Applying Bayesian techniques to probe AGN geometry

Santiago de Chile / Mar 2015

Johannes Buchner / MPE

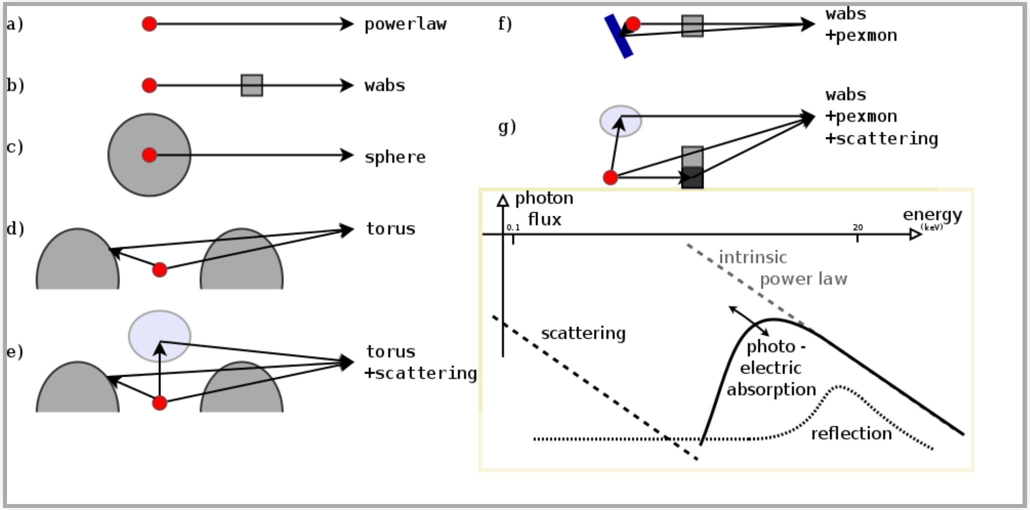
in collaboration with A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni and M. Salvato

Buchner et al. 2014 - **arxiv:1402.0004**

Problem

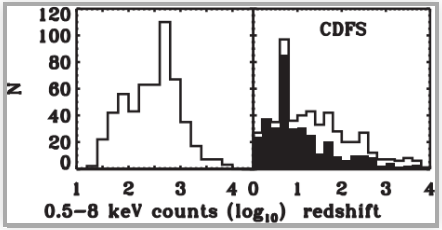
- at z=0.5-3 : peak of star formation & AGN activity
- What is the right spectral model of AGN
  - To derive luminosities, N_H, ...
  - and to understand the "average" AGN
- local AGN: very complex spectra

Set-up

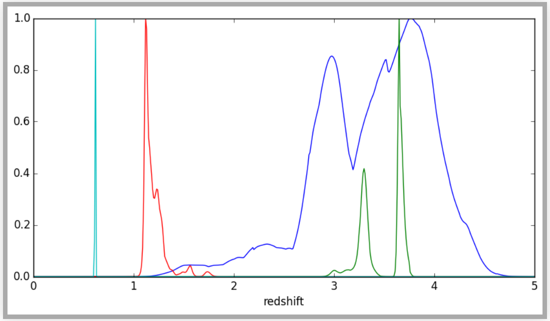


Johannes Buchner

Data: CDFS



Brightman et al. 2014

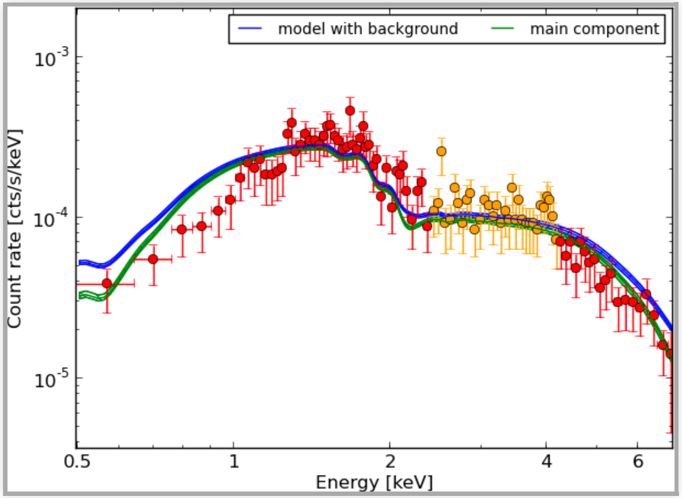


Hsu et al. 2014

Johannes Buchner

Example - Source 179 z=0.605, 2485 counts

powerlaw

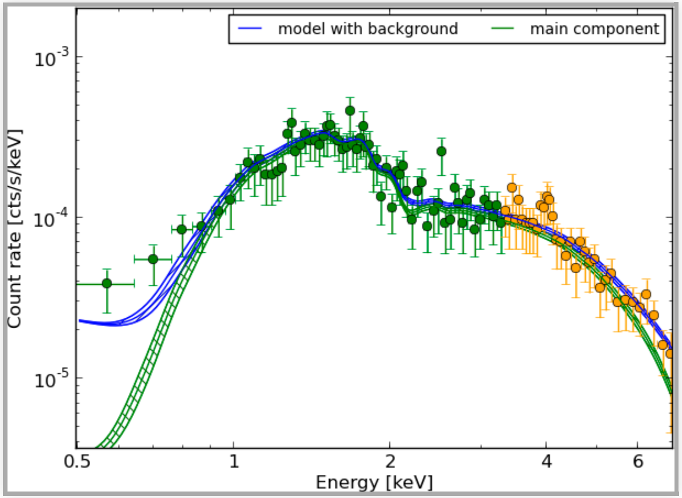


$$\log Z = -64.3, N_H = 20$$

Johannes Buchner

Example - Source 179 z=0.605, 2485 counts

wabs

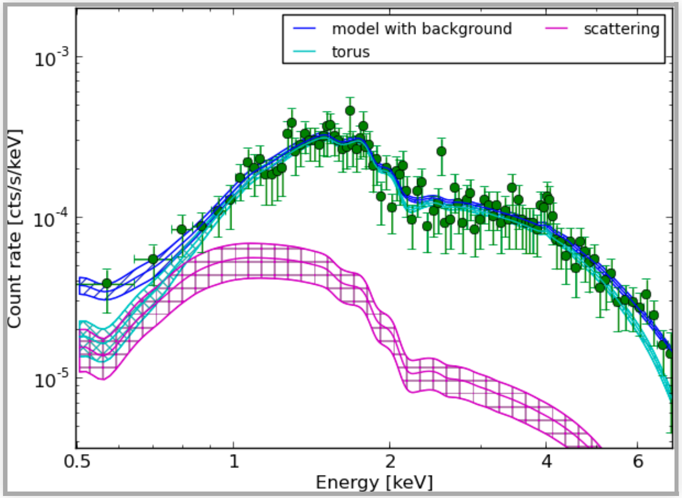


$$\log Z = -12.7, N_H = 22.28 \pm 0.04$$

Johannes Buchner

Example - Source 179 z=0.605, 2485 counts

torus+scattering

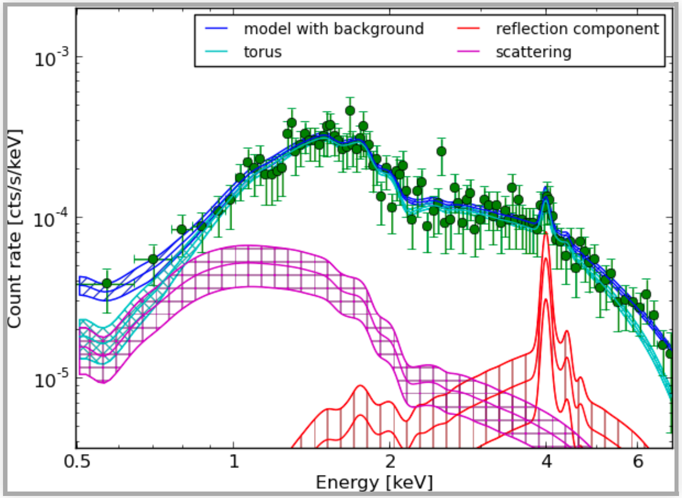


$$\log Z = -4.7, N_H = 22.45 \pm 0.05$$

Johannes Buchner

Example - Source 179 z=0.605, 2485 counts

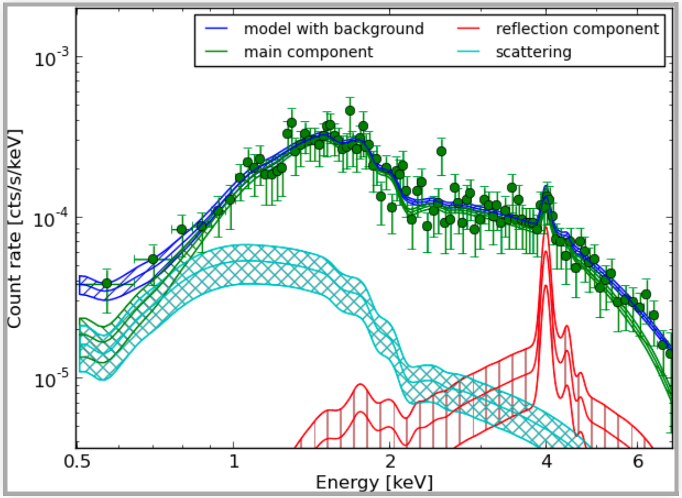
torus+pexmon+scattering



$$\log Z = 0.0, N_H = 22.44 \pm 0.05$$

Example - Source 179 z=0.605, 2485 counts

wabs+pexmon+scattering



$$\log Z = -0.3, N_H = 22.49 \pm 0.05$$

Johannes Buchner

Model comparison

Which physical effects are important? Which model is probably the right one for this data?

- Goodness of Fit?
  - Does the model "match" the data?
  - Requires binning
  - Can not decide between two models which "match".
- Likelihood ratio?
  - Tells probability to mistake one model for another model
  - Requires simulation in every situation

Another way: Bayesian inference

- For each model, compute integral Z of Poisson likelihood (C-stat) over parameter space
- (instead of just maximum likelihood, minimal C-stat)

$$\frac{p(M_1|D)}{p(M_2|D)} = \frac{p(M_1)}{p(M_2)} \frac{Z_1}{Z_2}$$

if all equally probable without data:

$$p(M_1|D) = \frac{Z_1}{\sum Z_i}$$

Parameters: of torus+scattering

- Normalisation ~ luminosity
- Column density $N_H = 10^{20-26}\mathrm{cm}^{-2}$
- Soft scattering normalisation: $f_{scat} < 10\%$
- Photon index: Γ
- Redshift

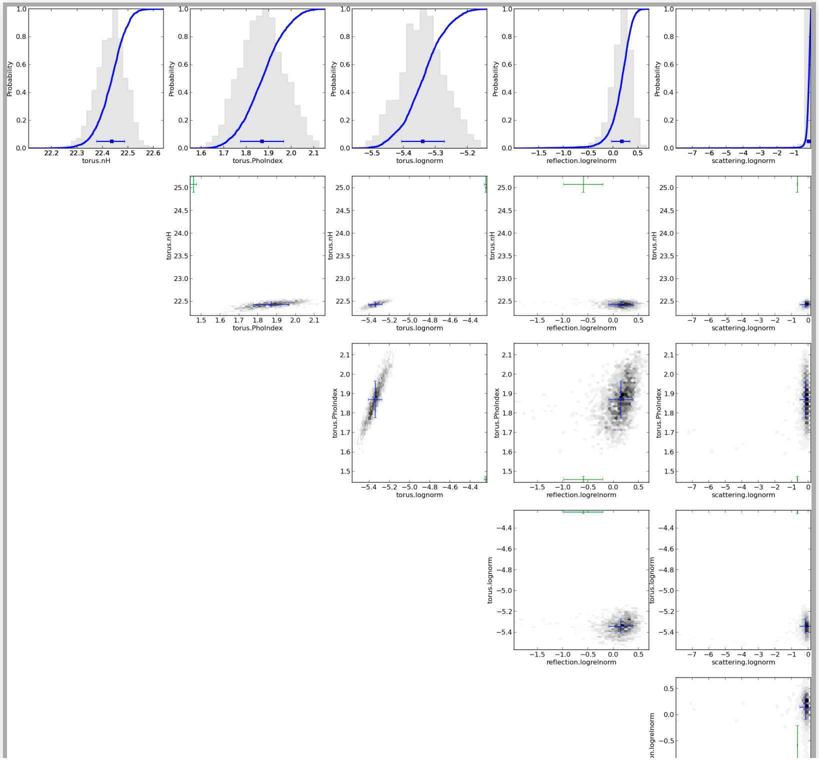
Parameters: of torus+scattering

- Uniform: log Normalisation ~ log Luminosity
- Uniform: Column density $\log N_H/\mathrm{cm}^{-2} = 20 - 26$
- Uniform: Soft scattering normalisation: $\log f_{scat} = 10^{-10...-1}$
- Photon index: $\Gamma \sim 1.95 \pm 0.15$ (Nandra&Pounds 1994)
- Redshift: from pdz

Model comparison results

Model (1)	$L_{2-10\mathrm{keV}}$ (2)	Γ (3)	N_H (4)	$\mathrm{KL}\|_{N_H}$ (5)	$\log Z$ (6)	$p(M\|D)$ (7)
`torus+pexmon+scattering`	$43.35^{+0.01}_{-0.01}$	$1.87^{+0.1}_{-0.1}$	$22.44^{+0.05}_{-0.06}$	1.43	0.0	40.0%
`sphere+pexmon+scattering`	$43.35^{+0.01}_{-0.01}$	$1.88^{+0.1}_{-0.1}$	$22.44^{+0.05}_{-0.05}$	1.46	-0.1	30.2%
`wabs+pexmon+scattering`	$43.34^{+0.01}_{-0.01}$	$1.88^{+0.1}_{-0.1}$	$22.49^{+0.05}_{-0.06}$	1.44	-0.1	29.0%
`torus+scattering`	$43.39^{+0.01}_{-0.01}$	$1.76^{+0.1}_{-0.1}$	$22.45^{+0.05}_{-0.06}$	1.42	-2.0	0.4%
`sphere+scattering`	$43.39^{+0.01}_{-0.01}$	$1.77^{+0.1}_{-0.1}$	$22.46^{+0.05}_{-0.06}$	1.45	-2.1	0.3%
`wabs+scattering`	$43.39^{+0.01}_{-0.01}$	$1.75^{+0.1}_{-0.1}$	$22.51^{+0.05}_{-0.06}$	1.44	-2.7	0.1%
`sphere`	$43.35^{+0.01}_{-0.01}$	$1.57^{+0.1}_{-0.1}$	$22.23^{+0.05}_{-0.04}$	1.52	-4.7	0.0%
`torus`	$43.35^{+0.01}_{-0.01}$	$1.56^{+0.1}_{-0.1}$	$22.22^{+0.05}_{-0.04}$	1.54	-4.7	0.0%
`wabs`	$43.35^{+0.01}_{-0.01}$	$1.55^{+0.1}_{-0.1}$	$22.28^{+0.04}_{-0.04}$	1.58	-5.5	0.0%
`powerlaw`	$43.35^{+0.01}_{-0.01}$	$0.80^{+0.0}_{-0.0}$			-27.9	0.0%

Model comparison results



Johannes Buchner

Model comparison

- Can distinguish some models, others are equally probable
- Have made use of
  - local AGN information $\leftrightarrow \Gamma$
  - other λ information $\leftrightarrow z$
- Have derived parameter distributions and Z simultaneously
- incorporating uncertainty in Γ, z
- Likelihood averaging punishes model diversity not realised in the data
- Can tell that it does not know

How can geometries be distinguished?

- How to increase strength of data?
- Combine objects
- Traditionally: stacking
  - lose information in averaging
  - averaging of different redshifts?
- Here, multiply Z to combine information

$$Z_1 = \prod_{\mu} Z_1^{\mu}$$

- Buchner et al. 2014 - http://arxiv.org/abs/1402.0004

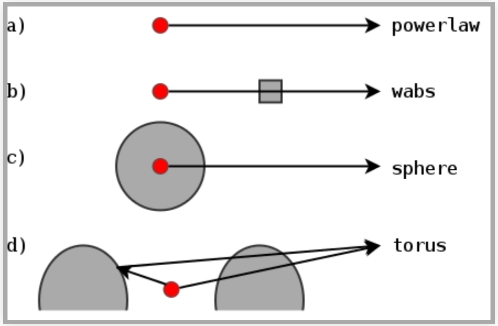
Results for the full CDFS sample

Model	Sample results		Bootstrapped results		
	# rej	$\sum \log Z$	# rej	$\sum \log Z$	ruled out
(1)	(2)	(3)	(4)	(5)	(6)
All (334 sources)					
`torus+pexmon+scattering`	11	−85.6	10 ± 3.4	−85.4 ± 8.0	9 ± 28%
`wabs+pexmon+scattering`	8	−93.4	8 ± 2.7	−93.6 ± 7.6	85 ± 35%
`sphere+pexmon+scattering`	11	−95.3	10 ± 3.3	−95.1 ± 7.8	100 ± 5%
`torus+scattering`	16	−116.8	16 ± 4.0	−117.3 ± 11.8	100 ± 0%
`sphere+scattering`	21	−127.3	21 ± 4.6	−127.3 ± 10.3	100 ± 0%
`wabs+scattering`	37	−167.6	37 ± 5.7	−168.6 ± 14.2	100 ± 0%
`torus`	39	−188.2	39 ± 5.9	−190.4 ± 20.0	100 ± 0%
`sphere`	55	−233.1	55 ± 6.8	−235.3 ± 22.3	100 ± 0%
`wabs`	59	−270.9	59 ± 6.8	−274.1 ± 26.9	100 ± 0%
`powerlaw`	187	−2657.9	186 ± 9.6	−2704.8 ± 412.6	100 ± 0%
log N_{H} = 24–26 *(14 sources)*					
`torus+scattering`	0	−1.5	0 ± 0.0	−1.5 ± 0.5	7 ± 26%
`torus+pexmon+scattering`	0	−1.7	0 ± 0.0	−1.7 ± 0.3	7 ± 26%
`torus`	3	−9.8	3 ± 2.1	−10.3 ± 7.1	93 ± 26%
z > 1 (229 sources)					
`torus+pexmon+scattering`	8	−57.6	8 ± 3.0	−58.9 ± 7.7	11 ± 32%
`wabs+pexmon+scattering`	5	−64.5	5 ± 2.3	−65.2 ± 7.7	80 ± 40%
`torus+scattering`	7	−67.7	7 ± 2.5	−68.5 ± 7.7	100 ± 7%

Johannes Buchner

Conclusions

- Absorbed powerlaw necessary
- Soft scattering powerlaw detected
- Additional Compton-reflection detected
- Torus model preferred (also in CT AGN)



Practicalities

- **https://github.com/JohannesBuchner/BXA**
- uses MultiNest (Feroz&Hobson 2008)
- via PyMultiNest

Fitting in PyXspec

```
from xspec import *
Fit.statMethod = 'cstat'
Plot.xAxis = 'keV'
s = Spectrum('example-file.fak')
s.notice("0.2-8.0")

m = Model("pow")
# set parameters range : val, delta, min, bottom, top, max
m.powerlaw.norm.values = ",,1e-10,1e-10,1e1,1e1" # 10^-10 .. 10
m.powerlaw.PhoIndex.values = ",,1,1,5,5"         #      1 .. 3

m.fit()
```

Johannes Buchner

Analysis in PyXspec

```
from xspec import *
Fit.statMethod = 'cstat'
Plot.xAxis = 'keV'
s = Spectrum('example-file.fak')
s.ignore("**"); s.notice("0.2-8.0")

m = Model("pow")
# set parameters range : val, delta, min, bottom, top, max
m.powerlaw.norm.values = ",,1e-10,1e-10,1e1,1e1" # 10^-10 .. 10
m.powerlaw.PhoIndex.values = ",,1,1,5,5"         #      1 .. 3

import bxa.xspec as bxa
transformations = [
        bxa.create_uniform_prior_for( m, m.powerlaw.PhoIndex),
        bxa.create_jeffreys_prior_for(m, m.powerlaw.norm)      ]

bxa.standard_analysis(transformations,
        outputfiles_basename = 'simplest-',)
```

Summary

- Bayesian inference is useful for
  - estimating physical parameters (yields probability distributions)
  - comparing physical models
- Drawbacks: Computational cost
- Geometry of AGN obscurer is a torus (not a sphere, not a thin disk)
- Additional components
  - Soft scattering powerlaw -- Thomson scattering off ionised clouds?
  - Compton reflection -- Accretion disk? Thicker parts of the torus?

Backup slides

Johannes Buchner

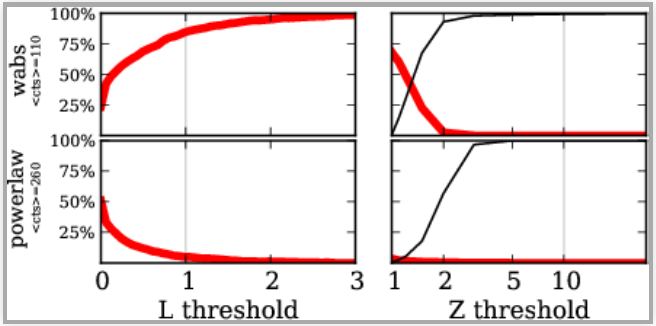
Pragmatic viewpoint

distinguish two models via data can use any statistic

does not have to be probabilistic

- can be counts in the 6keV bin
- determine discriminating threshold via simulations
  - false association rate ($B \rightarrow A, A \rightarrow B$)
  - correct association rate ($A \rightarrow A, B \rightarrow B$)

Comparison $\hat{L}$ vs. Z



red: falsely choose powerlaw, for wabs input

red: falsely choose wabs, for powerlaw input

(Appendix 2)

Z more effective than $\hat{L}$

what is Z? Integral of likelihood / average

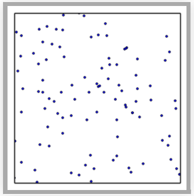
Bayesian evidence

$$\frac{P(A|D)}{P(B|D)} = \frac{P(A)}{P(B)} \times \frac{Z_A}{Z_B}$$

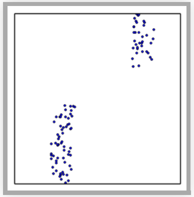
interpretation of Z-ratio under flat priors:

- prob. that this model is the right one, rather than the other.
- A, B or equal

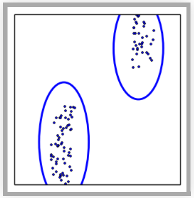
Nested Sampling



draw randomly uniformly 200 points
always remove least likely point, replace with a new draw of higher likelihood



converges to maximum likelihood, stops when flat



how to draw new points efficiently?
MultiNest does it via clustering and ellipses

Nested Sampling with MultiNest

explores the problem in

- high dimensions (3-20)
- handles multiple maxima
- handles peculiar shapes
- runs efficiently to convergence (typically 10000-40000 points)

measures and describes shapes (like MCMC)

Connecting

Need to connect C-stat calculation with algorithm

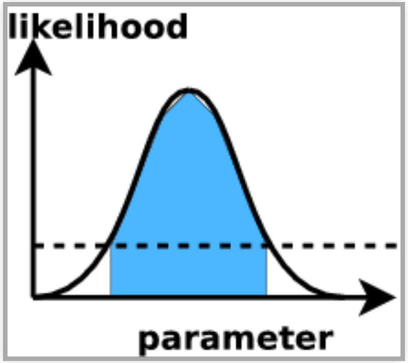
BXA: Bayesian X-ray Analysis

- MultiNest for Sherpa / (Py)Xspec
- installed on ds42 under `/utils/bxa`
- documentation: **github.com/JohannesBuchner/BXA**

Analysis

Likelihood value evaluated "everywhere"

- ML analysis: find confidence intervals
  - defined via: how often the estimator (maximum) gives the right answer
  - different for different estimators - property of the method
- credible intervals
  - defined via: prob. that the true value is inside this range rather than outside is x%.

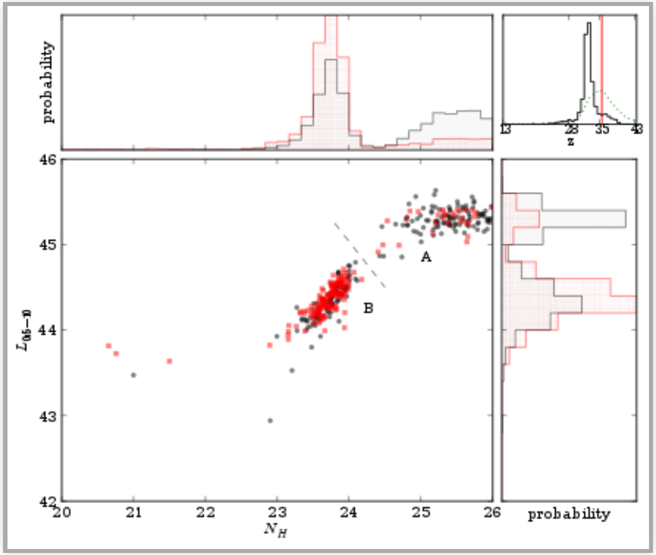


results coincide for some choice of prior (usually "flat").

How?

posterior "chain" from MCMC/nested sampling: representation through point density

norm	N_H	z
-4.1	22.4	2.3
-4.3	22.45	2.4
-4.3	22.38	2.5
...	...	



just make histogram of 1/2 columns

contains all correlations

Error propagation: example

norm	N_H	z
-4.1	22.4	2.3
-4.3	22.45	2.4
-4.3	22.38	2.5
...	...	

for every parameter vector (norm, N_H, z, ...), set the model

- set $N_H = 0$
- compute intrinsic flux
- compute luminosity using z and flux

---

incorporates uncertainty in z and the parameters!

- just do your calculation with every value instead of one

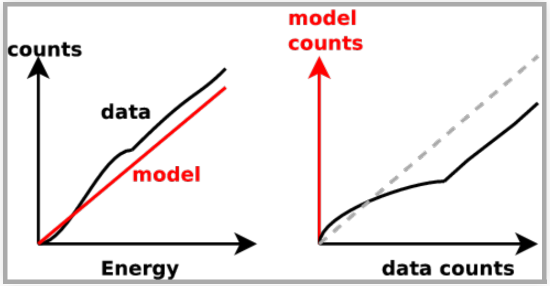
model inadequacy

Have cool tools now, but:

- Is the model right?
- Where is the model wrong?
- Systematic effects?
- Discover new physics beyond the model

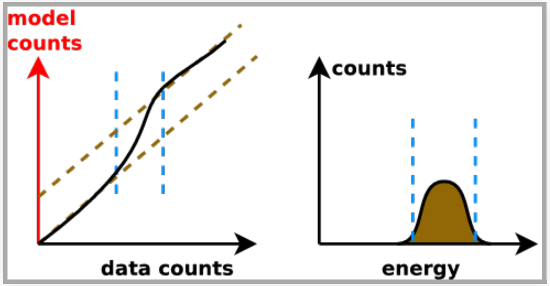
Common route: residuals

New idea: Q-Q plot (no binning) (Appendix 1)

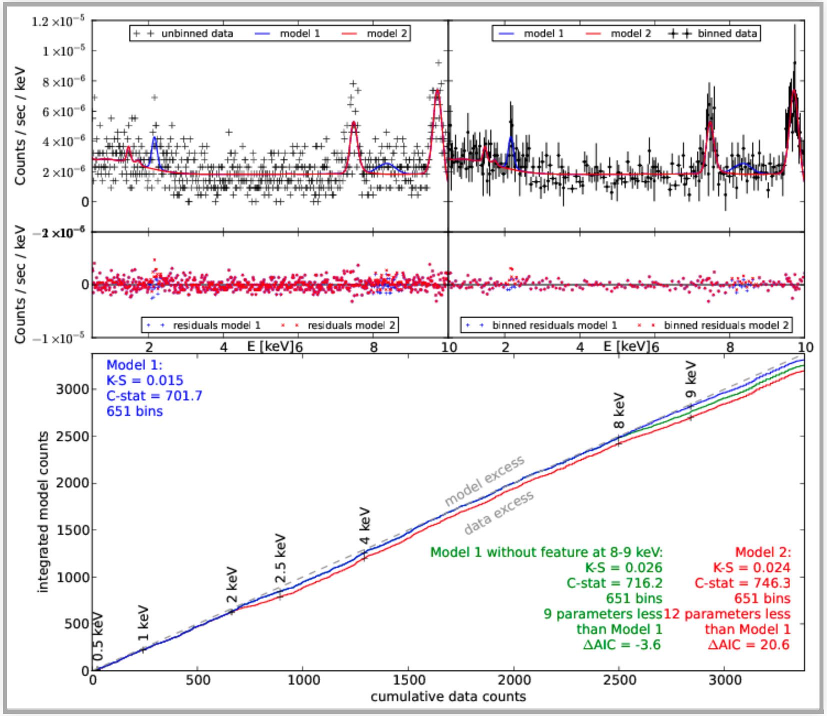


good fit if straight line

Q-Q plot primer

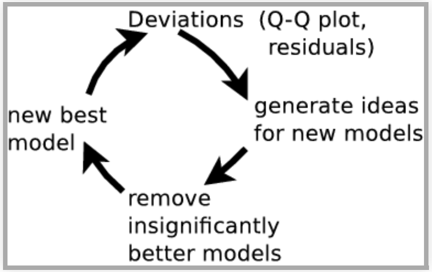


Johannes Buchner



Johannes Buchner

Generate ideas for new models



Johannes Buchner

Summary

- **Parameter estimation**: (see 5.1, Appendix 3)
  explore multiple maxima
  general solution with nested sampling
- **Model comparison**:
  Likelihood ratio is less effective than Z ratios (see 5.2, Appendix 2)
  computed by nested sampling; has right interpretation
- **Model discovery**: (see 5.3, Appendix 1)
  Q-Q plots + model comparison

arxiv:1402.0004 github.com/JohannesBuchner/BXA