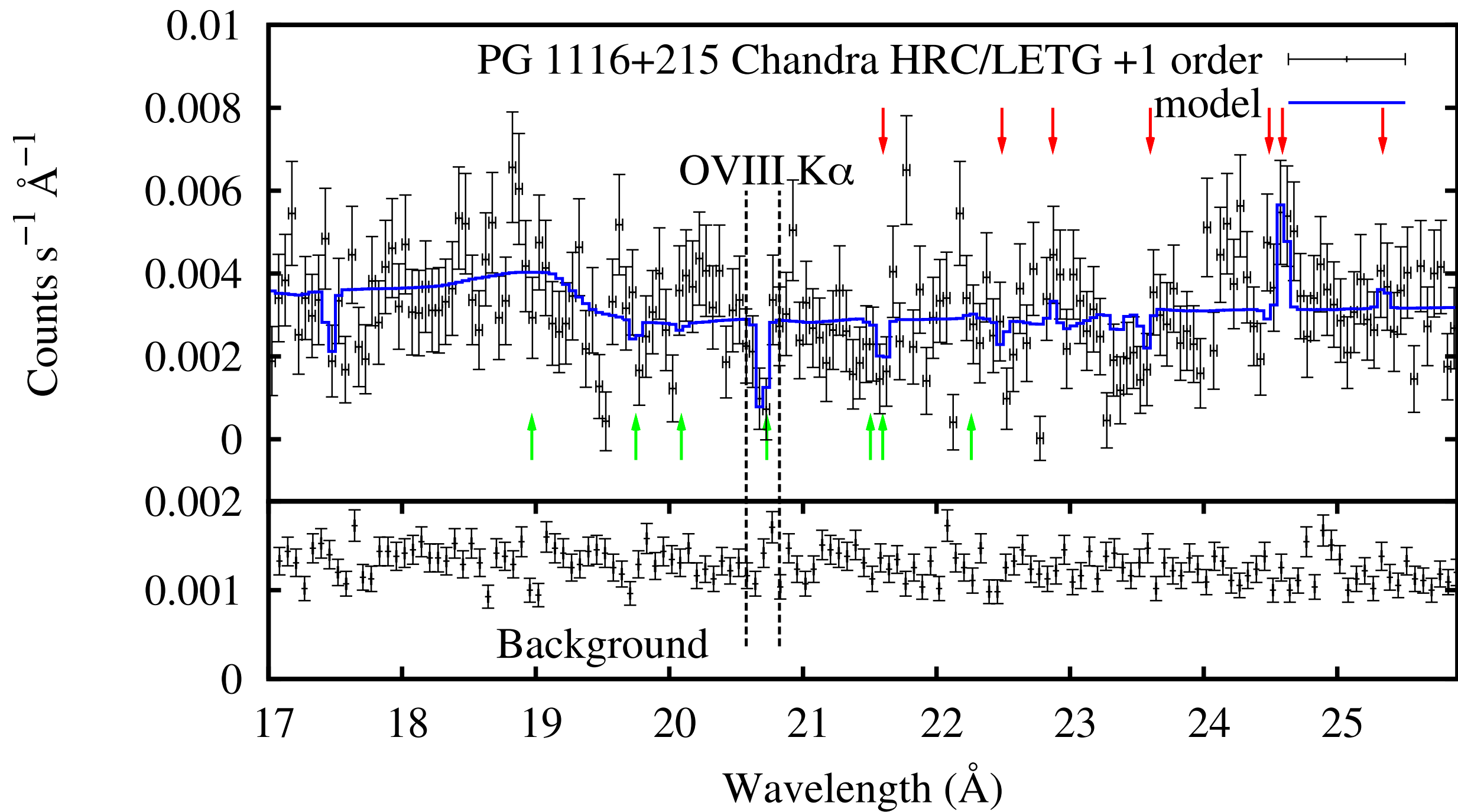
# WHIM Search

- PG 1116+125 (Bonamente et al. 2016, MNRAS, 457)
  - How solid was the initial claim, and how much time did we really need to confirm a detection?
- H1821+643 (Kovács et al. 2019, ApJ, 872)
  - How solid was the initial claim, and do how much do we have to obsess over systematics?



PG 1116+125, HRC-LETG  
Bonamente et al. (2016)

88 ksec

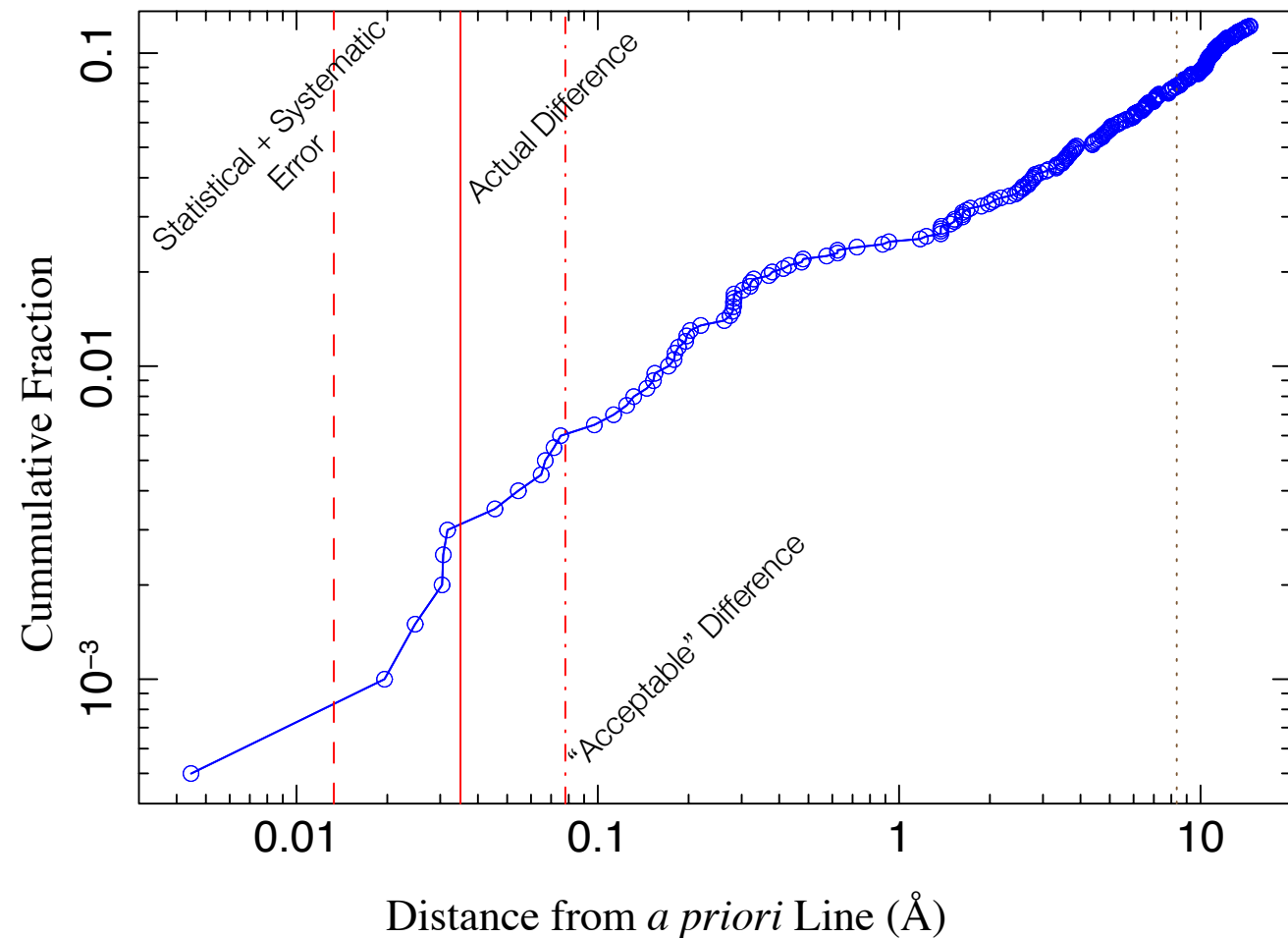
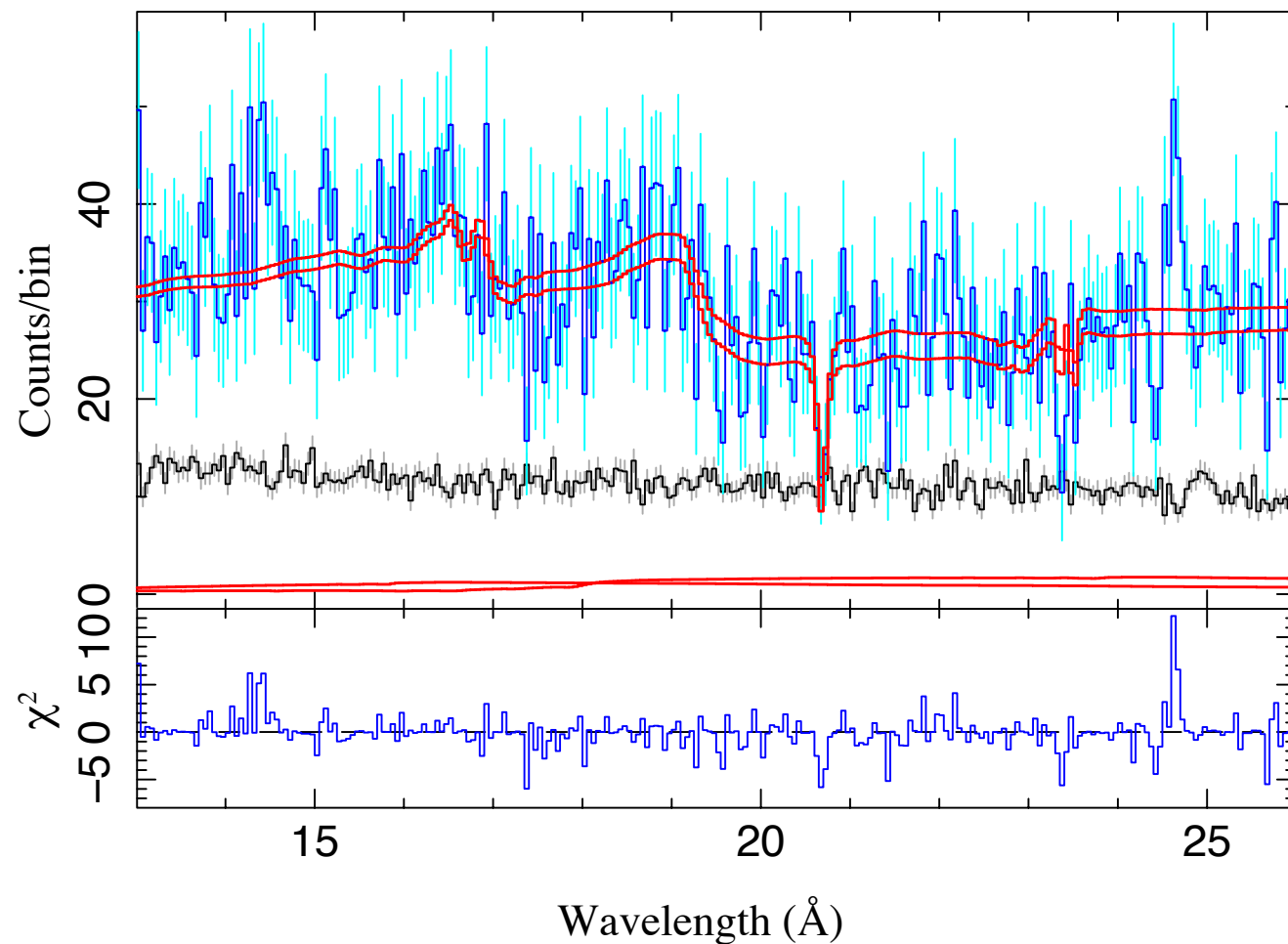


PG 1116+125, HRC-LETG  
Bonamente et al. (2016)

88 ksec

- Arrows are expected locations of possible OVII and OVIII systems, based upon *HST*-COS measurements of OVI
- Most significant residual is claimed to be  $5.2\sigma$
- Fitted wavelength has to be shifted by  $\gtrsim 0.03 \text{ \AA}$  to agree with expectations — *systematics important!*
- Methodological Issues to Worry About:
  - LETG: *analysis must include higher orders*
  - Cash statistics: *background should be modeled*
  - Always use care extrapolating  $1\sigma$  errors to zero bound

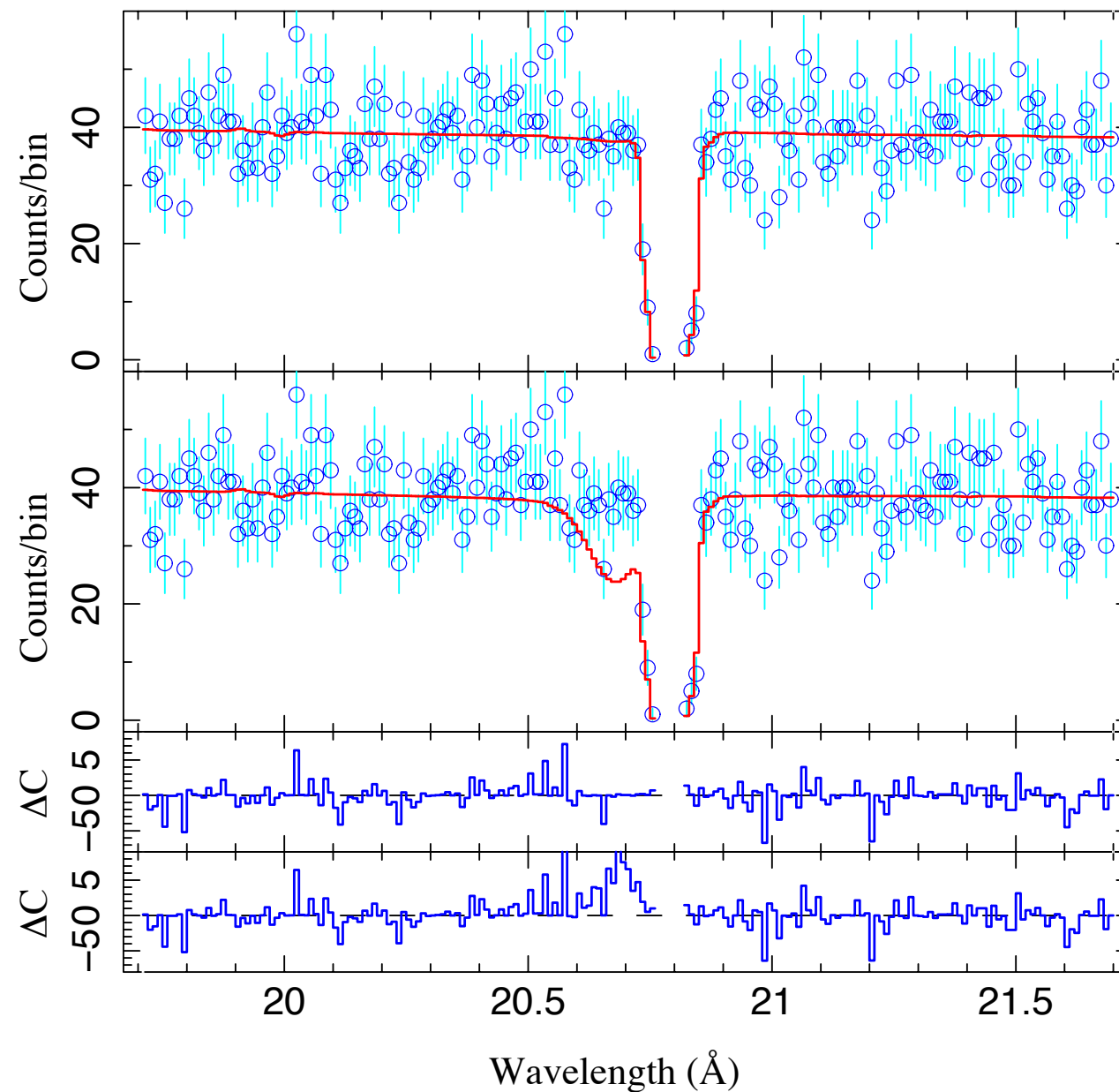
# PG 1116+125 Reanalysis



- Individual LETG orders shown; residuals with line removed
- 1.4 million simulations (without line), with comparable equivalent width line found at given distance from “known” source.
- 99.4%—99.7% significant, *ignoring systematics*, but ...

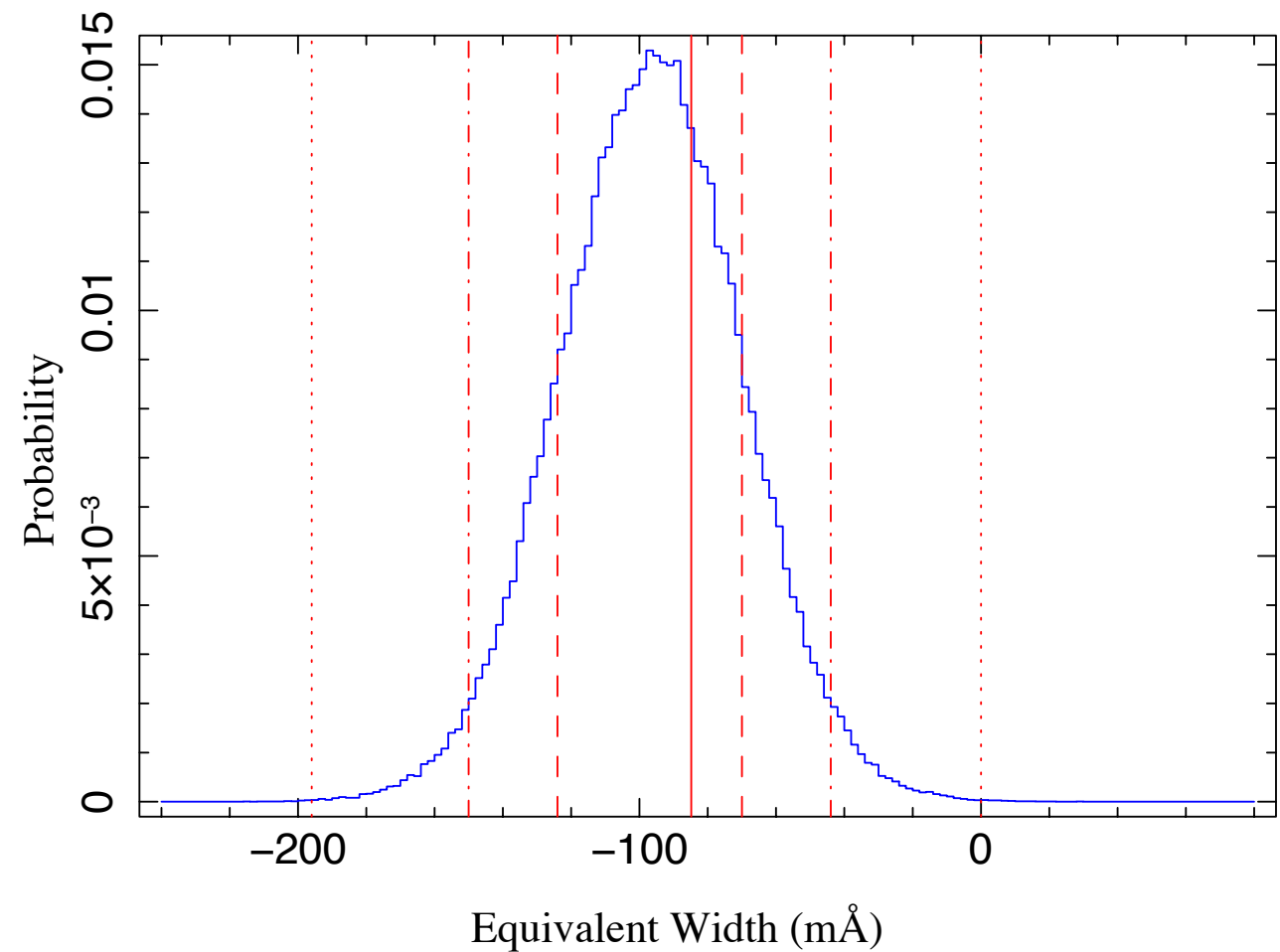
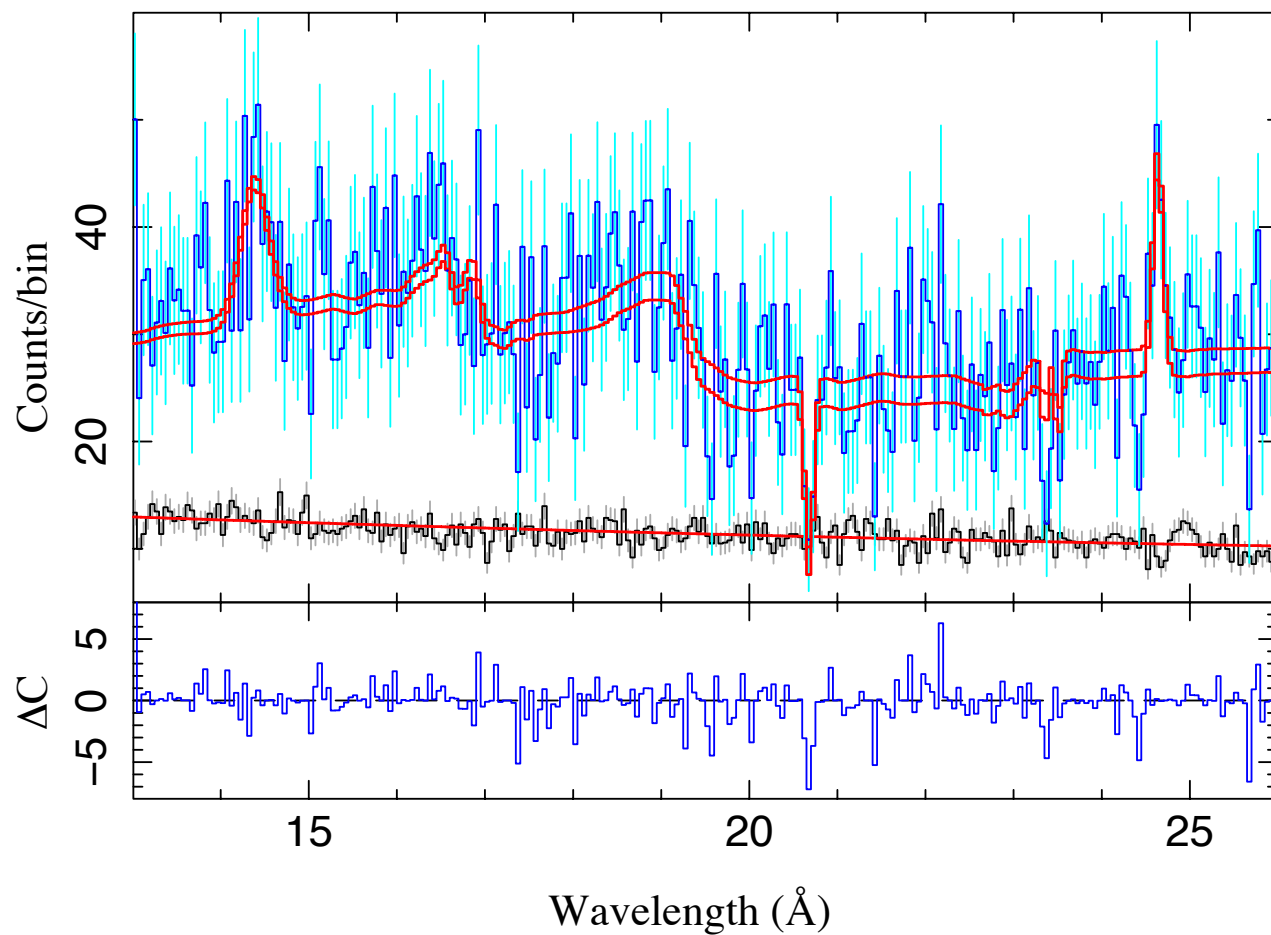


# PG 1116+125 Reanalysis



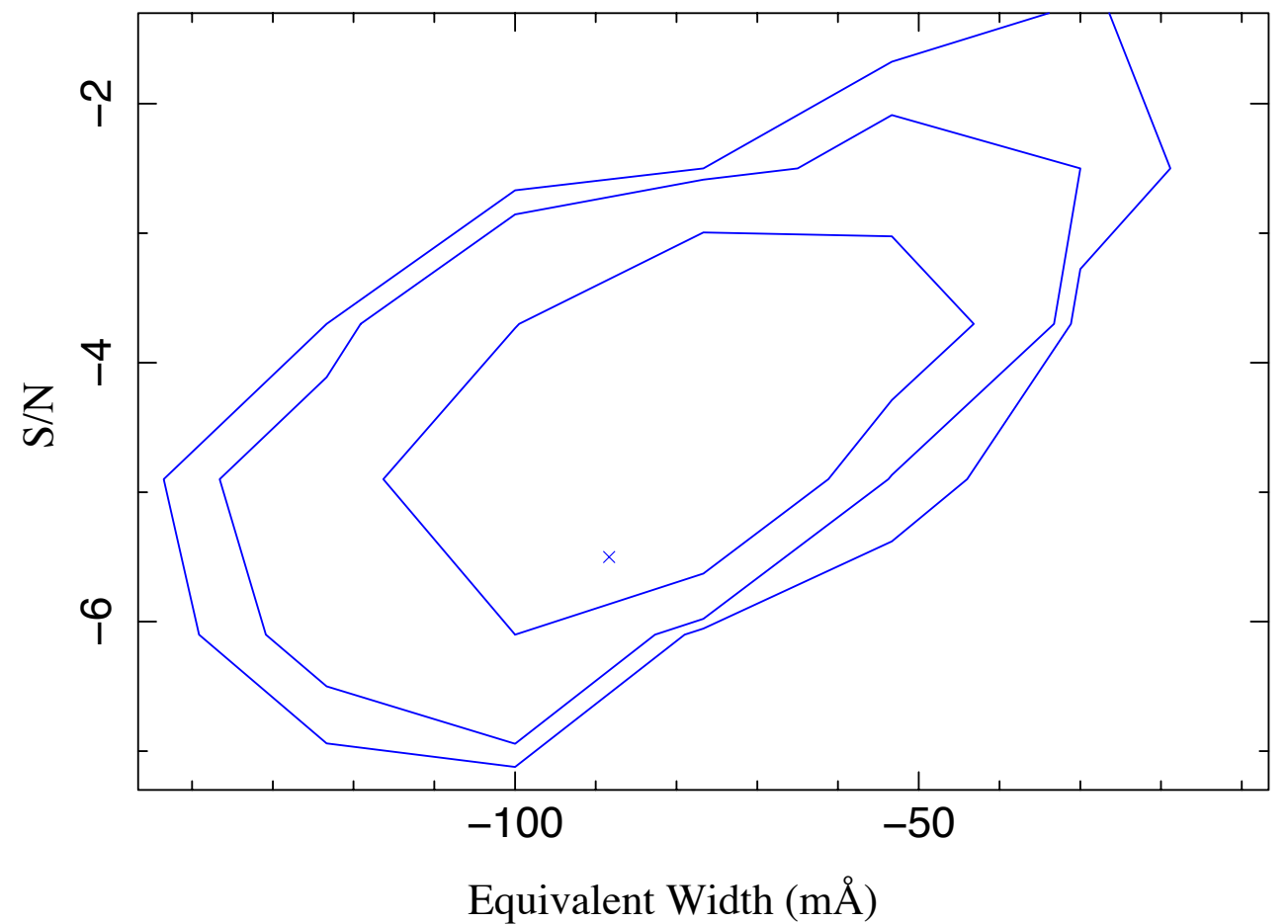
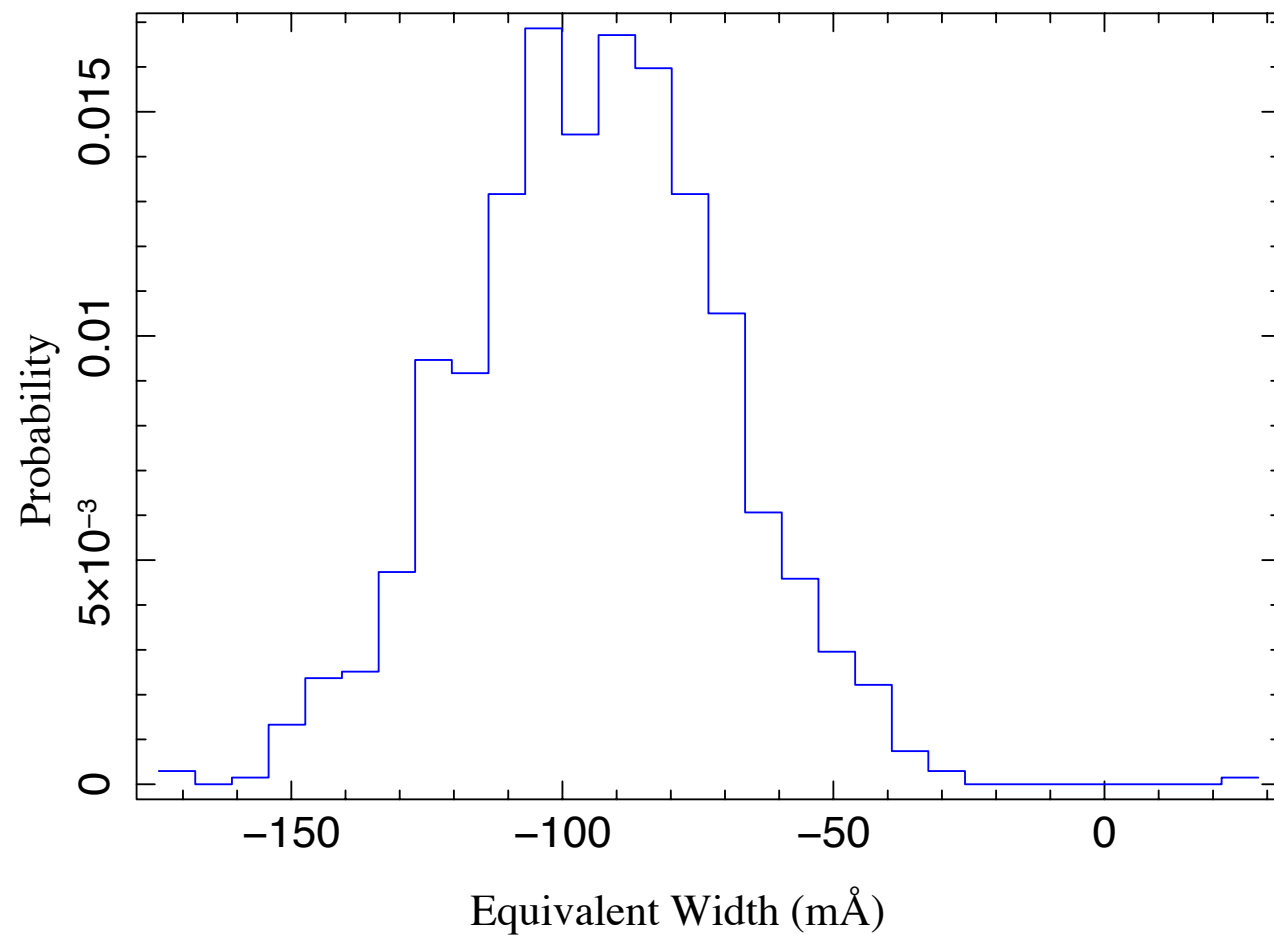
- XMM-RGS fit (tricky, but doable): probably should have seen the line if real. Systematic issues in LETG? Random chance? Or systematic issues in RGS so you shouldn't have seen it?

# How to Verify?



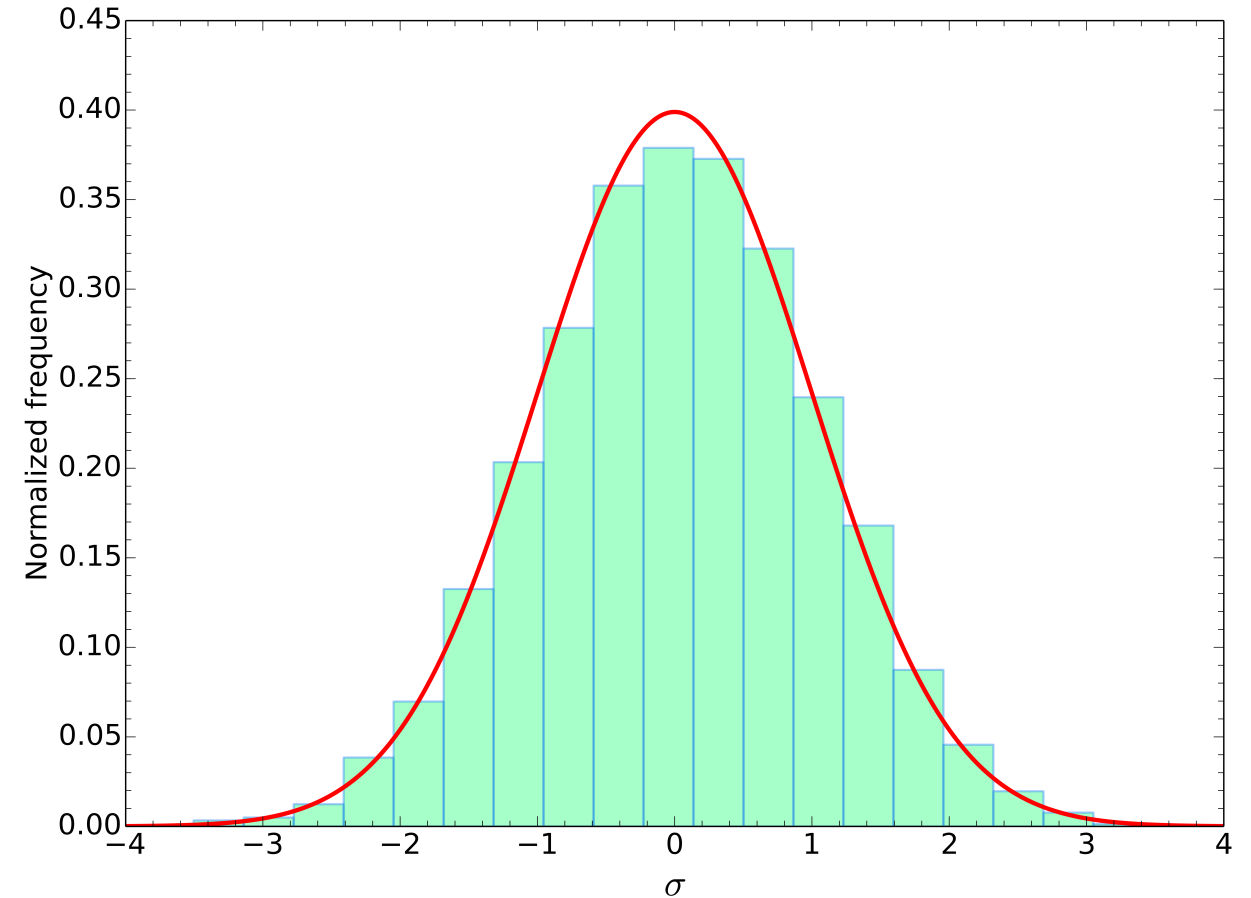
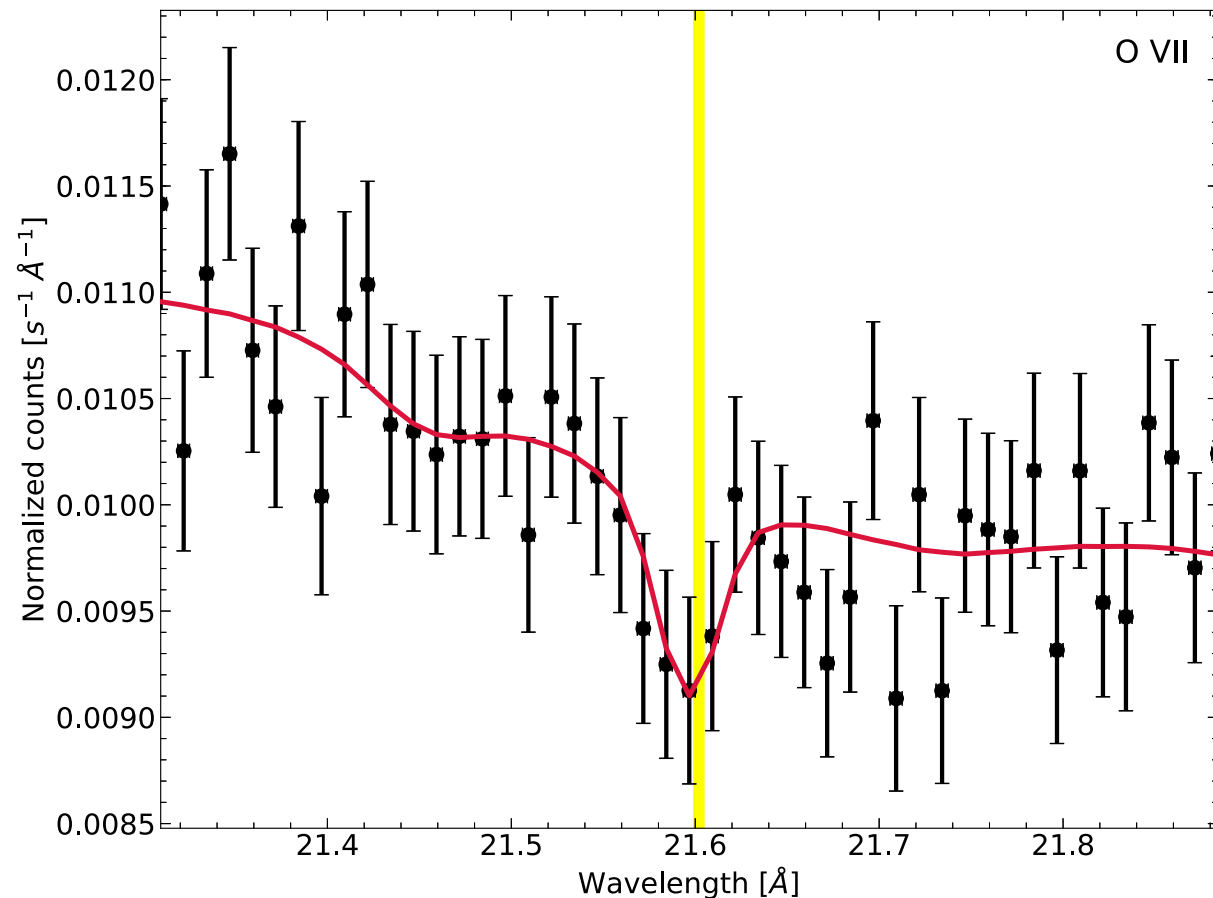
- Cash statistics, modeled background, MCMC analysis:  $3.4\sigma$
- The MCMC posterior should be used to assess further observations! What does an extra 280 ks buy?

# How to Verify?



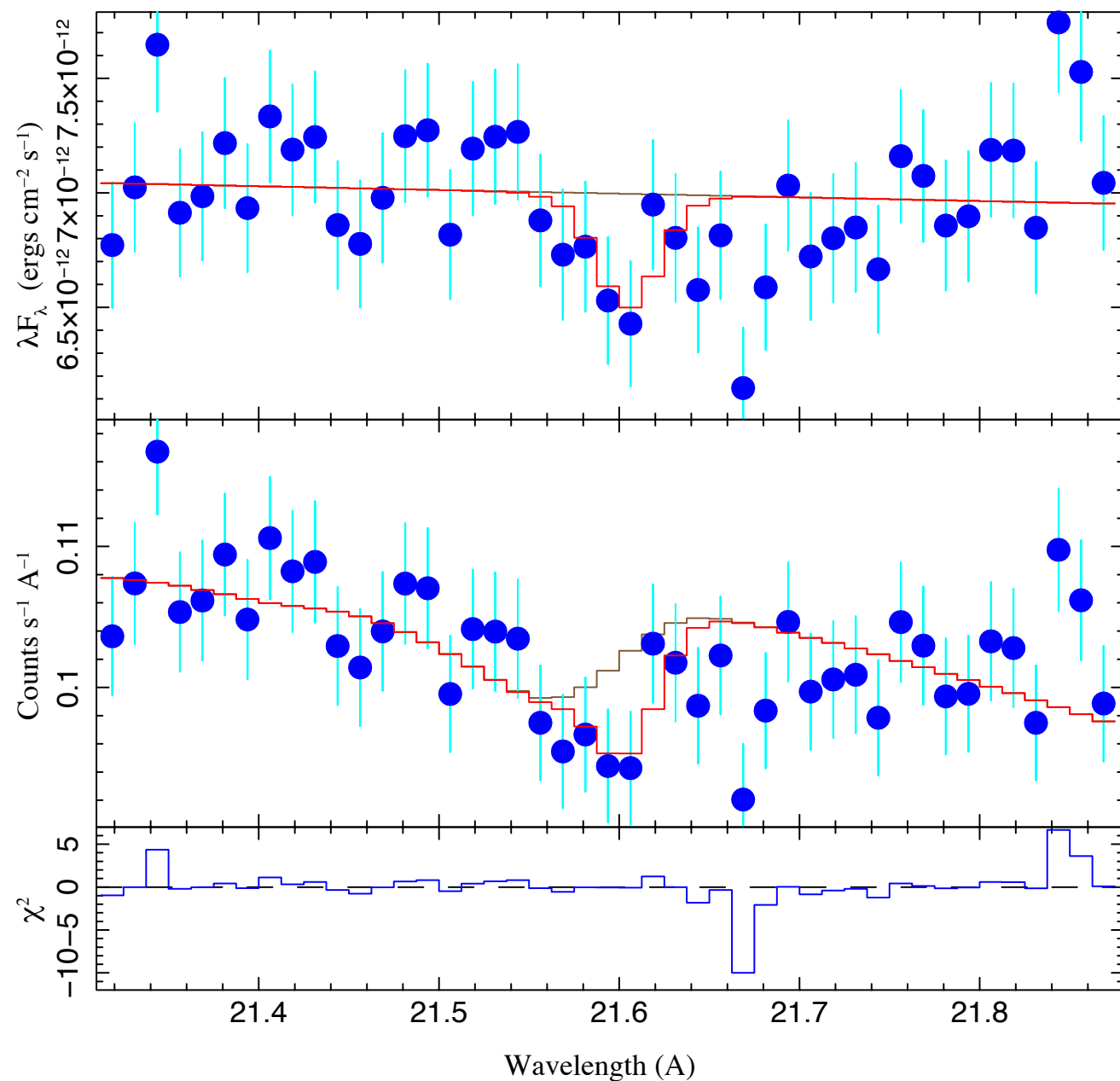
- 1000 simulations with fits & error bar search
- 48% chance that further observations leads to  $S/N > 5$
- 3% chance that further observations decreases S/N

# H1821+643



- Stacked 17 site lines with galaxies with known redshifts/impact parameters:  $3.3\sigma$  (Kovács et al. 2019).
- Monte Carlo analysis to assess significance by summing 17 *random site lines*; consistent with  $3.3\sigma$ .

# H1821+643 Reanalysis



- Stacking the data (differently? not 100% sure):  $2.6\sigma$
- Near a “stacked” detector feature: dip in effective area.
- *I would not choose random stacks in this case; replicate exact procedure of analysis*
- General question: how do we account for (usually uncounted) human decisions?

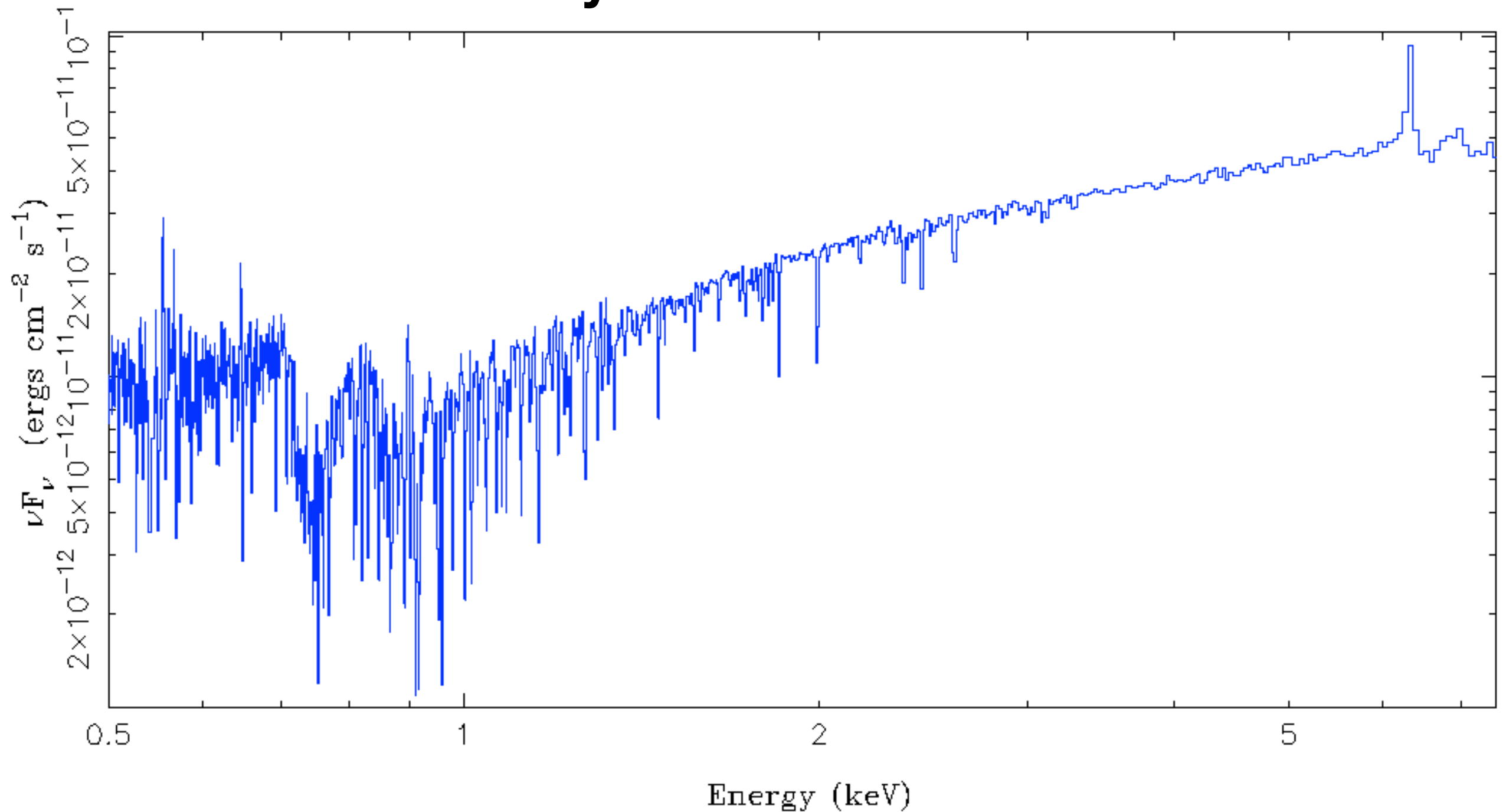
# Line Searches

- High Energy Astrophysicists are slightly insane:
  - Fit continua over 2–3 orders of magnitude in energy, with 1–6 groups of lines (3–4 orders of magnitude in ionization parameter), shifted by  $-0.3c$ – $0.3c$
- Example of high S/N (visual, TGCat) I won't touch (but see poster by Anna Ogorzałek for High S/N case of NGC 4051)
- Example of lower S/N — visual line id in TGCat data of M81\*
- Example of low(ish) S/N Blind Line Search in PG 1211+143
- Thoughts on Future Extensions for Line Searches

# TGCAT: Transmission Gratings Catalog

- Location: <http://tgcat.mit.edu/>
- Description: Huenemoerder et al. 2011, AJ, 141, 129
- Examples & video demos:  
[http://tgcat.mit.edu/tgHelp.php?guide=help/tgcat\\_demos.html](http://tgcat.mit.edu/tgHelp.php?guide=help/tgcat_demos.html)
- Help desk: [tgcat@space.mit.edu](mailto:tgcat@space.mit.edu)
- Software: <http://space.mit.edu/cxc/analysis/tgcat/index.html>
  - S-lang script running under ISIS, implementing CIAO tools

# NGC 3783 — Nearly 1 Ms Summed & Plotted via TGCat

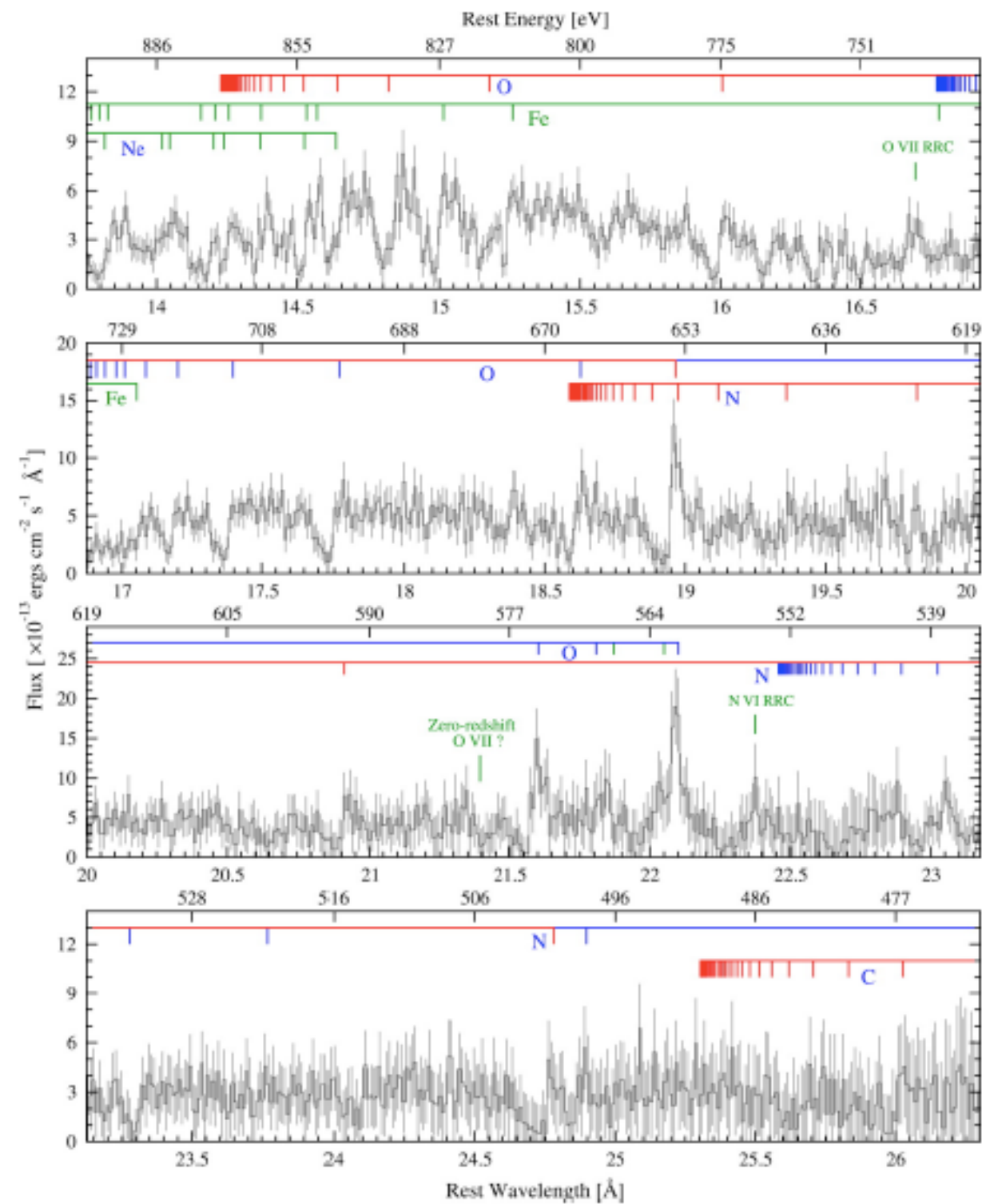
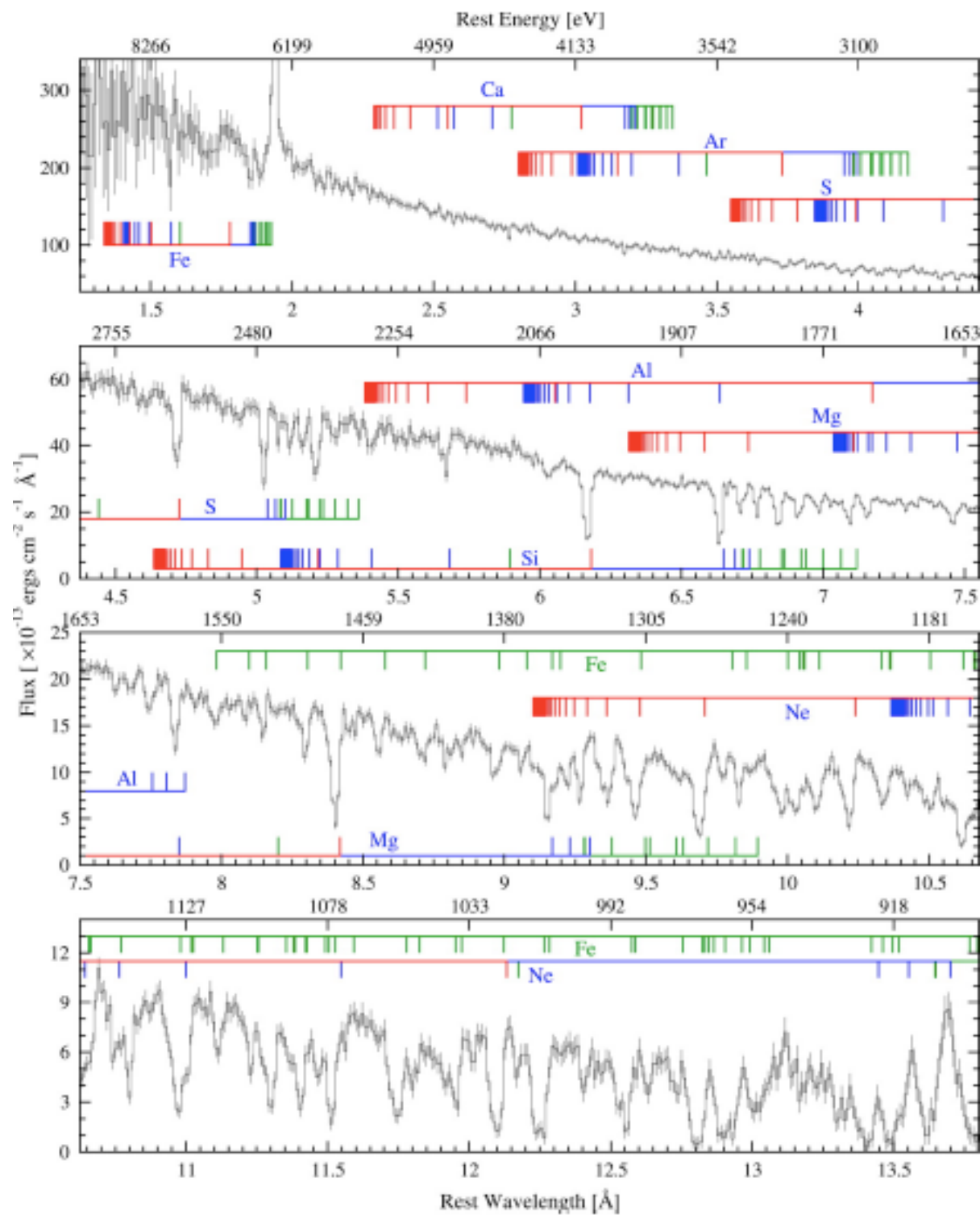


- TGCat makes it easy to visually inspect & download spectra
- This spectrum is too complicated for me!



# NGC 3783 900 ksec Chandra observation

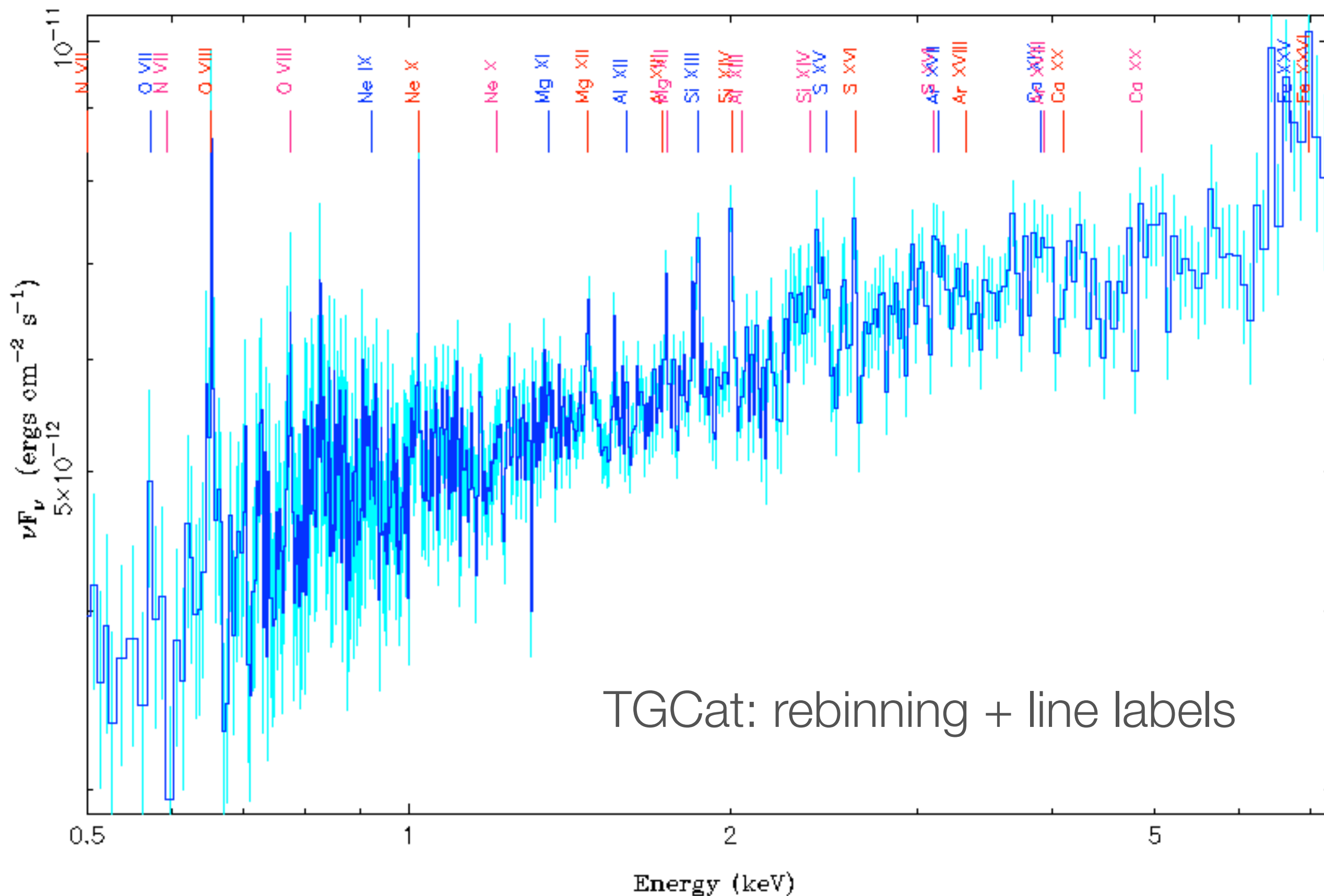
## Absorption



135 absorption lines identified

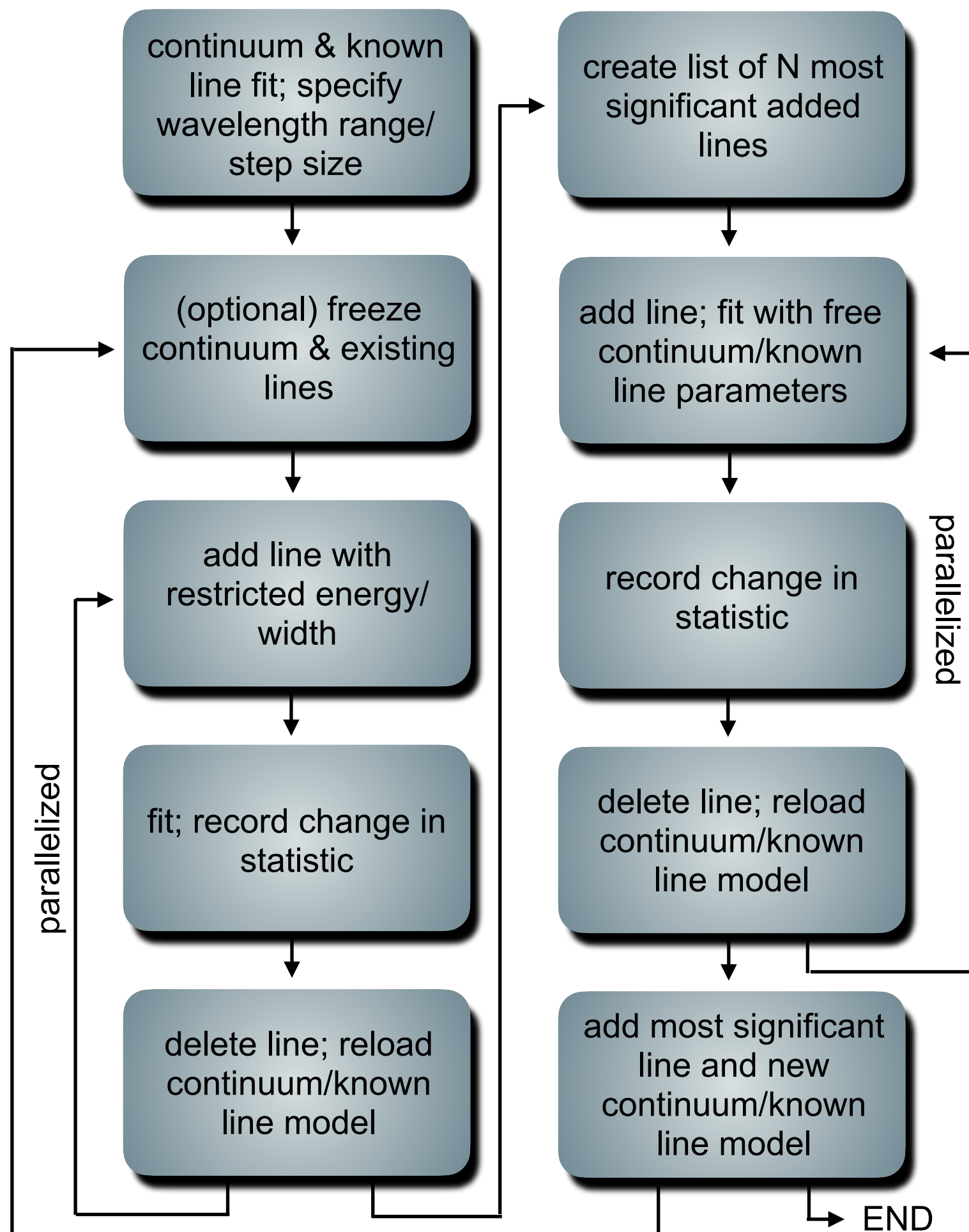
Kaspi et al. 2003

# M81\* — 450 ks Summed & Plotted via TGCat



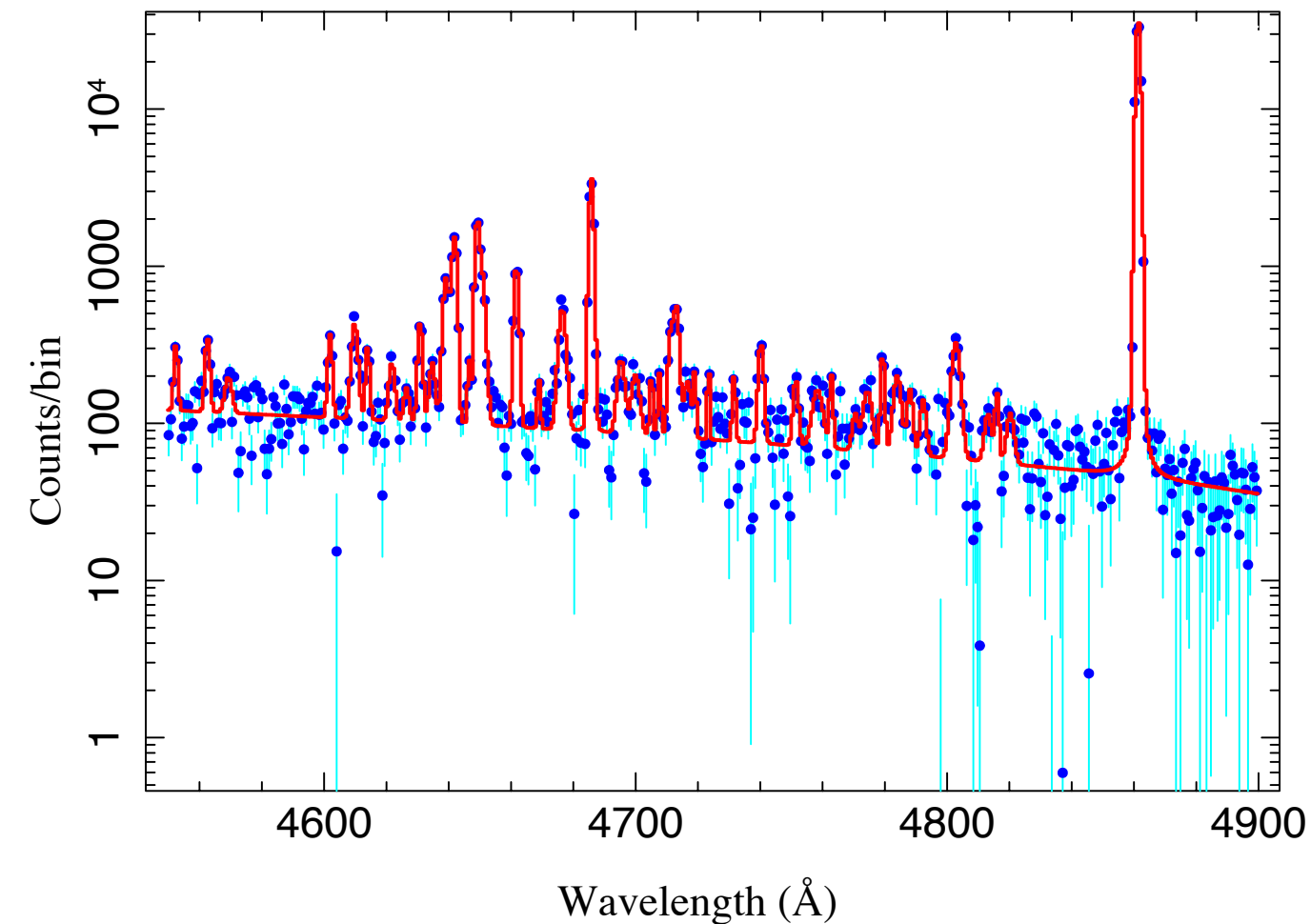
# Blind Line Searches

- M81\* early HETG test case (Young et al. 2007, ApJ, 669)
  - “Bayesian Blocks” technique (analogous to timing)
  - Blind line search scan & identify (Julia Lee; Andy Young)
- Working on improving this — concept is to live between:
  - continuum+(gaussian+gaussian+gaussian+...)
  - (warmabs\*warmabs\*...)\*continuum (wait, grow old, die)



- Prototypes available from Remeis ISIS Scripts page  
<https://www.sternwarte.uni-erlangen.de/isis/>
- Similar concepts often use “fixed grid” searches
- Putting in customization: fixed vs. free components; “symmetric” line searches; multiple functional ties; no negative counts!
- “Comb Searches” needed
- *Parallelization important!*

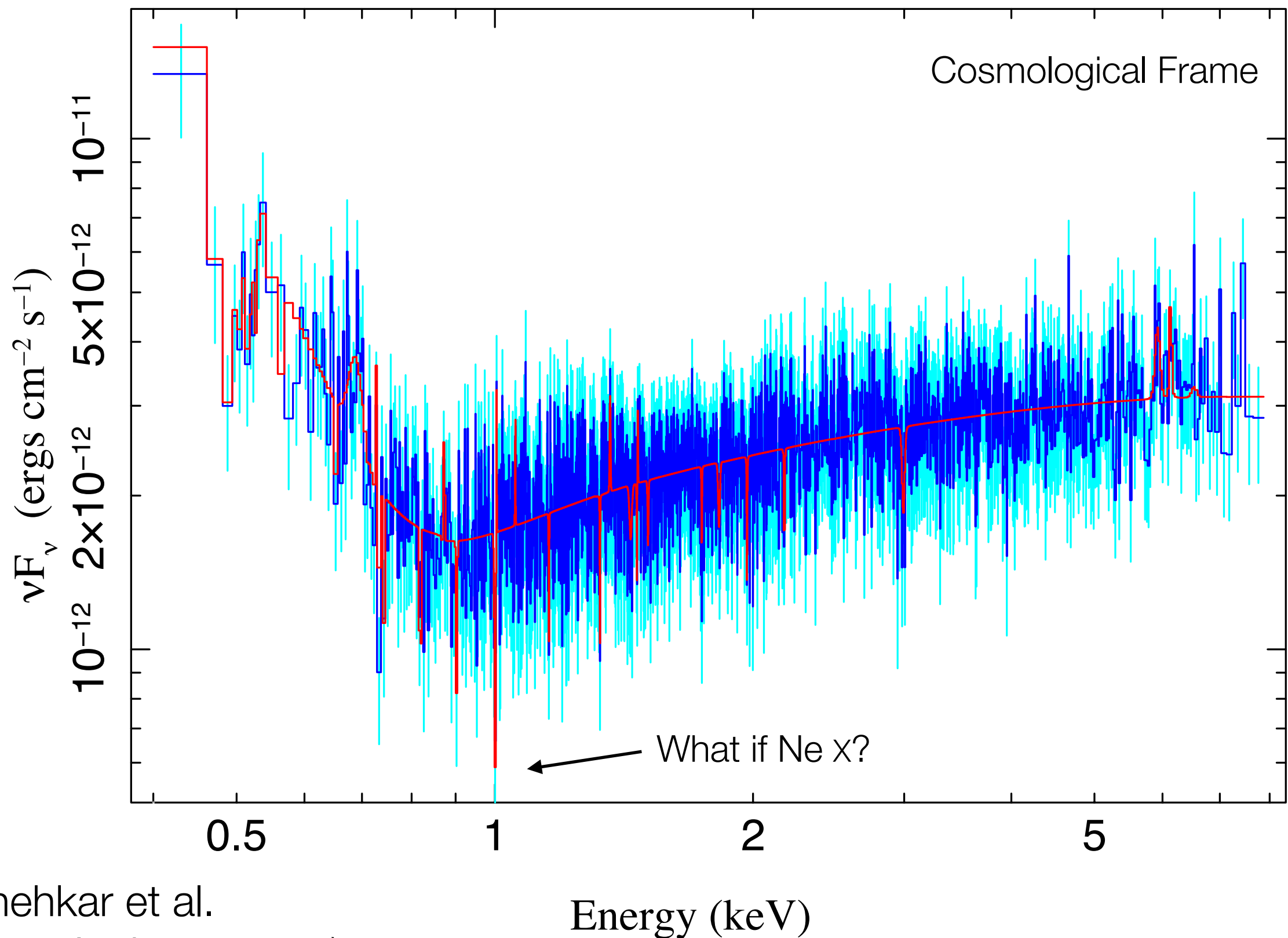
# Early (Optical!) Example



See optical line code challenge:  
Wesson (2016, MNRAS, 456, 3774)

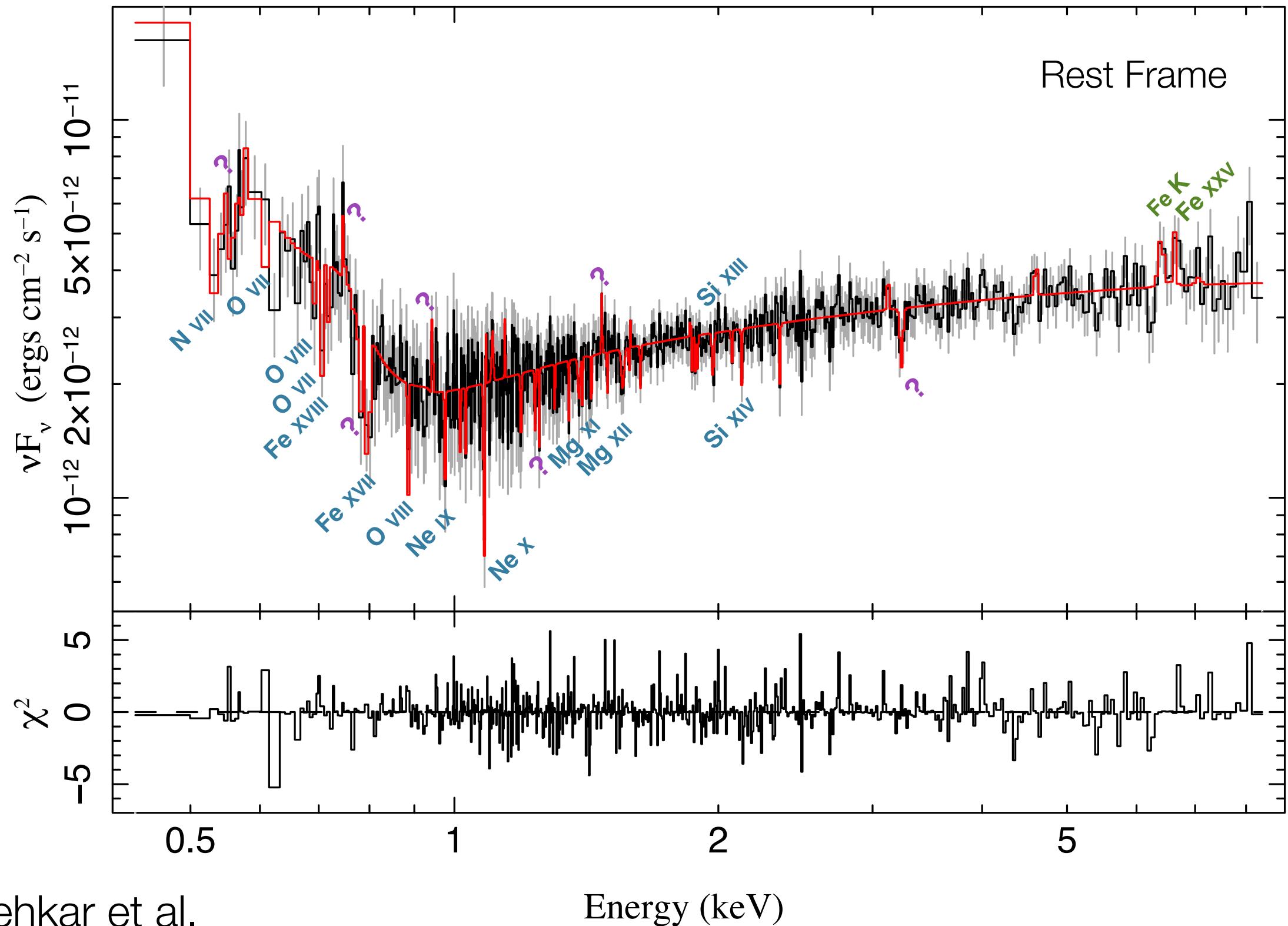
- I truly know nothing about optical spectroscopy ...
- Had to “make up”: response, background, (some) statistics
- Found 45 lines, ~40 “real”; multiple line blends. Most associated with “known” (What? Not 0.3c away??! Easy!!!)
- “Next best” code found ~30 lines; almost all codes missed blends
- But ... “X-ray” way was ~100× slower (few hours vs. few minutes)

# X-rays: PG 1211+143



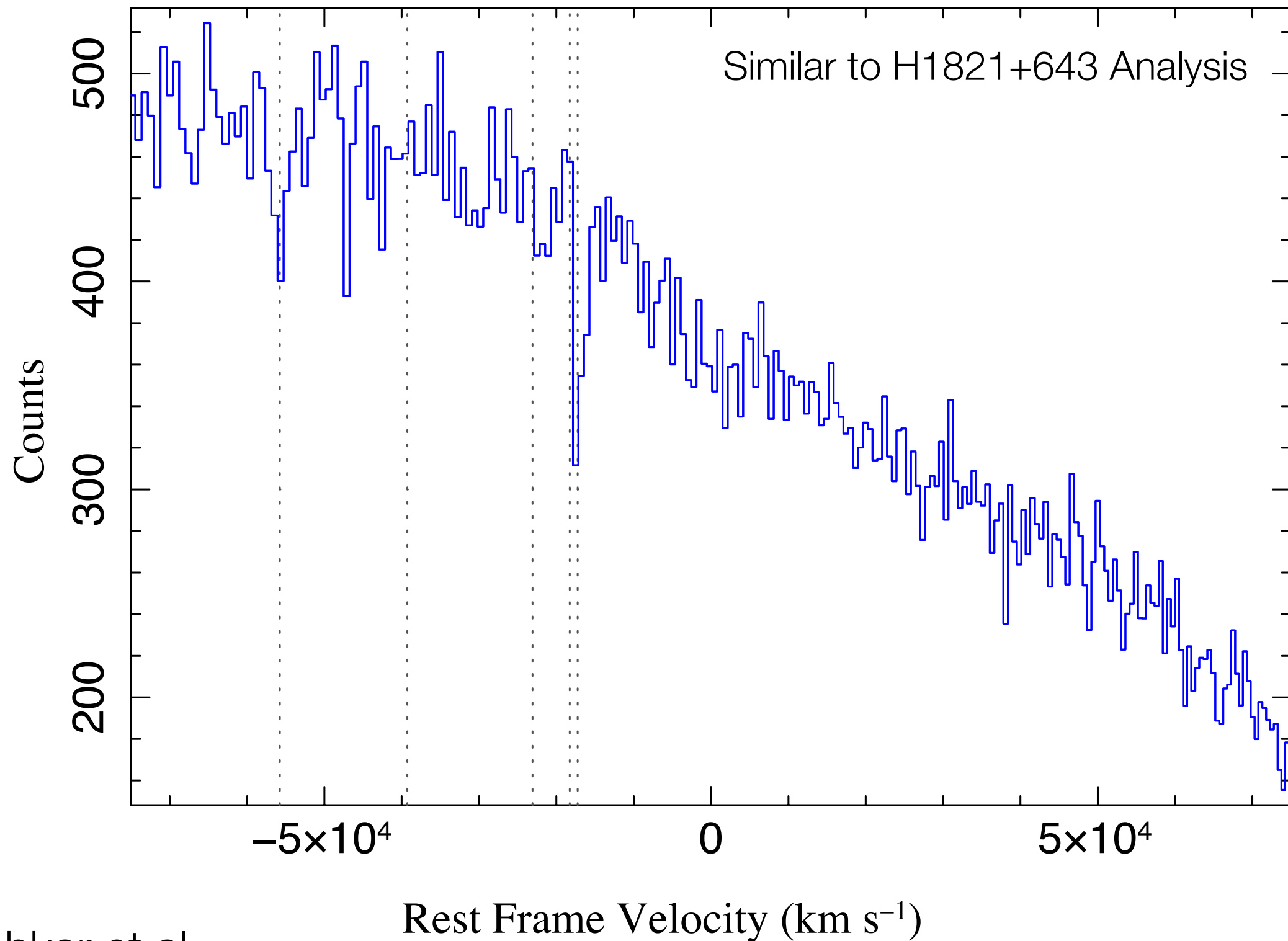
Danehkar et al.  
(2018, ApJ, 853, 165)

# X-rays: PG 1211+143



Danehkar et al.  
(2018, ApJ, 853, 165)

# X-rays: PG 1211+143



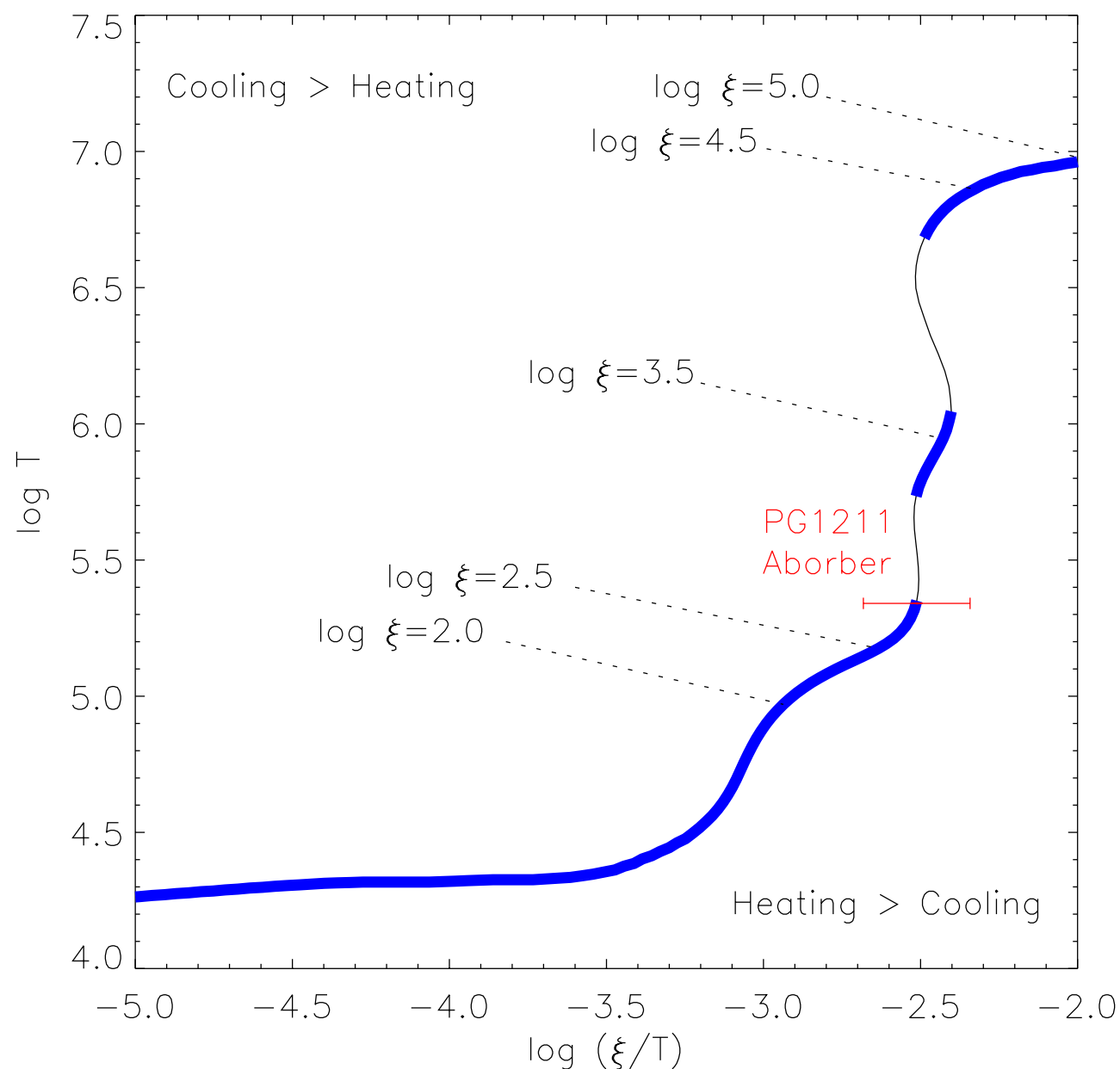
Danehkar et al.  
(2018, ApJ, 853, 165)



# Next Steps:

- Individual Lines => “Comb Searches” => Redshift Families
- Steal concepts from LIGO folks: initial search fairly simple “matched filter” grid search on a few parameters.
  - Is it enough to create continuum spline, “match filter” residuals with ( $\log \xi$  vs.  $N_H$ ) grids at “known” redshifts?
- Then jump to MCMC analysis with “slow”, detailed models
  - Need to “open up” slow models for “derived” parameters

# E.g., Stability Curves

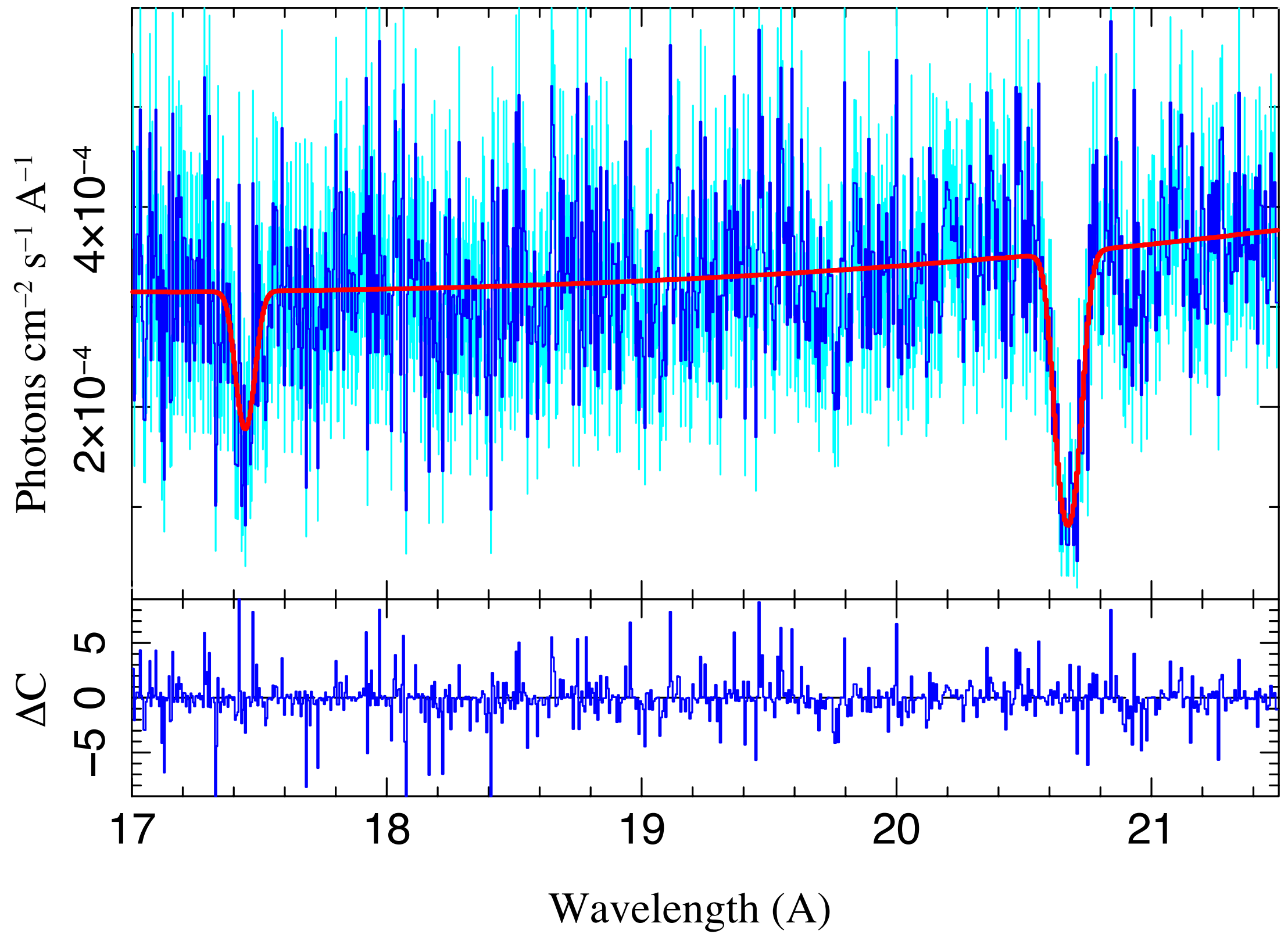


Danehkar et al.  
(2018, ApJ, 853, 165)

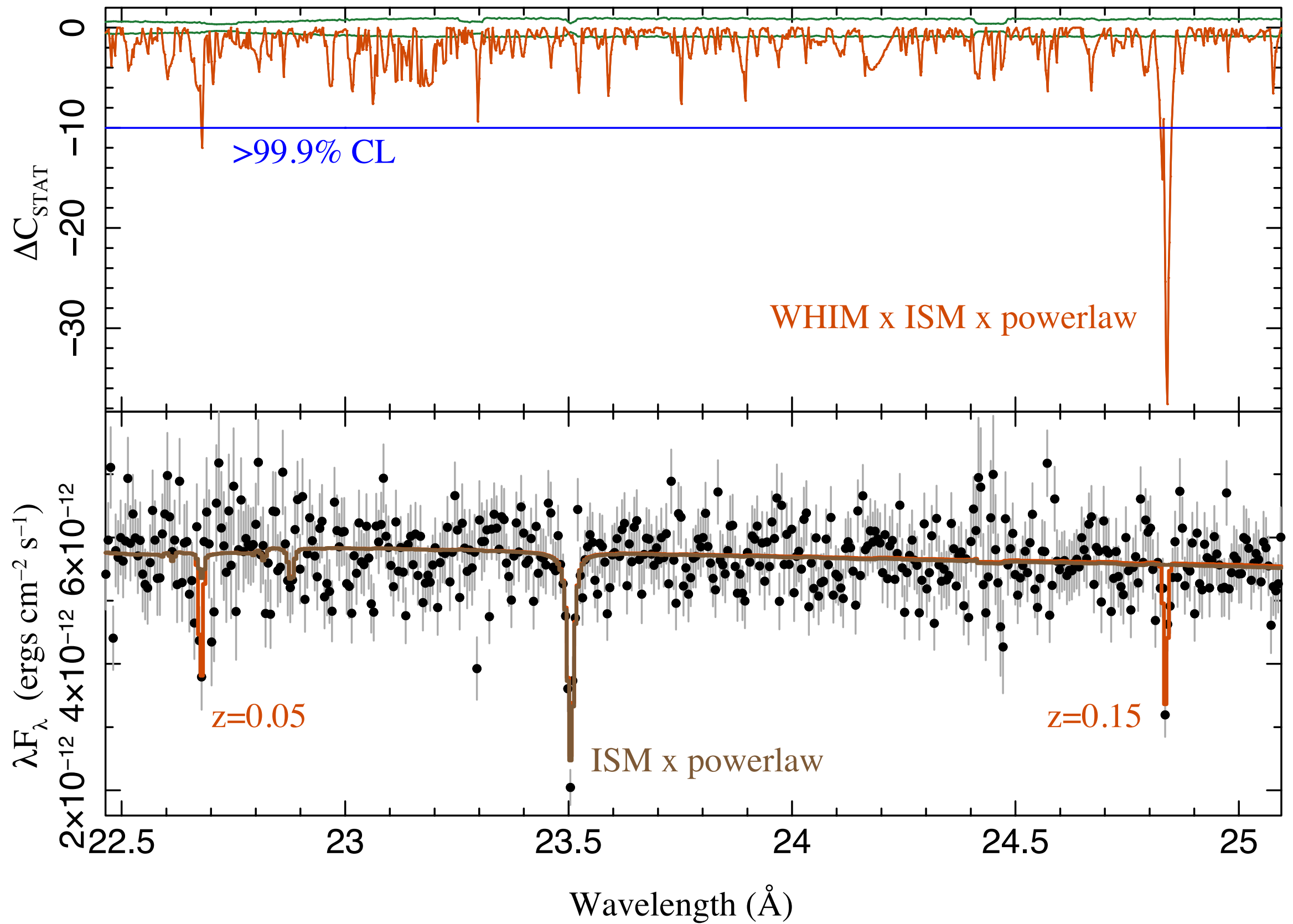
- The long pole is often the model calculations during the MCMC.
- “Derived” quantities might be a small perturbation (time wise) to the base calculation.
- Need to store interesting quantities with overall parameter chain.
- Interfaces to *XSTAR*/*warmabs* good!
- See, e.g., work by D. Huenemoerder & Lia Corrales: XSTARDB <http://space.mit.edu/cxc/analysis/xstardb/>

# Summary

- High Energy, Hi Res Spectroscopy has the nasty combination of needing lots of counts, but having small(ish) detectors, with very complex models.
- We need to be especially careful with our statistics & simulations & systematics. (No more WHIM until *Arcus*!)
- We need to try to make our models more accessible & faster: more open interfaces, parallelization, ...
- We need bigger instruments! *XRISM*! *Arcus*!



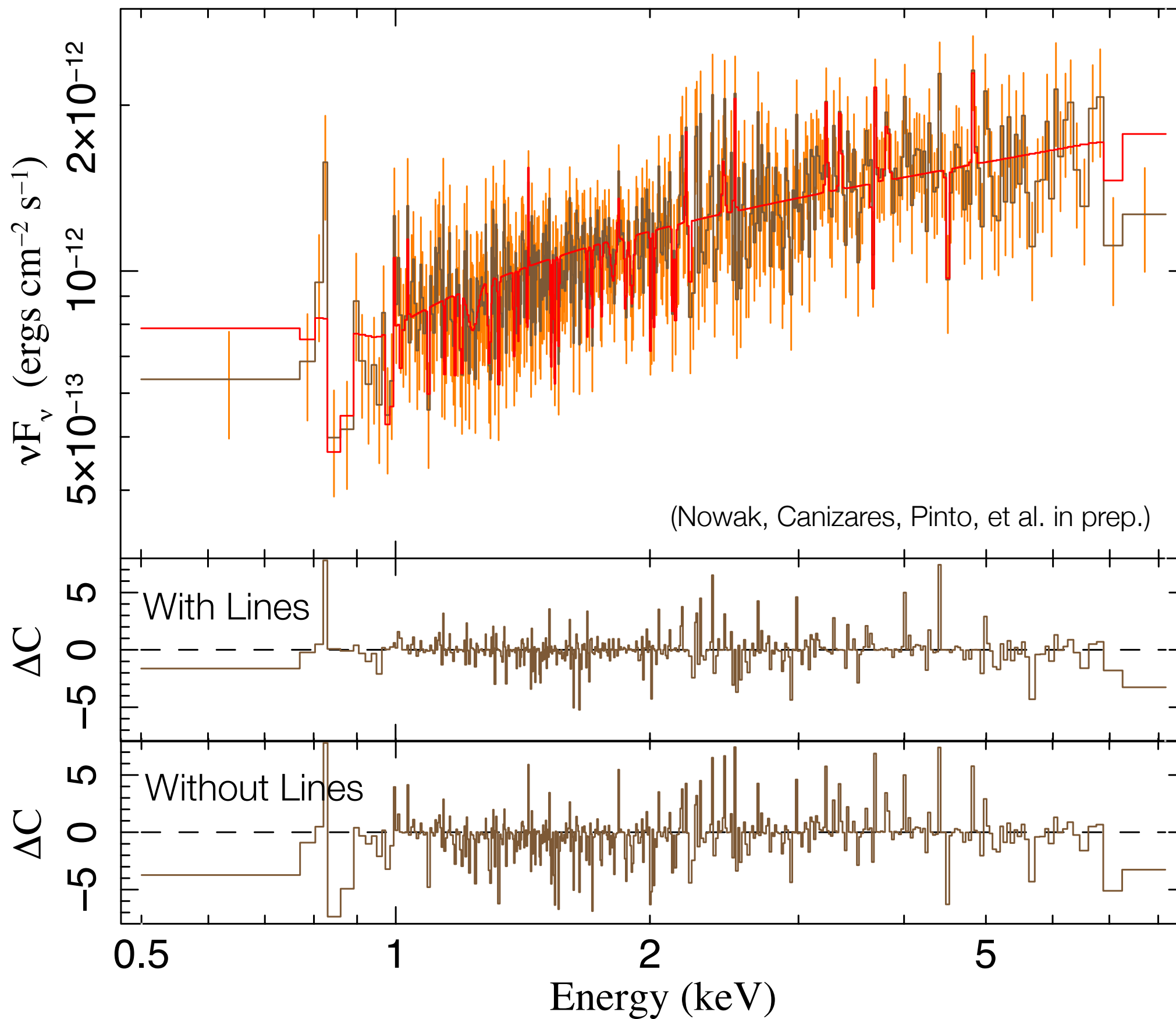
(Simulation of 10 ks Arcus observation of PG 1116+125, if line were real.)



(But of course, *Arcus* will be going after much weaker features...)

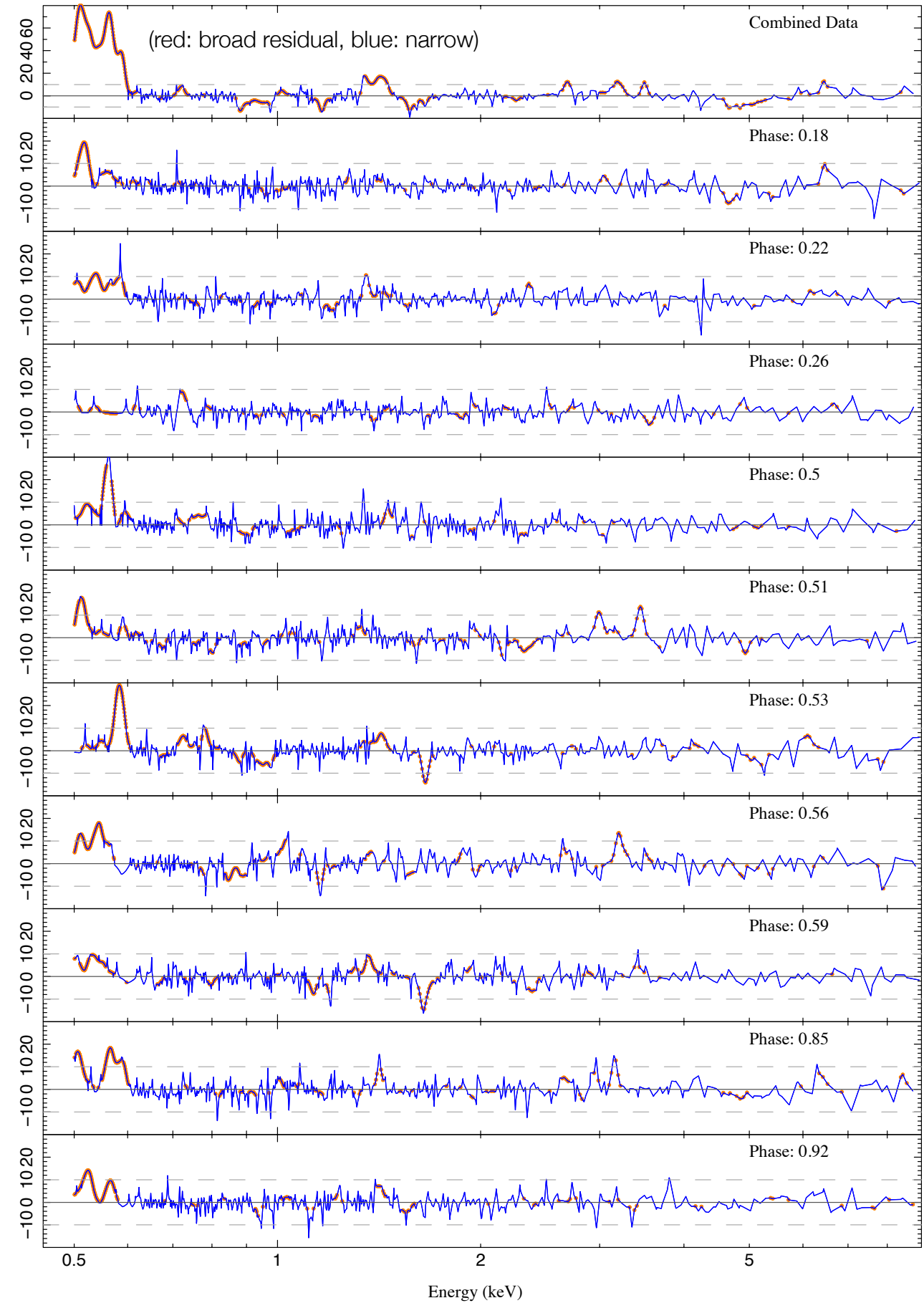
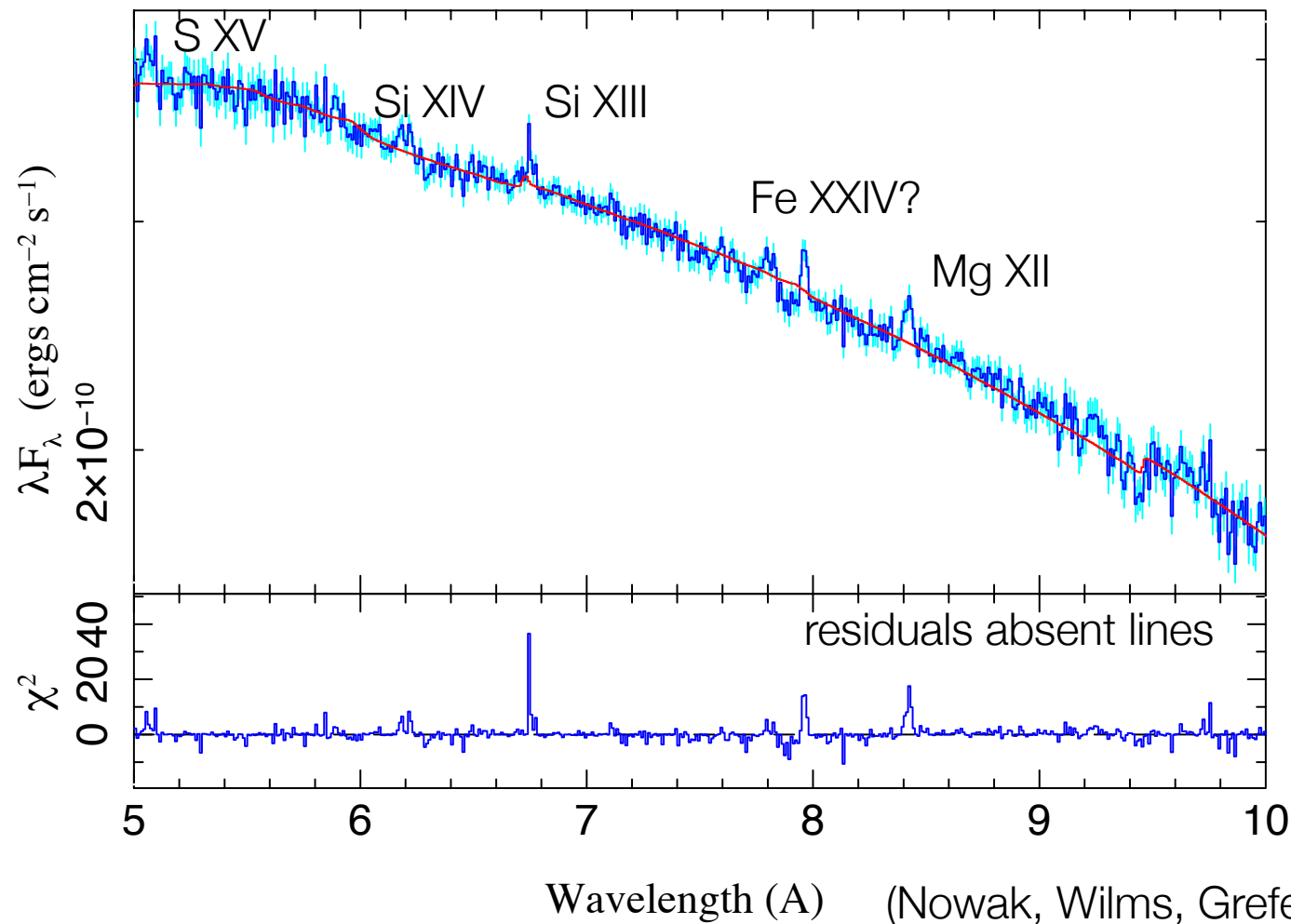
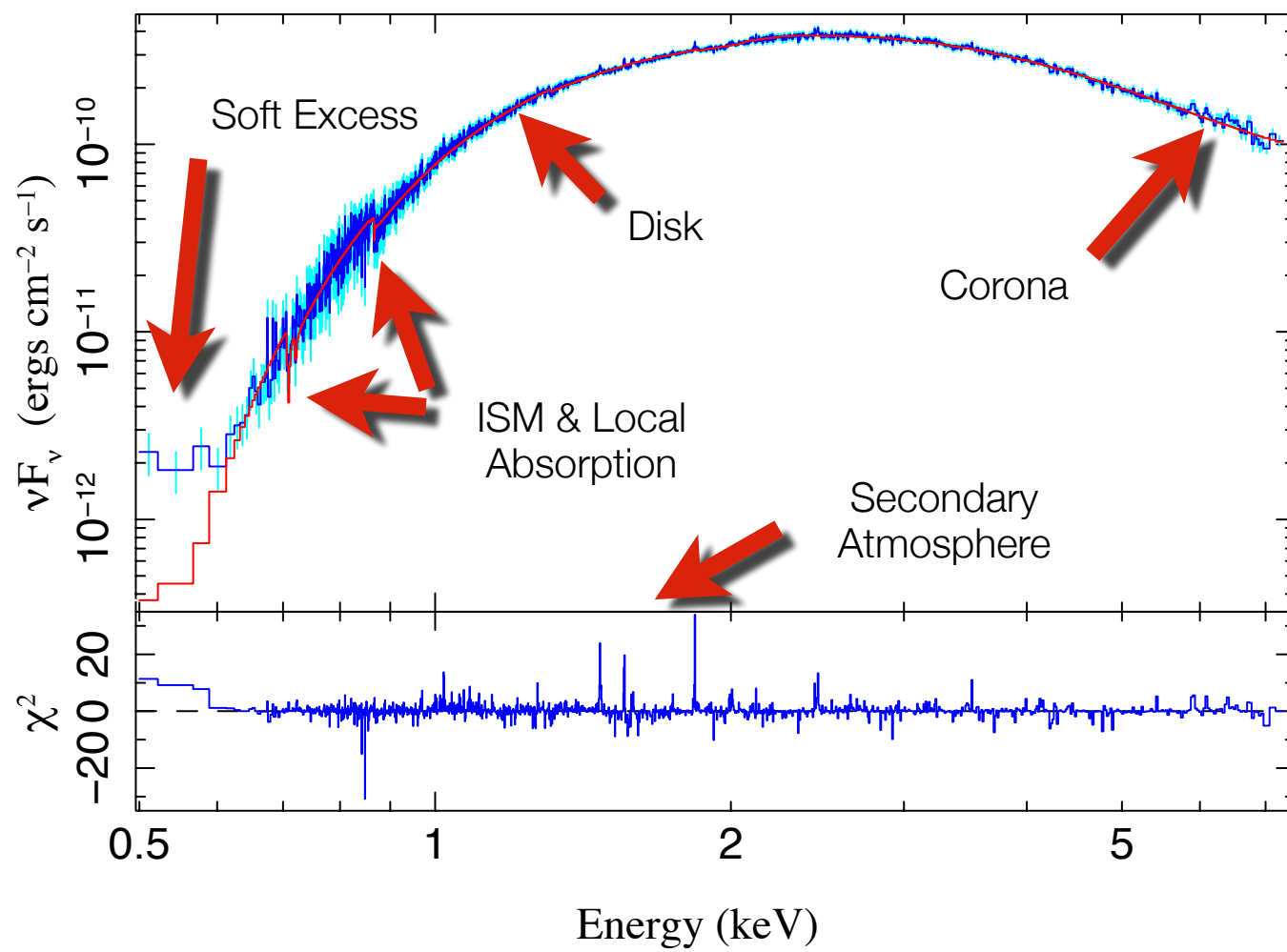
**— Extra Slides —**

# Blind Line Search: NGC 1313 X-1 (any lines weak & unidentified)



# Blind Line Search: LMC X-1

(after major lines modeled)



(Nowak, Wilms, Grefenstette, et al. 2020 in prep.)