Systematic Uncertainty: What's next?

Vinay Kashyap
(CfA/CXC)
Hyunsook Lee, Aneta Siemiginowska, David van Dyk, Jeremy Drake, Alanna Connors, Andreas Zezas, Taeyoung Park, Shandong Min, Jonathan McDowell, Arnold Rots, Pete Ratzlaff, Jason Kramer, Doug Burke, Brian Refsdal

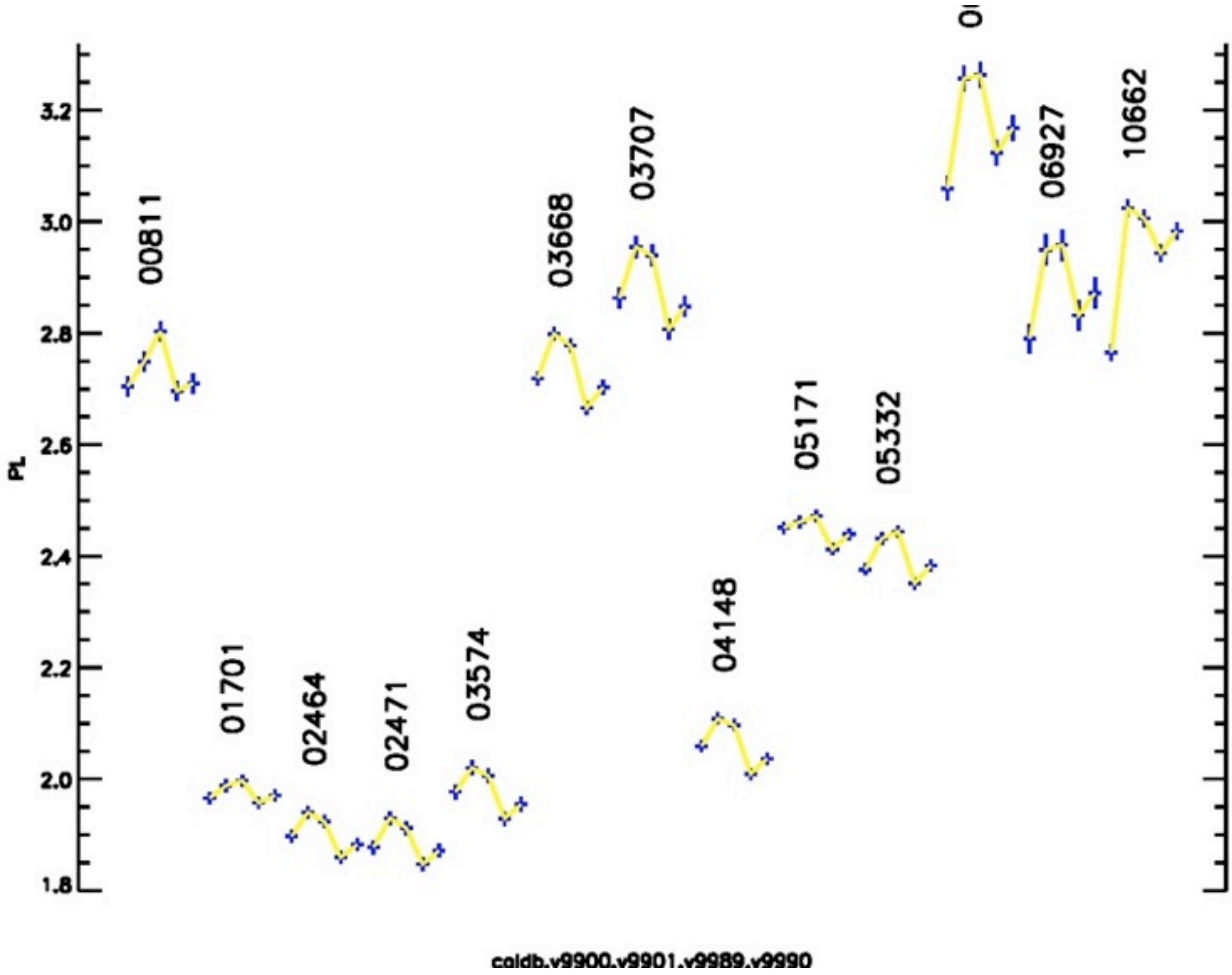
1. what Jeremy said
2. the quick and dirty method

   or, how to square-add errors via the Multiple Imputation Combining Rule

3. the theoretically completely general way to incorporate calibration uncertainty into spectral analysis
4. Theory vs Practice

   the AREF file specification

The Multiple Imputation Combining Rule



3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
PL
00811
01701
02464
02471
03574
03668
03707
04148
05171
05332
06927
10662
caldb.v9900.v9901.v9989.v9990

The Multiple Imputation Combining Rule

- Analyze a dataset using N separate cal products
- Derive parameters and errors $\theta_i \pm \sigma_i$
- Compute
  - S = mean of σ_i^2
  - V = variance of θ_i
- Compute total variance,

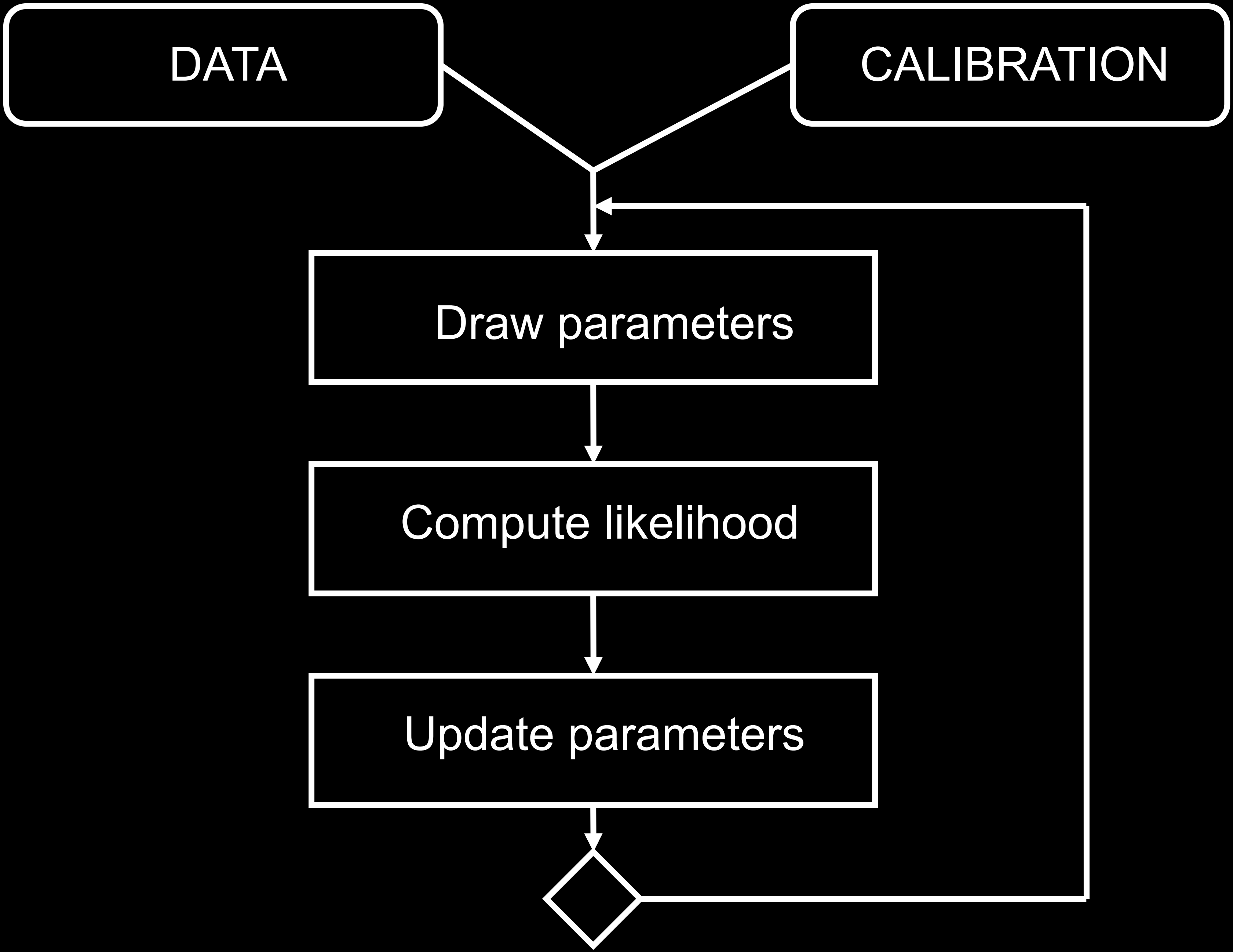
$$\Sigma = S + (1+1/N)*V$$

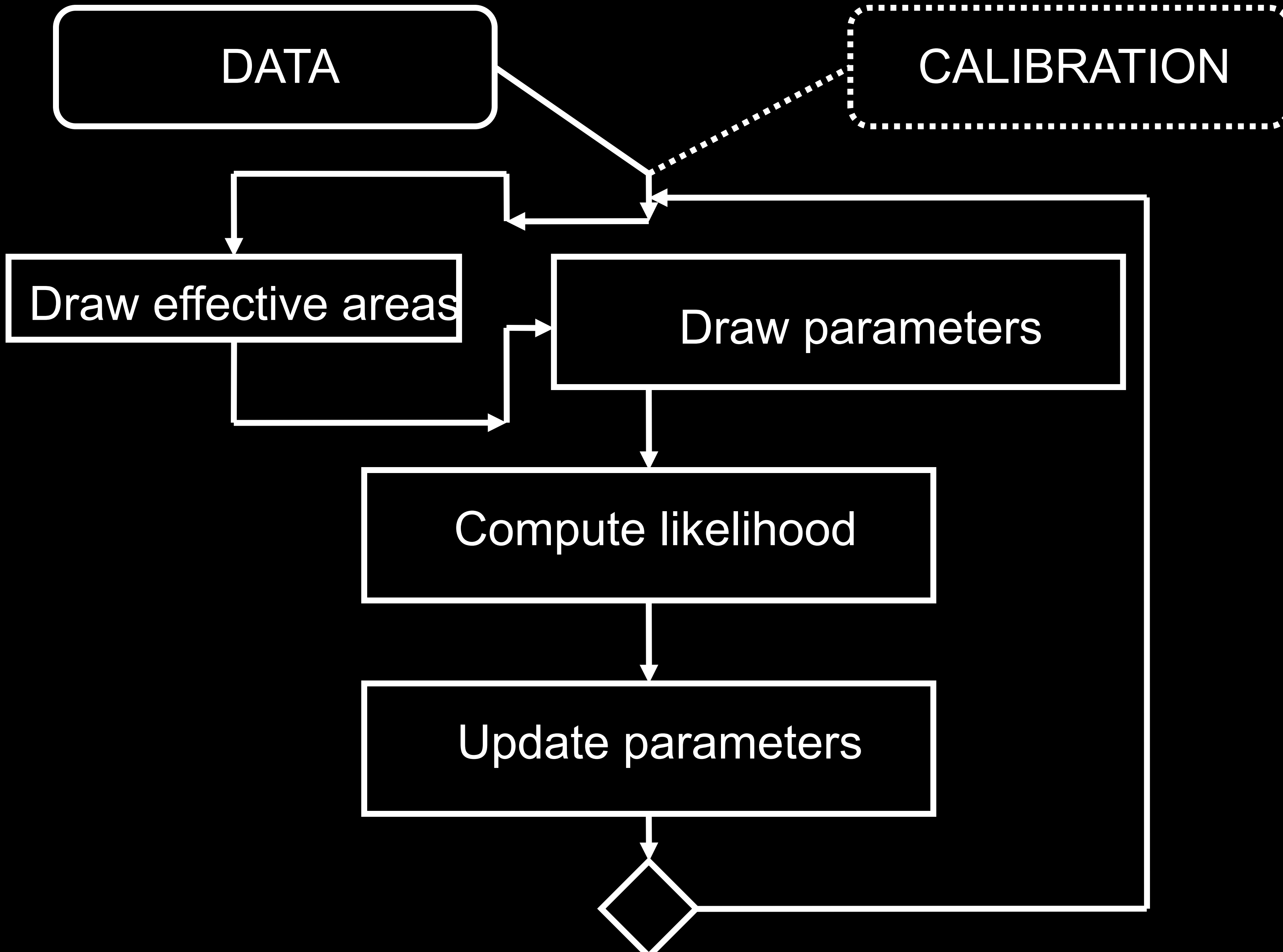
But..

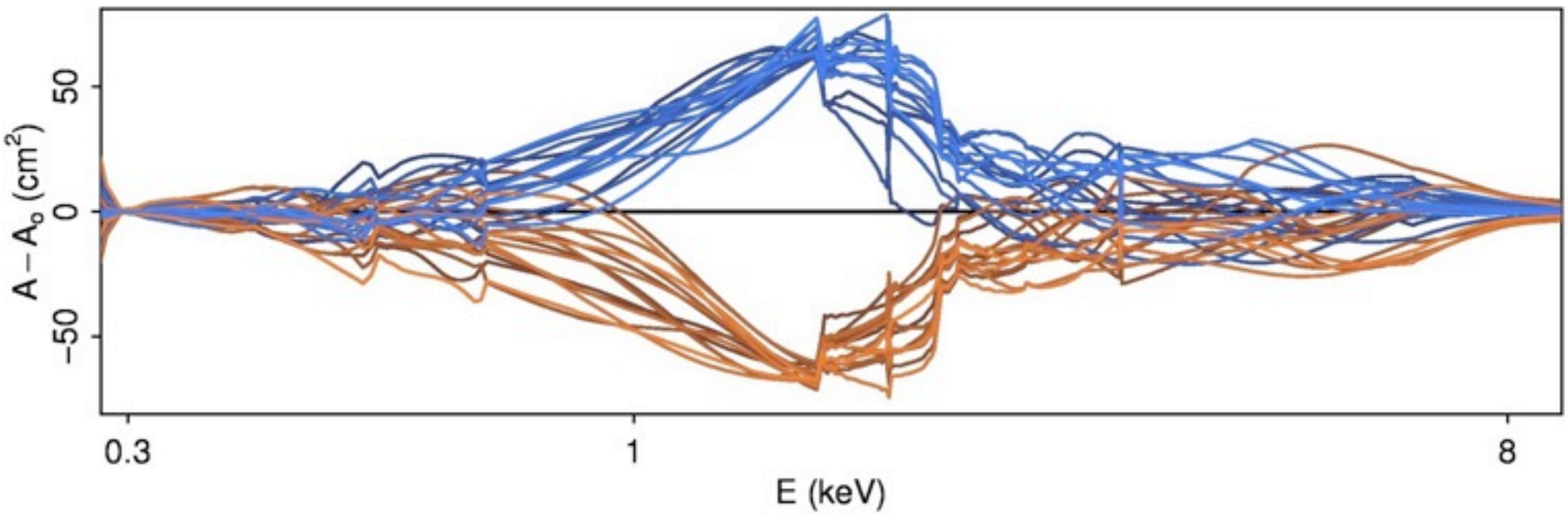
- assumes that
  - errors are symmetrical
  - error distributions are Gaussian
  - all fits are “good” fits
  - all replications are equally weighted
- “1sigma” is not necessarily a 68% interval

MCMC based fitting

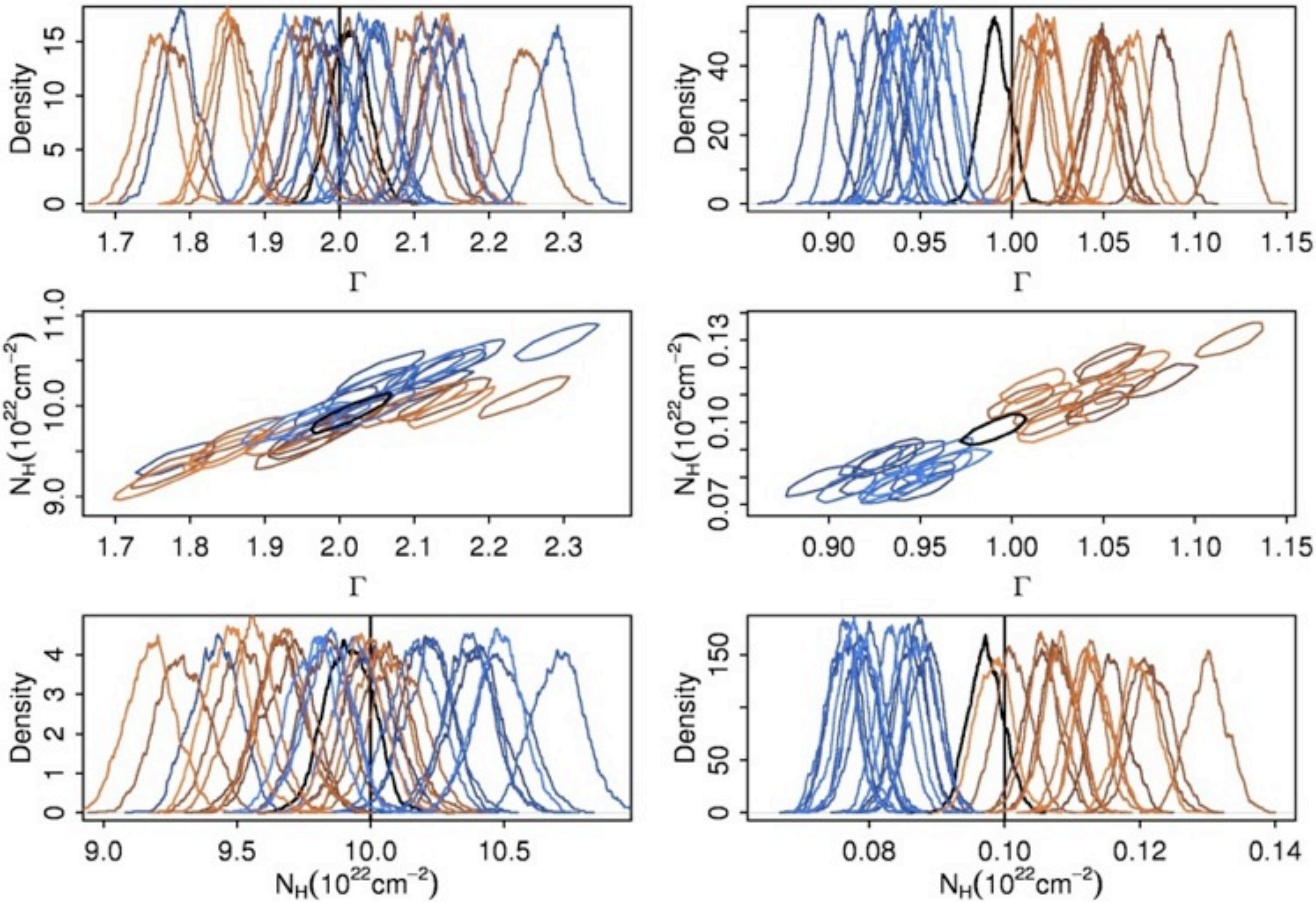
- Usual MCMC based fitting
  - construct a random chain of parameter values by intelligently directing the exploration of the parameter space
  - new parameter values are kept or discarded after comparing to most recent
- the wrinkle
  - randomly vary the cal product at each iteration step

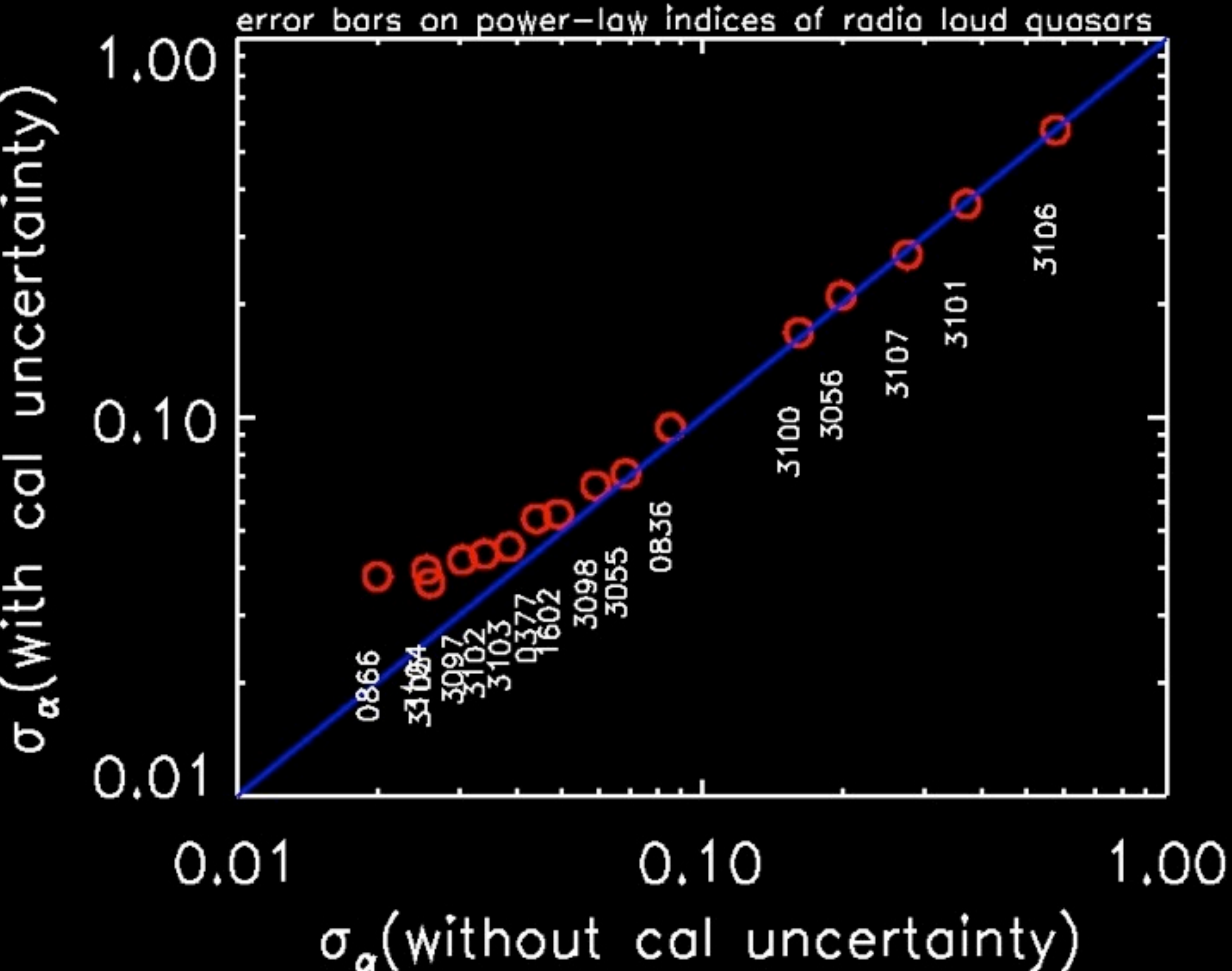

DATA
CALIBRATION
Draw parameters
Compute likelihood
Update parameters


DATA
CALIBRATION
Draw effective areas
Draw parameters
Compute likelihood
Update parameters



50
0
−50
A − A₀ (cm²)
0.3
1
8
E (keV)


Density
Γ
$N_H(10^{22}cm^{-2})$
1.7 1.8 1.9 2.0 2.1 2.2 2.3
0.90 0.95 1.00 1.05 1.10 1.15
9.0 9.5 10.0 10.5
0.08 0.10 0.12 0.14



error bars on power-law indices of radio loud quasars
σ_α(with cal uncertainty)
σ_α(without cal uncertainty)
1.00
0.10
0.01
0.01
0.10
1.00
0866
3097
3102
3103
0377
1602
3098
3055
0836
3100
3056
3107
3101
3106

the good, the bad, and the ugly

✓ the compleat solution

fast

robust to fluctuations in data

can even work in reverse to select best cal product

✷ requires MCMC

convergence not guaranteed

cannot be used as a black box

➡ where you gonna get the random cal product from?

$$A = A_0 + \text{bias} + \text{components} + \text{residual}$$

$A = A_0 + \text{bias} + \text{components} + \text{residual}$

store in same
format as A_0

e.g.,
SPECRESP

$$A = A_0 + \text{bias} + \text{components} + \text{residual}$$

store in same format as A_0

e.g.,
SPECRESP

case specific secondary FITS extension

e.g.,
SAMPLE
PCA1DADD
PCA1DMUL
POLY1D
SPL1DMUL
MULTSCAL

this is where you come in

- AREF

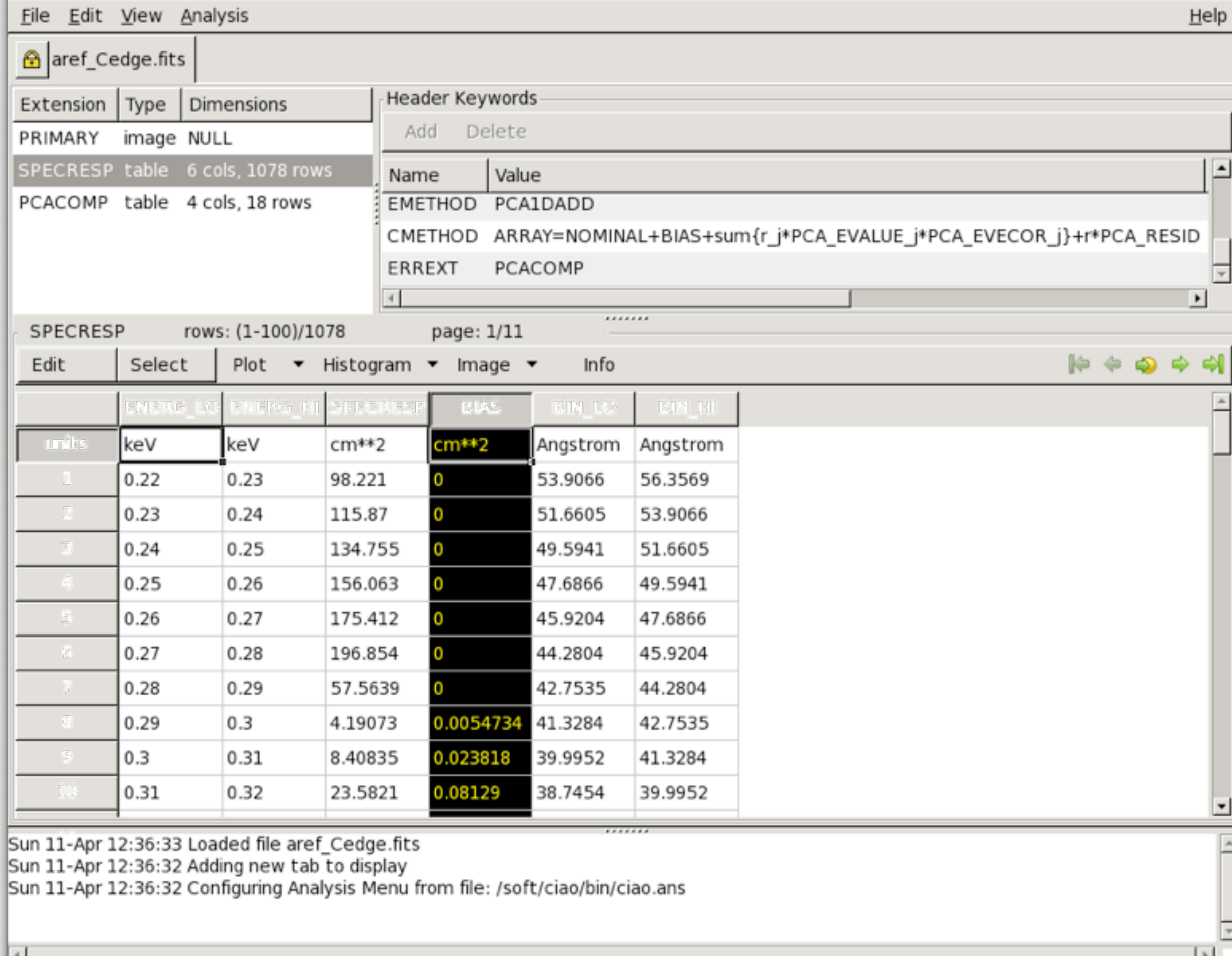
  like ARF, on steroids

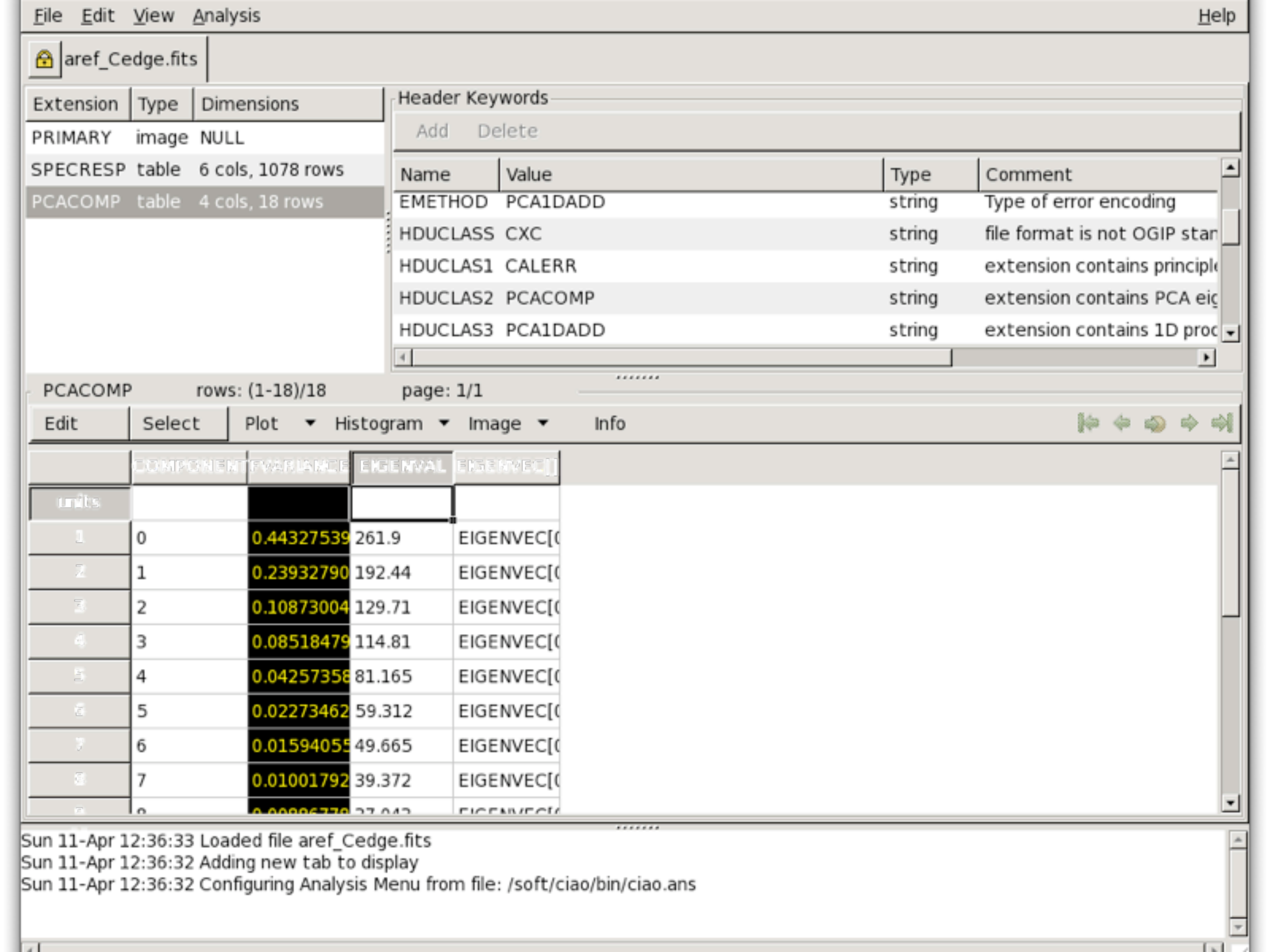
- basic assumption is that the calibration may be time and location dependent, but the uncertainty on it is not

  make once and store

- support for a wide range of encoding

  PCA, splines, samples, etc.


File Edit View Analysis Help
aref_Cedge.fits
Extension Type Dimensions
PRIMARY image NULL
SPECRESP table 6 cols, 1078 rows
PCACOMP table 4 cols, 18 rows
Header Keywords
Add Delete
Name Value
EMETHOD PCA1DADD
CMETHOD ARRAY=NOMINAL+BIAS+sum{r_j*PCA_EVALUE_j*PCA_EVECOR_j}+r*PCA_RESID
ERREXT PCACOMP
SPECRESP rows: (1-100)/1078 page: 1/11
Edit Select Plot Histogram Image Info
ENERG_LO ENERG_HI SPECRESP BIAS BIN_LO BIN_HI
units keV keV cm**2 cm**2 Angstrom Angstrom
1 0.22 0.23 98.221 0 53.9066 56.3569
2 0.23 0.24 115.87 0 51.6605 53.9066
3 0.24 0.25 134.755 0 49.5941 51.6605
4 0.25 0.26 156.063 0 47.6866 49.5941
5 0.26 0.27 175.412 0 45.9204 47.6866
6 0.27 0.28 196.854 0 44.2804 45.9204
7 0.28 0.29 57.5639 0 42.7535 44.2804
8 0.29 0.3 4.19073 0.0054734 41.3284 42.7535
9 0.3 0.31 8.40835 0.023818 39.9952 41.3284
10 0.31 0.32 23.5821 0.08129 38.7454 39.9952
Sun 11-Apr 12:36:33 Loaded file aref_Cedge.fits
Sun 11-Apr 12:36:32 Adding new tab to display
Sun 11-Apr 12:36:32 Configuring Analysis Menu from file: /soft/ciao/bin/ciao.ans


File Edit View Analysis Help
aref_Cedge.fits
Extension Type Dimensions
PRIMARY image NULL
SPECRESP table 6 cols, 1078 rows
PCACOMP table 4 cols, 18 rows
Header Keywords
Add Delete
Name Value Type Comment
EMETHOD PCA1DADD string Type of error encoding
HDUCLASS CXC string file format is not OGIP star
HDUCLAS1 CALERR string extension contains principl
HDUCLAS2 PCACOMP string extension contains PCA eig
HDUCLAS3 PCA1DADD string extension contains 1D proc
PCACOMP rows: (1-18)/18 page: 1/1
Edit Select Plot Histogram Image Info
COMPONENT FVARIANCE EIGENVAL EIGENVEC[]
units
0 0.44327539 261.9 EIGENVEC[(
1 0.23932790 192.44 EIGENVEC[(
2 0.10873004 129.71 EIGENVEC[(
3 0.08518479 114.81 EIGENVEC[(
4 0.04257358 81.165 EIGENVEC[(
5 0.02273462 59.312 EIGENVEC[(
6 0.01594055 49.665 EIGENVEC[(
7 0.01001792 39.372 EIGENVEC[(
Sun 11-Apr 12:36:33 Loaded file aref_Cedge.fits
Sun 11-Apr 12:36:32 Adding new tab to display
Sun 11-Apr 12:36:32 Configuring Analysis Menu from file: /soft/ciao/bin/ciao.ans

coming soon to a console near you

pyBLoCXS

(not just for cal)

load_arf()

RMFs,PSFs,ATOMDB

bottom line

there is a way
to include calibration uncertainty
in astrophysical data analysis
in a flexible way
for any instrument, mission, or detector.