BIRS Discovery Workshop - Banff, July 2010

Summary talk: a cosmologist's perspective

Roberto Trotta
Imperial College London
www.robertotrotta.com

Picture credit: Peter Liversidge

Thanks to Louis, Jim and Richard for organizing this workshop!

Lots of interesting talks, stimulating discussions... and some clashes!

A cosmologist's perspective

Imperial College London

- Physicists/statisticians and astro/cosmo vs particle physics
- Look elsewhere effect
- Model selection
- Relevance of 5 sigma for us folks

Particle physicists vs cosmologists/astro

Imperial College London

- **Methodological**
  - Repeatable experiments (counting) vs observations (there is only 1 Universe)
  - Frequentist vs mostly Bayesian
  - Profiling vs marginalizing
  - Priors: "What priors?" vs Often highly relevant prior information
  - Selection effects are usually important in cosmology
  - Combination of probes necessary in cosmology to break degeneracies (problem: what about systematics?)

FIRST-YEAR SLOAN DIGITAL SKY SURVEY-II (SDSS-II) SUPERNOVA RESULTS: HUBBLE DIAGRAM AND COSMOLOGICAL PARAMETERS

Richard Kessler,[1,2] Andrew C. Becker,[3] David Cinabro,[4] Jake Vanderplas,[3] Joshua A. Frieman,[2,1,5] John Marriner,[5] Tamara M Davis,[6,7] Benjamin Dilday,[8] Jon Holtzman,[9] Saurabh W. Jha,[8] Hubert Lampeitl,[10] Masao Sako,[11] Mathew Smith,[1012] Chen Zheng,[32] Robert C. Nichol,[10] Bruce Bassett,[12,13] Ralf Bender,[20] Darren L. Depoy,[14] Mamoru Doi,[15,16] Ed Elson,[12] Alexei V. Filippenko,[17] Ryan J. Foley,[17,18] Peter M. Garnavich,[19] Ulrich Hopp,[20] Yutaka Ihara,[15,21] William Ketzeback,[22] W. Kollatschny,[24] Kohki Konishi,[27] Jennifer L. Marshall,[14] Russet J. McMillan,[22] Gajus Miknaitis,[37,5] Tomoki Morokuma,[23] Edvard Mörtsell,[25] Kaike Pan,[22] Jose Luis Prieto,[26] Michael W. Richmond,[29] Adam G. Riess,[30,31] Roger Romani,[32] Donald P. Schneider,[33] Jesper Sollerman,[7,28] Naohiro Takanashi,[23] Kouichi Tokita,[15,21] Kurt van der Heyden,[35] J. C. Wheeler,[34] Naoki Yasuda,[27] and Donald York[1,36]

Accepted for publication in ApJS

Contributions by particle physcists

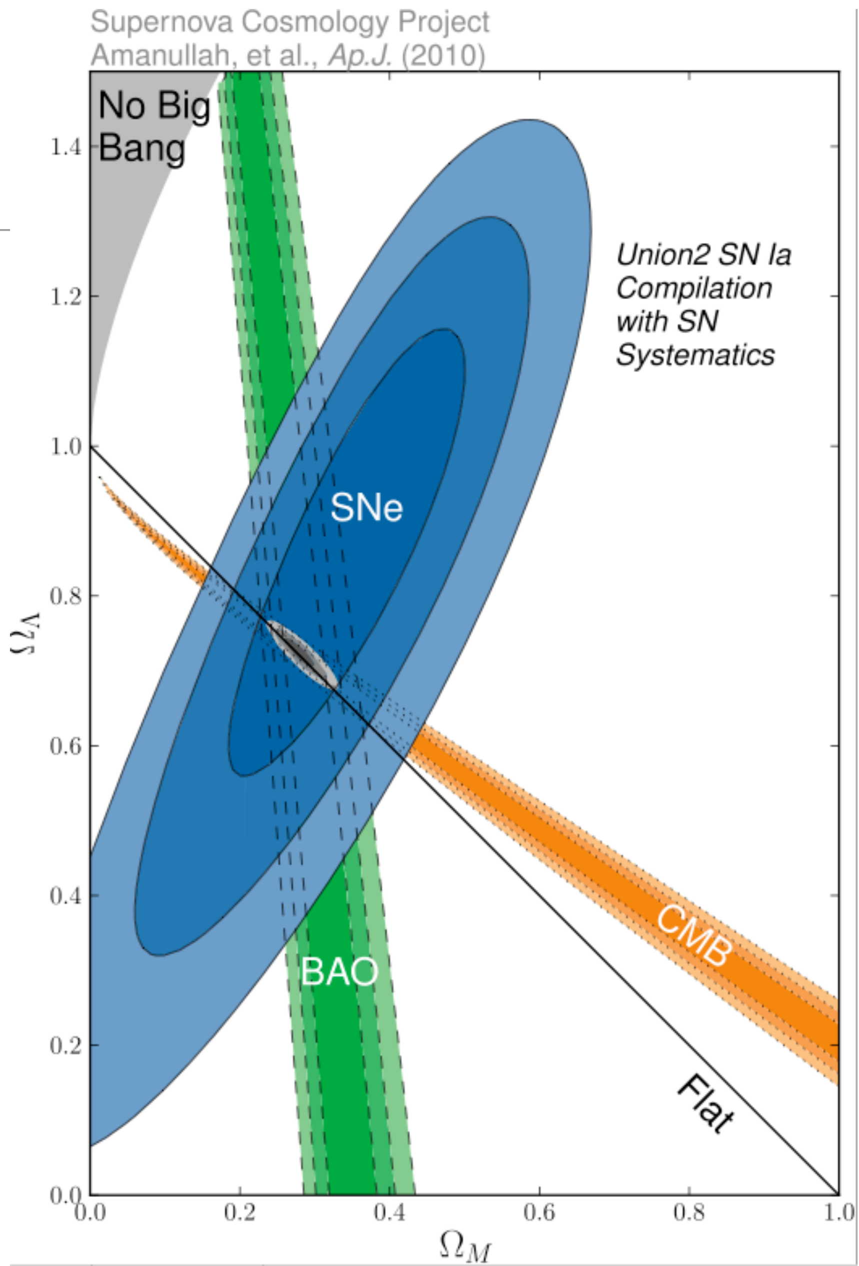
ABSTRACT

We present measurements of the Hubble diagram for 103 Type Ia supernovae (SNe) with redshifts $0.04 < z < 0.42$, discovered during the first season (Fall 2005) of the Sloan Digital Sky Survey-II (SDSS-II) Supernova Survey. These data fill in the redshift "desert" between low- and high-redshift SN Ia surveys. Within the framework of the MLCS2K2 light-curve fitting method, we use the SDSS-II SN sample to infer the mean reddening parameter for host galaxies, $R_V = 2.18 \pm 0.14_{\rm stat} \pm 0.48_{\rm syst}$, and find that the intrinsic distribution of host-galaxy extinction is well fit by an exponential function, $P(A_V) = \exp(-A_V/\tau_V)$, with $\tau_V = 0.334 \pm 0.088$ mag. We combine the SDSS-II measurements with new distance estimates for published SN data from the ESSENCE survey, the Supernova Legacy Survey (SNLS), the Hubble Space Telescope (HST), and a compilation of nearby SN Ia measurements. A new feature in our analysis is the use of detailed Monte Carlo simulations of all surveys to account for selection biases, including those from spectroscopic targeting. Combining the SN Hubble diagram with measurements of baryon acoustic oscillations from the SDSS Luminous Red Galaxy sample and with cosmic microwave background temperature anisotropy measurements from WMAP, we estimate the cosmological parameters w and $\Omega_{\rm M}$, assuming a spatially flat cosmological model (FwCDM) with constant dark energy equation of state parameter, w. We also consider constraints upon $\Omega_{\rm M}$ and Ω_Λ for a cosmological constant model (ΛCDM) with $w = -1$ and non-zero spatial curvature. For the FwCDM model and the combined sample of 288 SNe Ia, we find $w = -0.76 \pm 0.07({\rm stat}) \pm 0.11({\rm syst})$, $\Omega_{\rm M} = 0.307 \pm 0.019({\rm stat}) \pm 0.023({\rm syst})$ using MLCS2K2 and $w = -0.96 \pm 0.06({\rm stat}) \pm 0.12({\rm syst})$, $\Omega_{\rm M} = 0.265 \pm 0.016({\rm stat}) \pm 0.025({\rm syst})$ using the SALT–II fitter. We trace the discrepancy between these results to a difference in the rest-frame UV model combined with a different luminosity correction from color variations; these differences mostly affect the distance estimates for the SNLS and HST supernovae. We present detailed discussions of systematic errors for both light-curve methods and find that they both show data-model discrepancies in rest-frame U-band. For the SALT–II approach, we also see strong evidence for redshift-dependence of the color-luminosity parameter (β). Restricting the analysis to the 136 SNe Ia in the Nearby+SDSS-II samples, we find much better agreement between the two analysis methods but with larger uncertainties: $w = -0.92 \pm 0.13({\rm stat})^{+0.10}_{-0.33}\,({\rm syst})$ for MLCS2K2 and $w = -0.92 \pm 0.11({\rm stat})^{+0.07}_{-0.15}\,({\rm syst})$ for SALT–II.

Precision cosmology

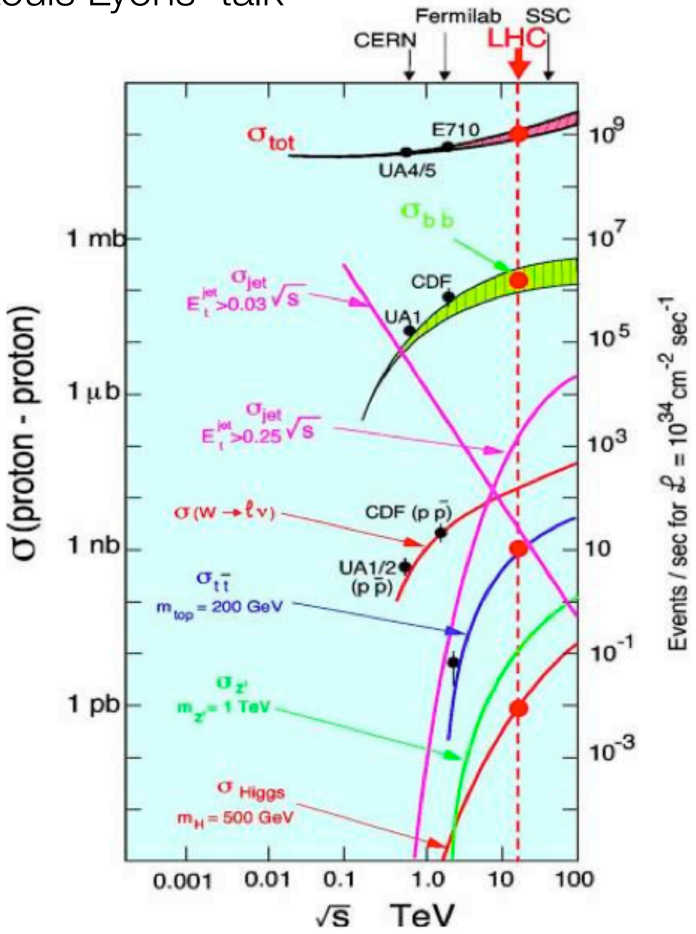
$$\Omega_{cdm}h^2 = 0.1099 \pm 0.0062$$
$$\Omega_m = 0.258 \pm 0.030$$

Those are NOT confidence regions!

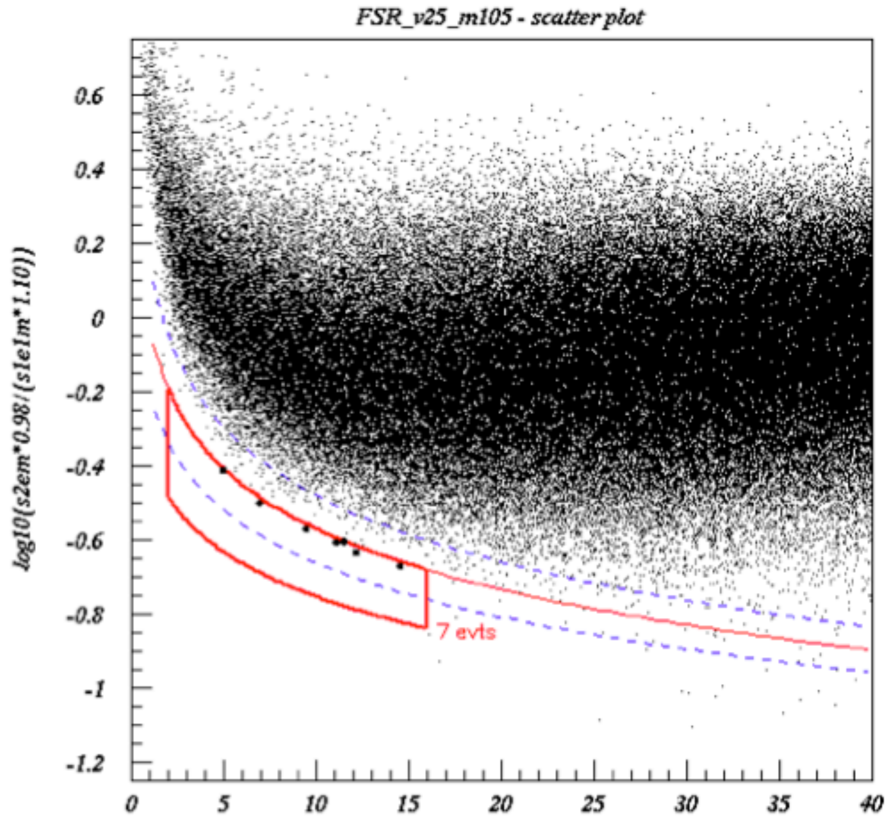


Needle in the haystack

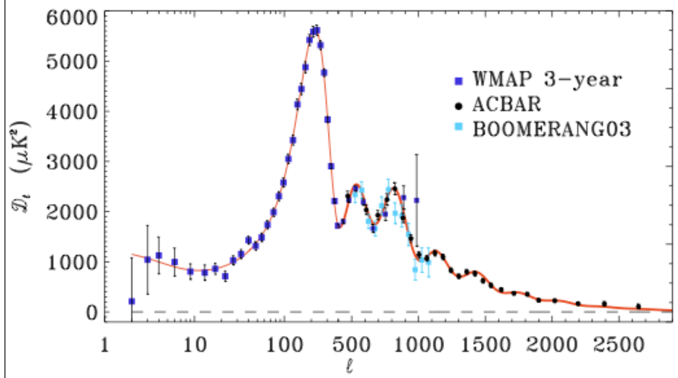
Louis Lyons' talk

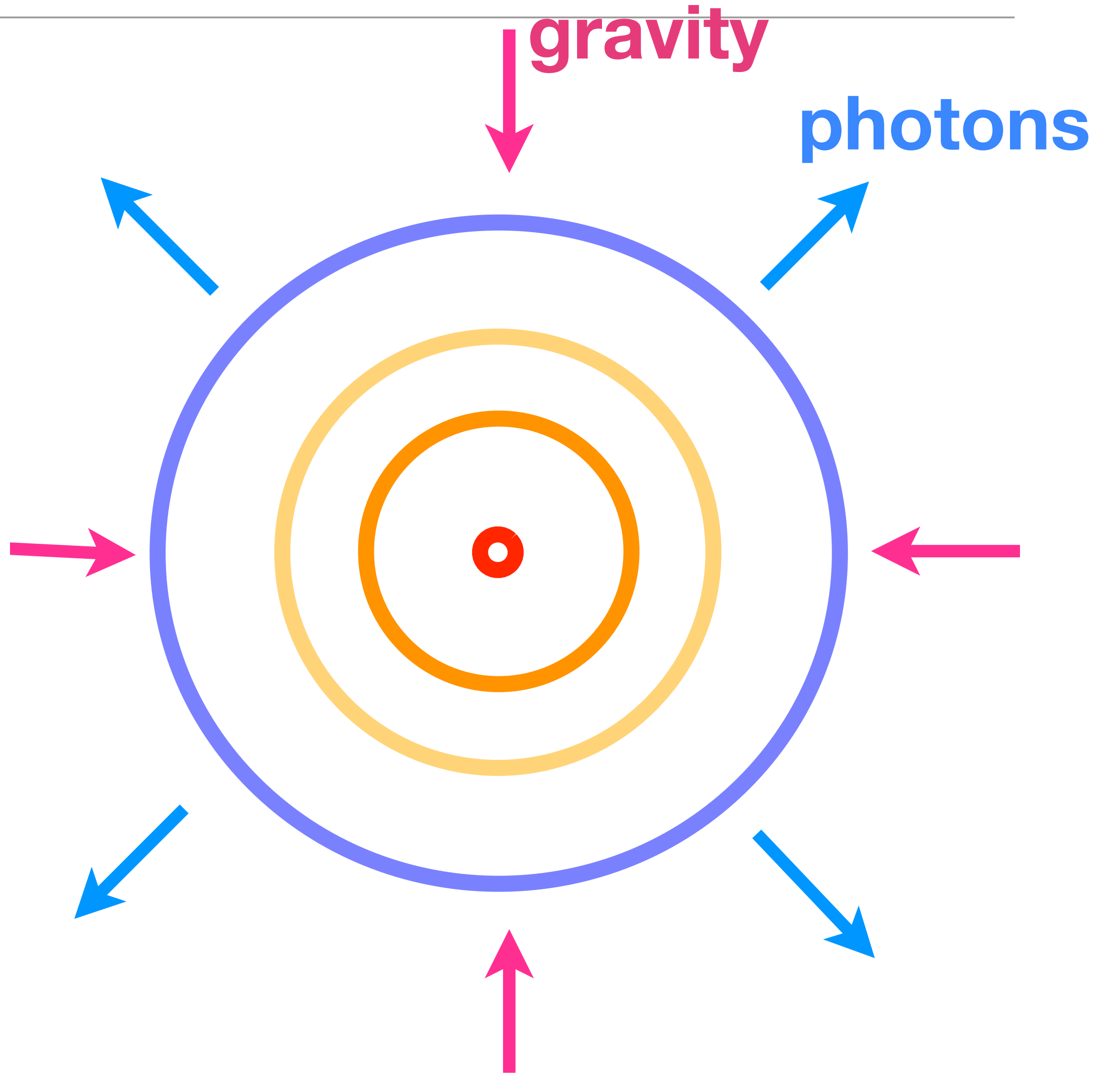


Henrique Araujo's talk



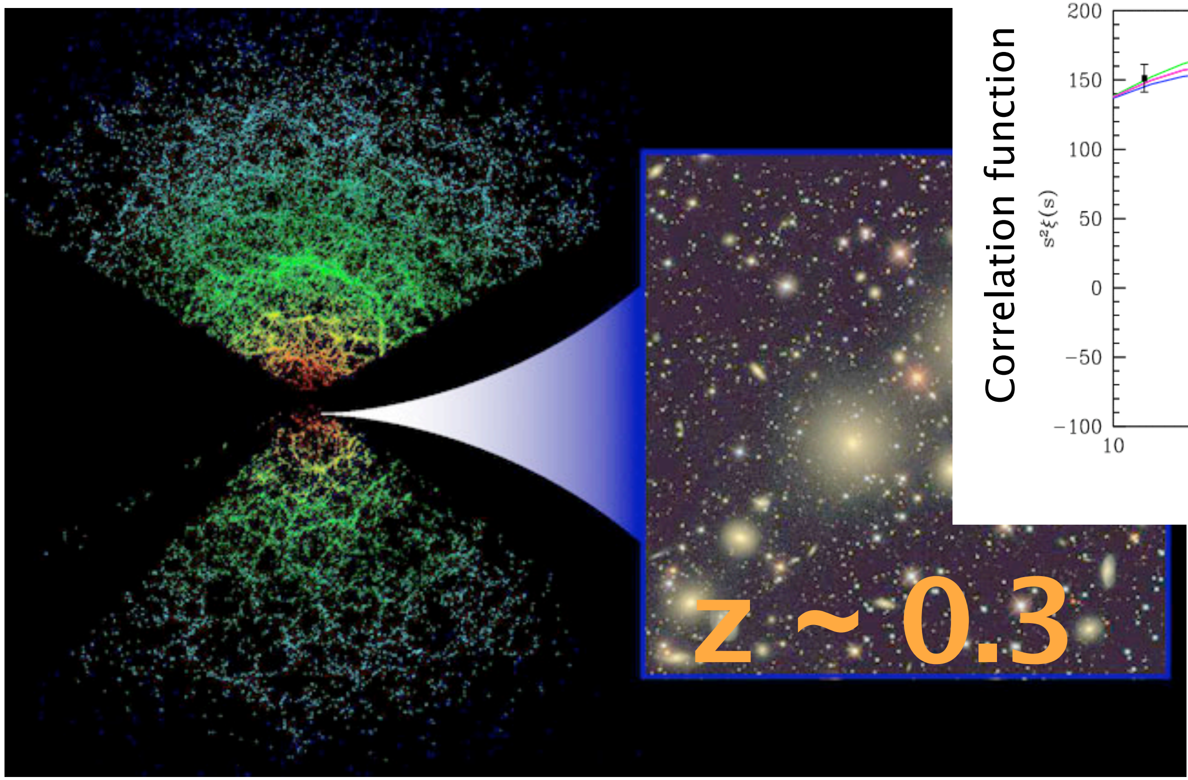
Cosmic sound

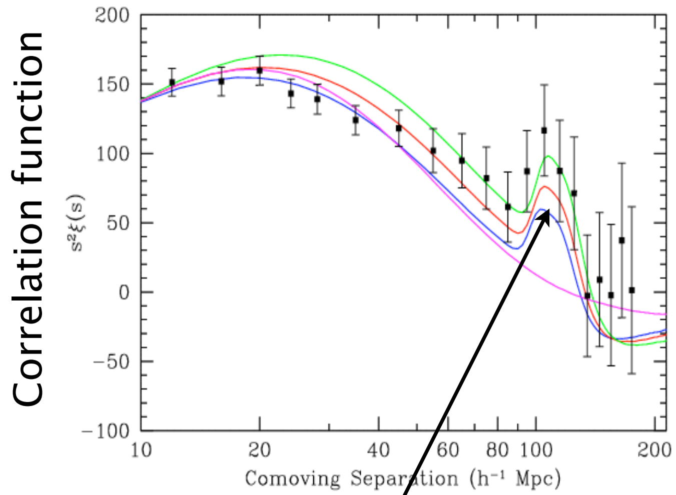




Needle in the haystack - cosmology

based on ~ 50’000 LRG





Baryonic acoustic oscillations signature
~ 2.5 sigma significance

We are looking for extra correlations between galaxies on scales ~ 150 Mpc: this corresponds to (on average) 1 extra galaxy at this preferential separation

Particle physicists vs cosmologists/astro

Imperial College London

- **Epistemological**
  - Particle physicists believe in the existence of THE TRUE MODEL.
  - Cosmology is often more pragmatic: “The cosmological concordance model” is more of a phenomenological description of the data (dark matter/dark energy), not necessarily fundamentally motivated in the same way as particle physics models are.
  - Frequentist error probabilities vs uncertainties representing degree of lack of knowledge/belief.

Particle physicists vs cosmologists/astro

Imperial College London

- **Community**
  - Data often published in summary form (although this is changing -> Kyle Cranmer's talk) vs full data made public (WMAP, SDSS)
  - Large, international collaborations vs smaller, more compact teams (although this is changing: WMAP team ~ 15 people; Sloan ~ 50 people; Planck ~ 500 people, Auger ~ 400 people)
  - Large codes often private vs codes usually made public, community input.
  - Both communities have now meetings where discussions with statisticians are encouraged (PHYSTAT, Banff, cosmostats, ...)

Language barriers: a dictionary

Imperial College London

Physicist		Statistician
cuts	⟷	filter, selection
large samples statistics	⟷	?
measurements	⟷	estimates
events	⟷	events (different meaning)
chi-squared	⟷	weighted sum of squares
look elsewhere	⟷	multiple comparisons
doh!	⟷	measurement error problem
upper limit	⟷	?
not on the boundary	⟷	the null in the interior

All of those came up during this workshop

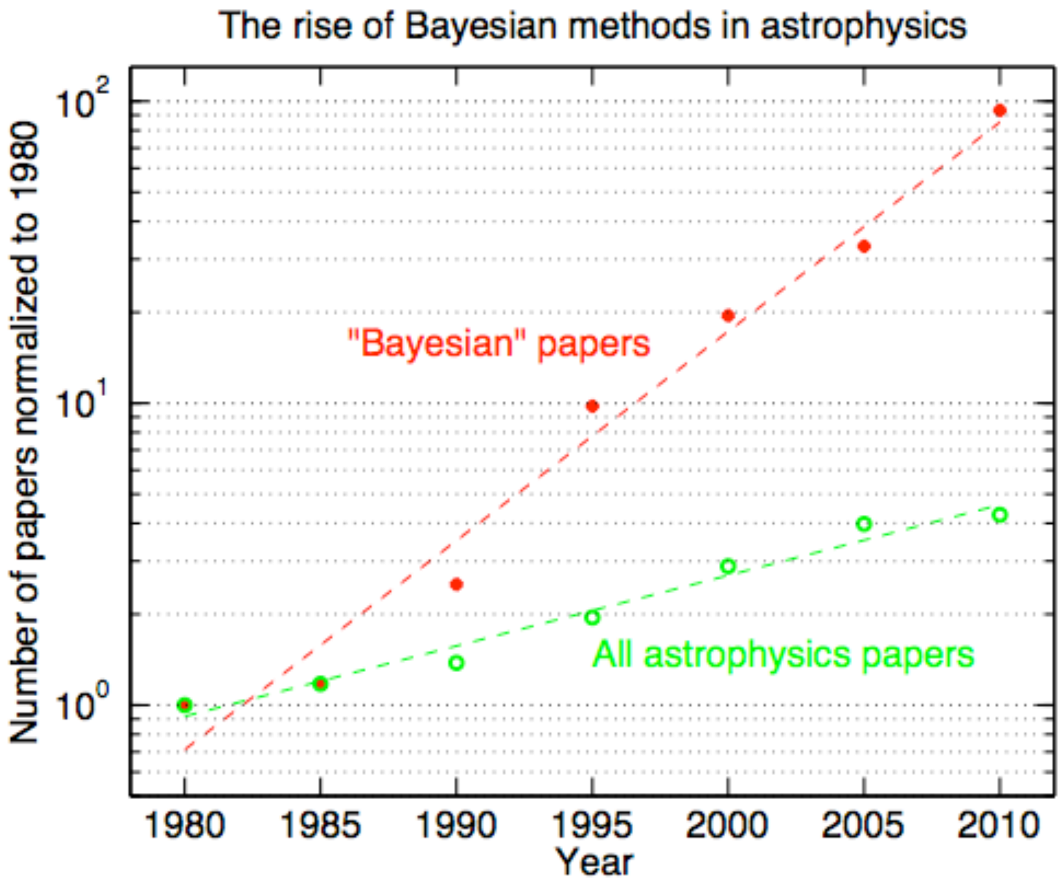
Why cosmology is different

Imperial College London

- An observational science (seems obvious, but it has profound consequences)
- Strong selection effects
- Often poorly understood nuisance parameters
- Cosmic variance limited in some cases
- Not clear what the ensemble would be in a frequentist sense!
- Often, somebody's noise is somebody else's signal: This means that often we are interested in P(signal, noise | data), so there is no obvious way to classify parameters as "nuisance parameters".
- We have the sexier pictures!
- As we never properly learnt statistics, we are mostly Bayesians.

Bayes in the sky

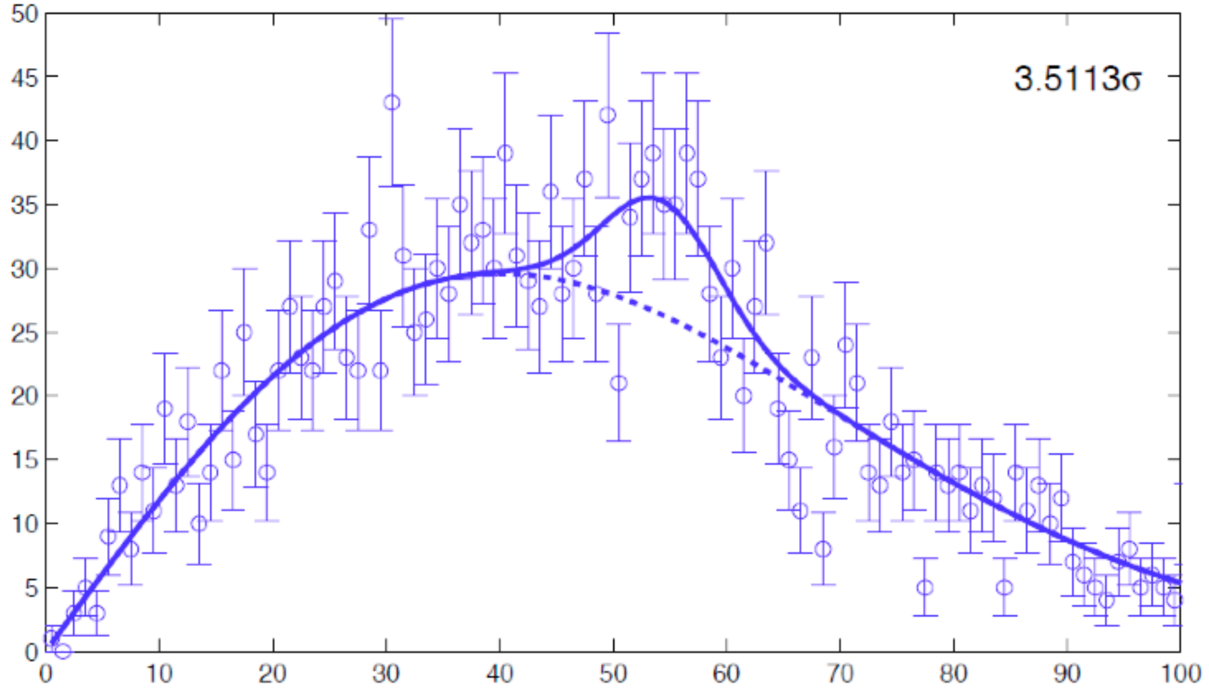
Imperial College London



Look elsewhere effect

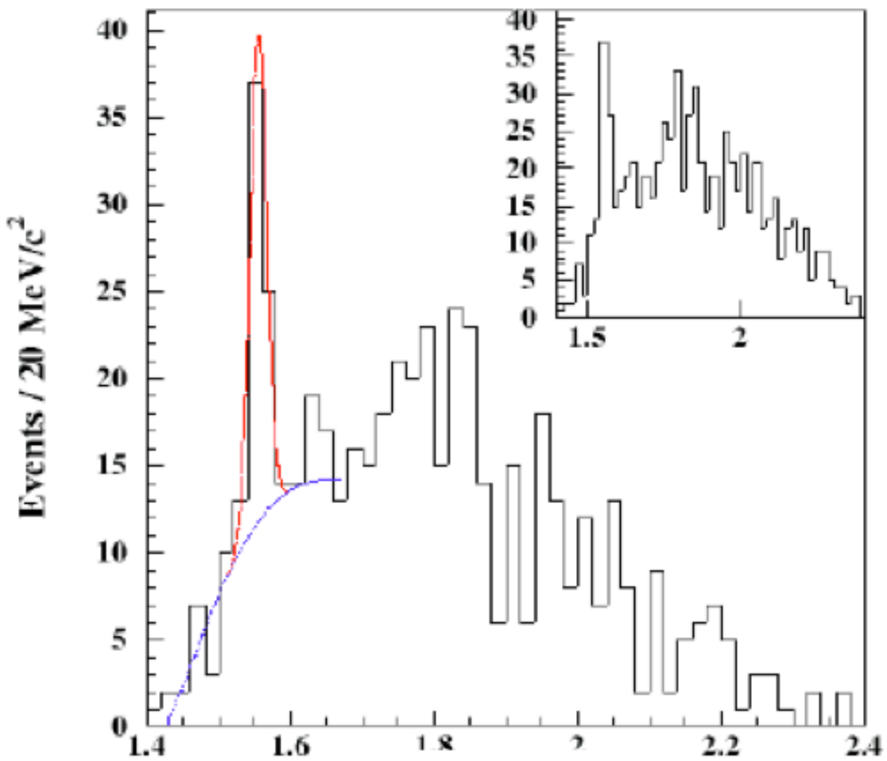
Example from Eilam Gross talk: a 3.5 sigma signal?

Eilam Gross' talk

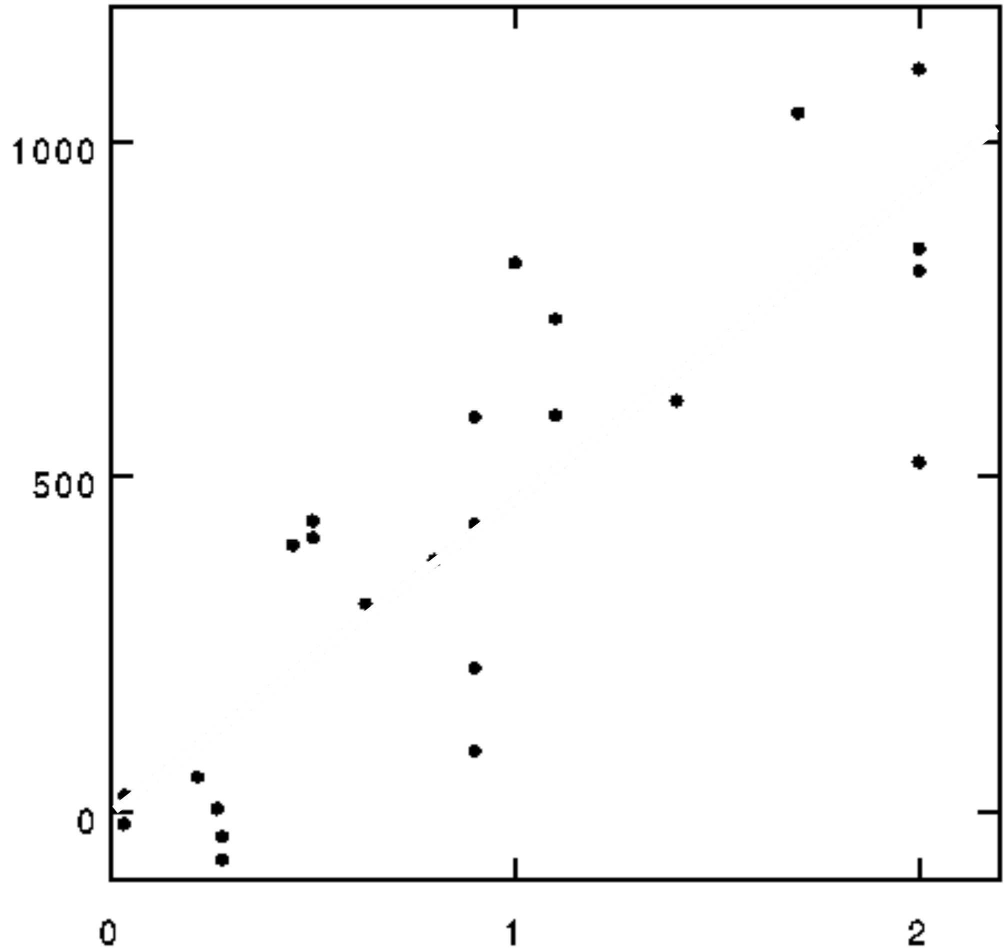


Pentaquark "discovery", ooops, maybe not...

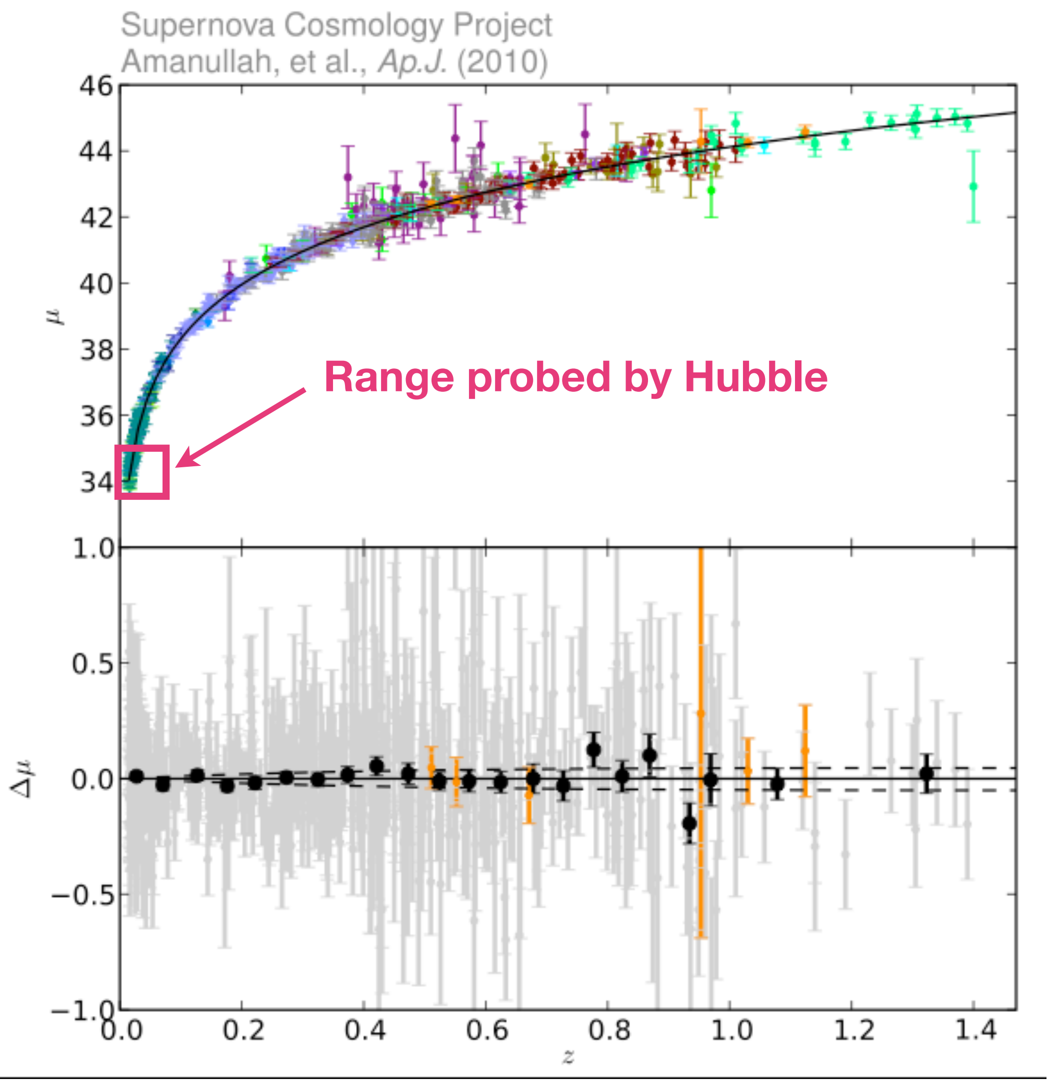
Bob Cousins,
Louis Lyons



$v = H_0 * R$ (Hubble, 1929)



Hubble diagram today



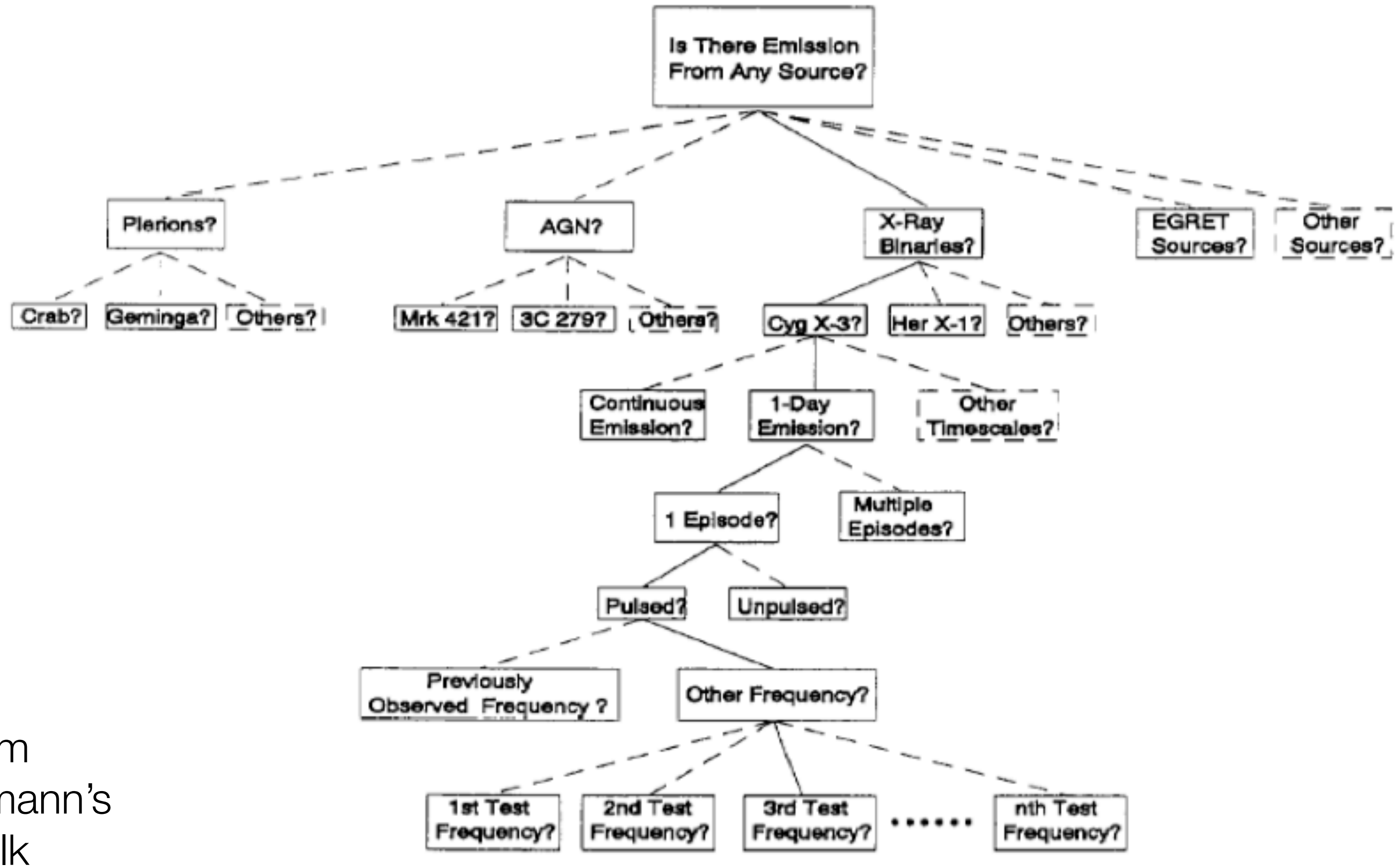
Look elsewhere effect

Imperial College London

- Where is elsewhere?
- Need to define what “elsewhere” means!
- Do future/possible searches matter? It seems to me that only things you have actually looked at should matter.
- This however brings in the Stopping rule problem. You have to make sure you follow your protocol through!

Importance ordering: write out a protocol

S.D. Biller / Astroparticle Physics 4 (1996) 285–291



Jim Linnemann's talk

Look elsewhere effect

- The whole discussion to me gives strong motivation for being a Bayesian about hypothesis testing/model selection. There seems to be no unique/well defined way of defining what “elsewhere” means.
- Jim Berger argued very strongly that the Bayesian answer corresponds to a specific (unique) choice of conditioning statistics for the frequentist testing.
- Things are going to get worse with more complex data sets for which a “single shot” discovery protocol is simply unfeasible.
- Cosmo/astro is typically much more exploratory, with pretty much everybody getting a free hand with the data (see later).

Dangers of looking elsewhere...

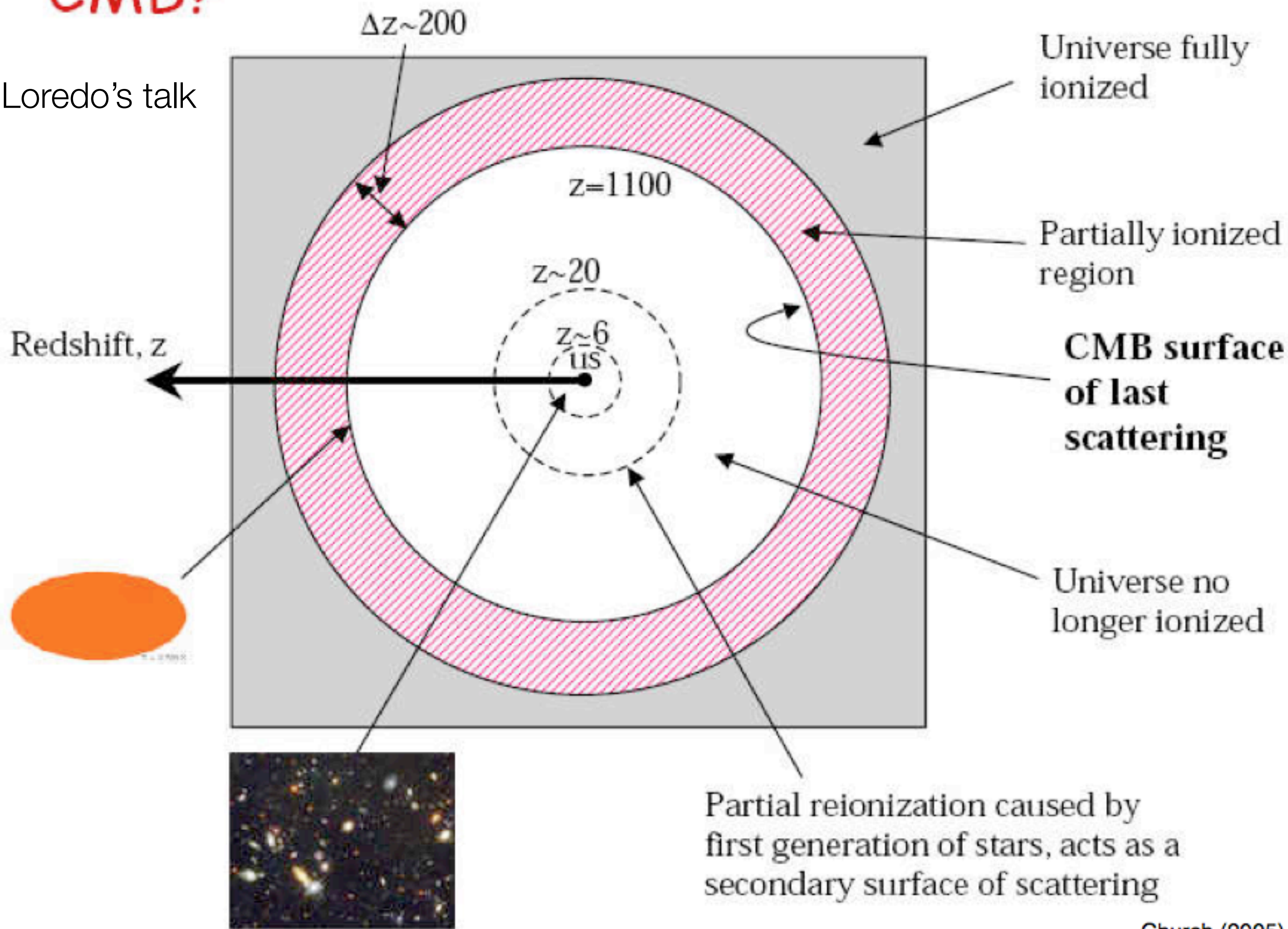
Picture credit: Elilam Gross

Look elsewhere effect in cosmology:

Large-scale anomalies in the cosmic microwave background

What are we seeing when we look at the CMB?

Tom Loredo's talk



Church (2005)

Multi-frequency observations

Imperial College London

WMAP: 5 bands from 23 GHz to 90 GHz

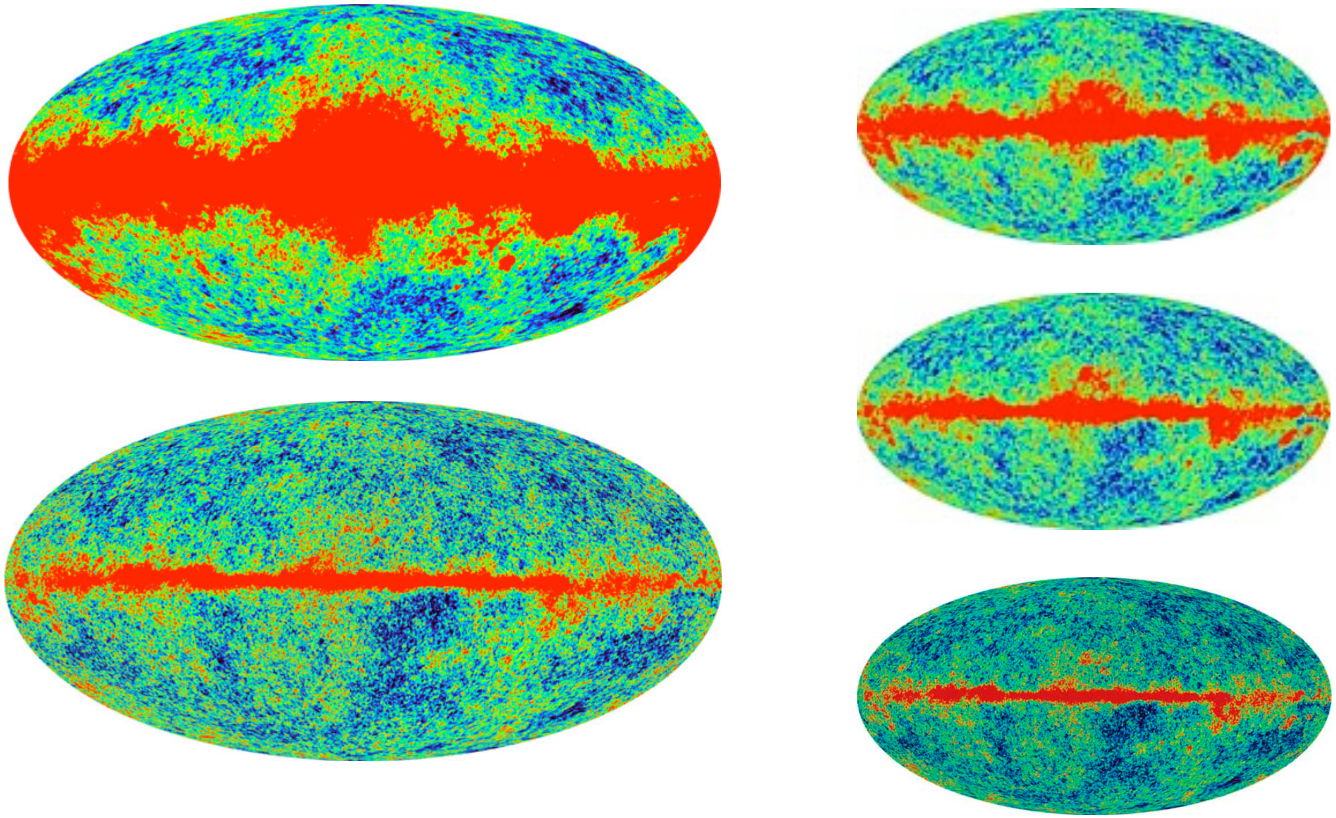
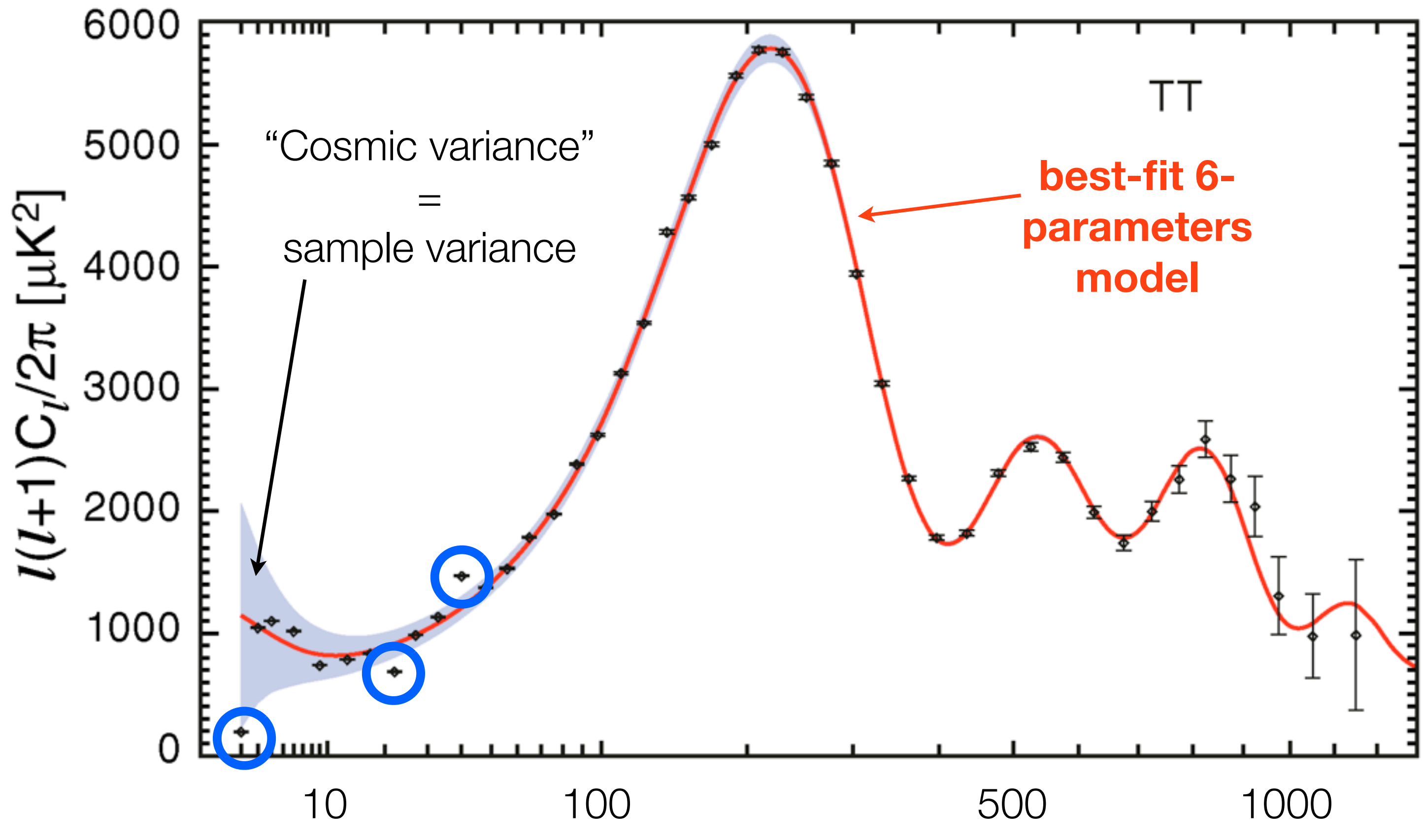


Image: WMAP team

WMAP 7-years temperature power spectrum



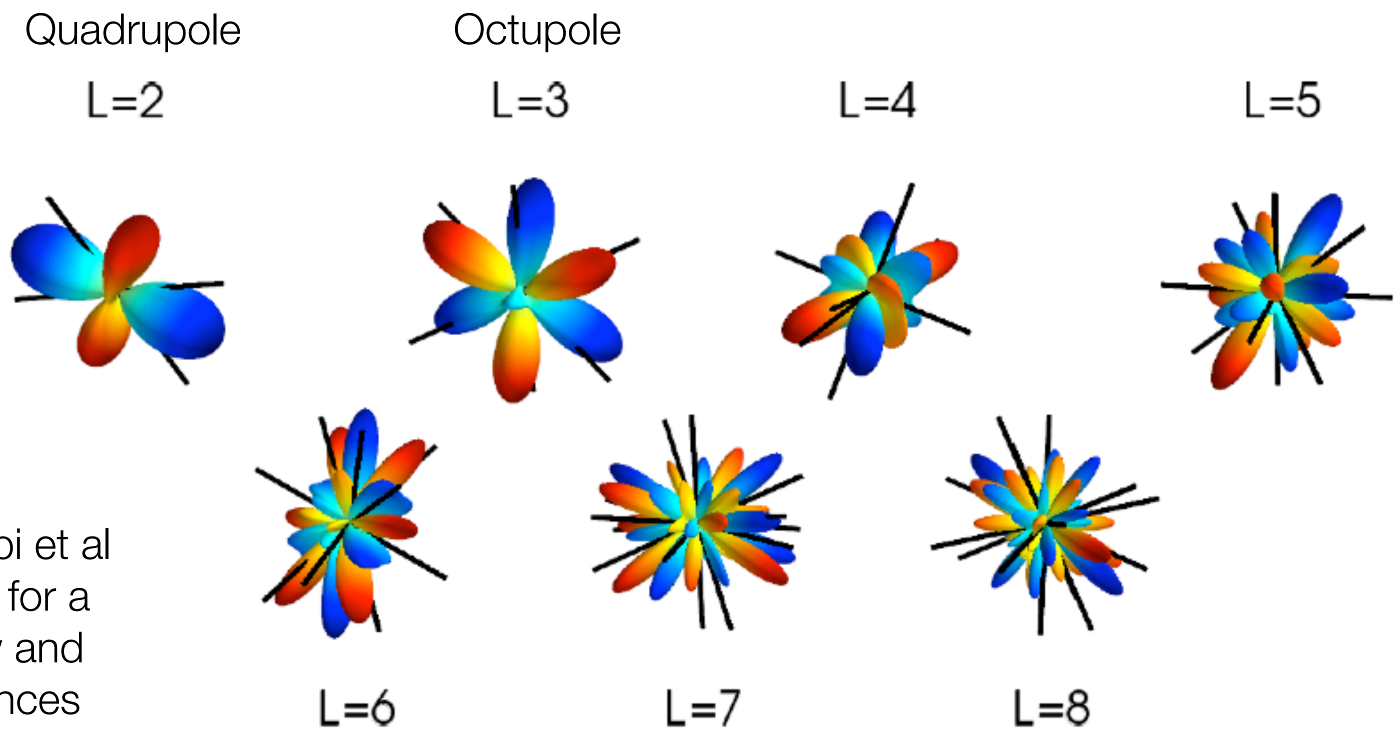
Multipole moment ell

Jarosik et al (2010)

Multipole vectors

$$\frac{\Delta T(\mathbf{n})}{\bar{T}} \equiv \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m} \equiv A^{(\ell)} \left[\prod_{i=1}^{\ell} (\hat{v}^{(\ell,i)} \cdot \hat{e}) - \mathcal{T}_\ell \right]$$

Characterize spherical harmonics by “multiple vectors”



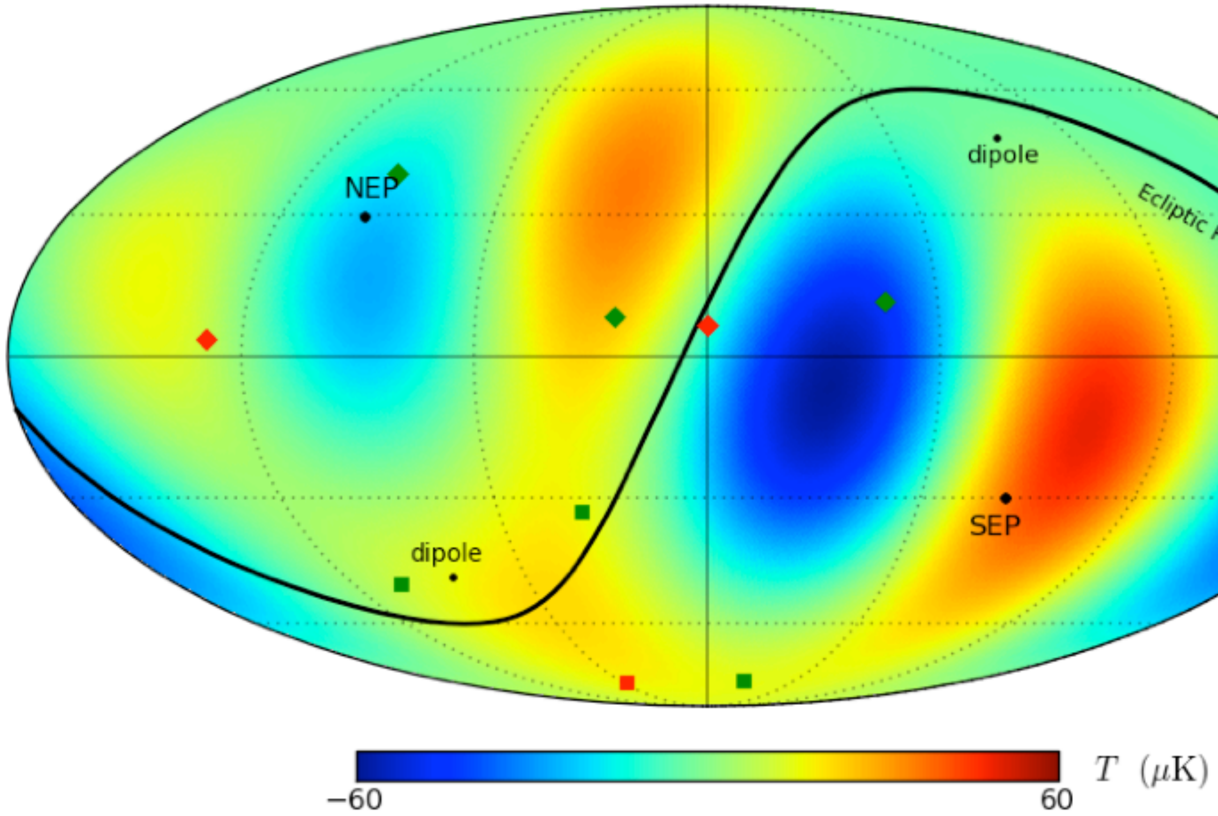
see Copi et al (2010) for a review and references

Large scale anomalies?

Imperial College London

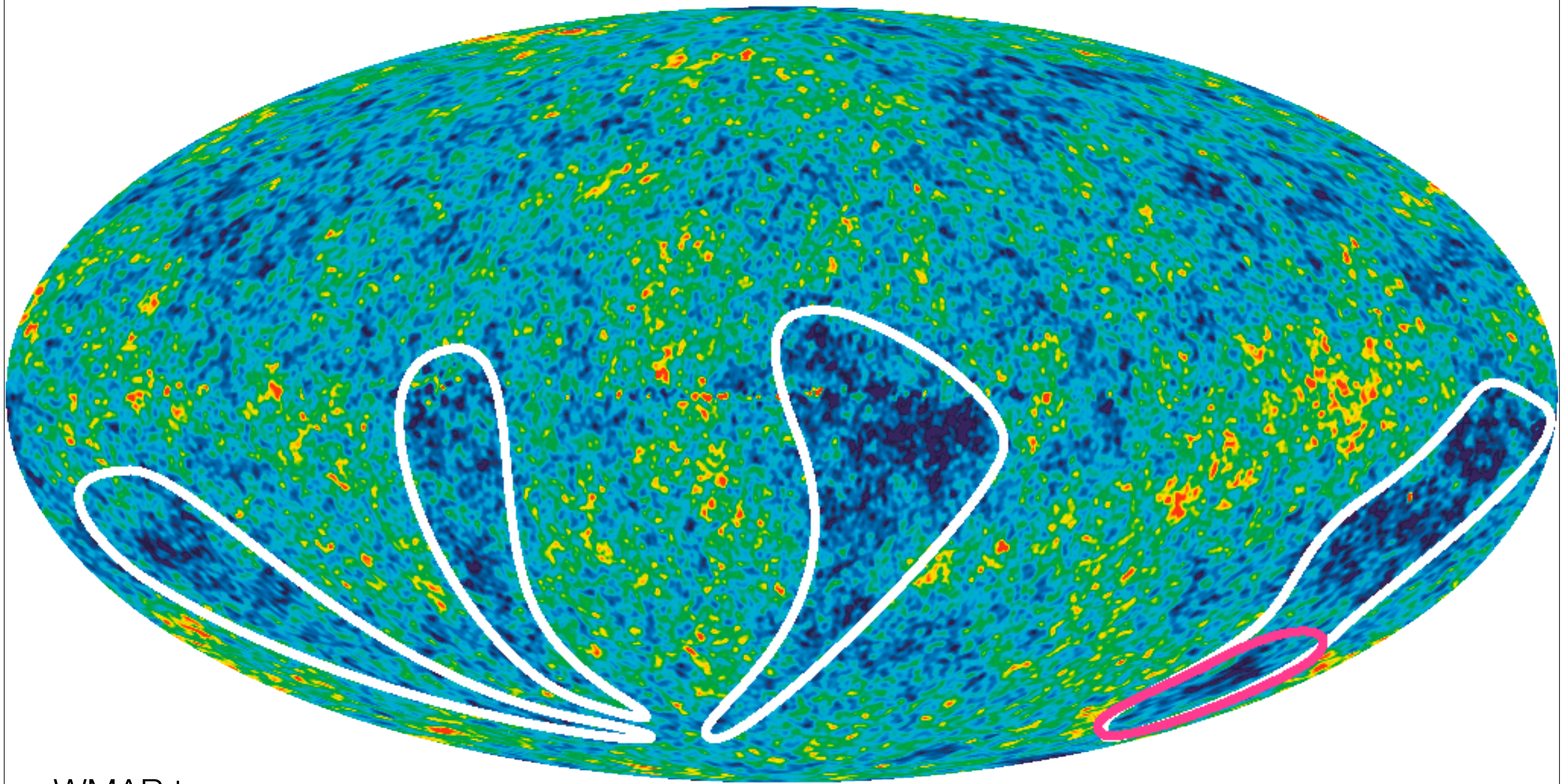
- Quadrupole (ell = 2) is low
- Four area vectors of Quadrupole and Octupole are mutually close (p-value = 0.004)
- Quadrupole and Octupole aligned with ecliptic (p-value = 0.041)
- Normals to area vector planes aligned with dipole (p-value = 0.003)
- Hot/Cold spots divided by the ecliptic (p-value ~ 0.05)
- Two-point correlation function vanishes above 60° (p-value ~ 0.0002)
- All of the above appear to disprove the cosmological principle (isotropy and homogeneity)

Large scale CMB anomalies



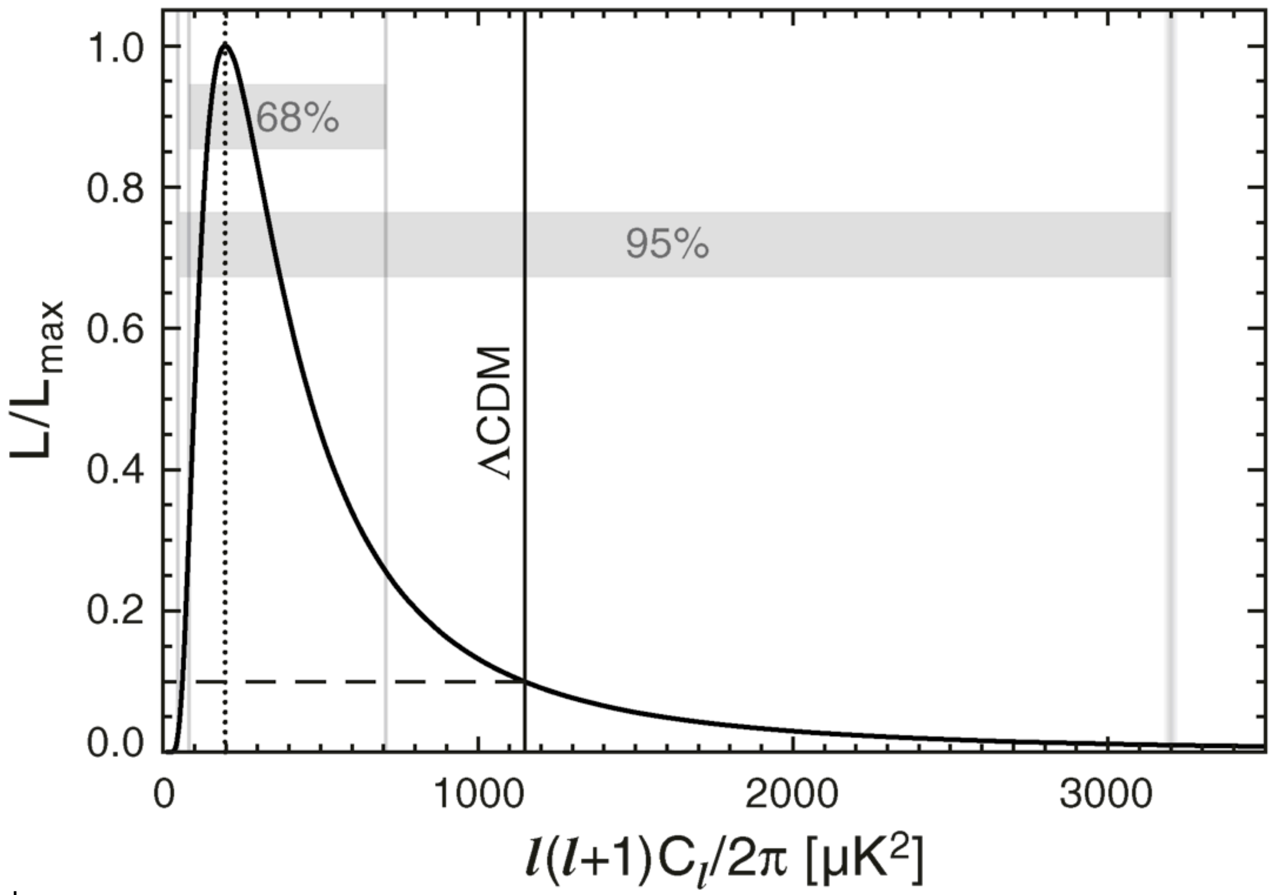
Copi et al (2010)

Anomalous cold regions in the southern hemisphere



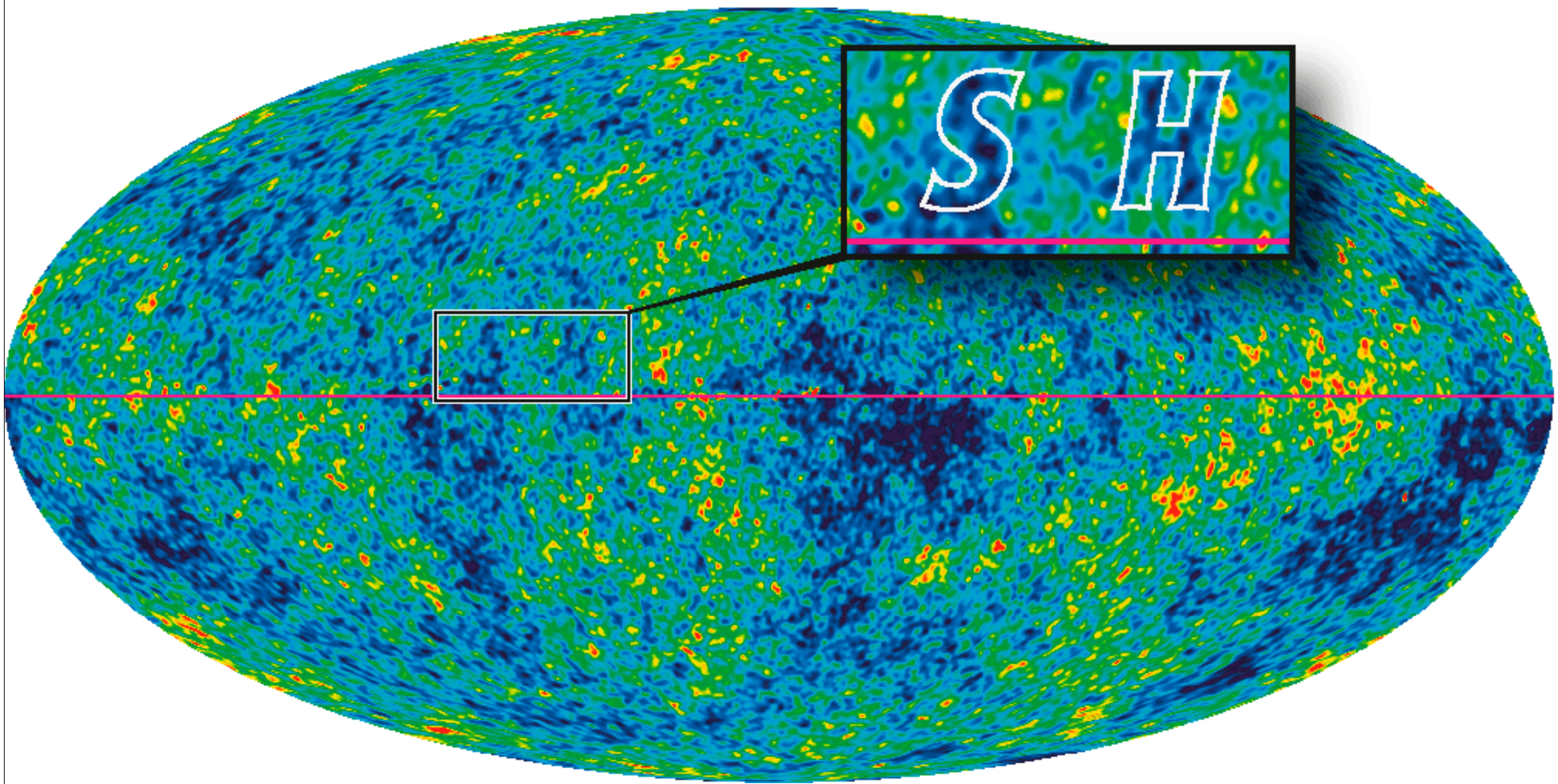
WMAP team

Is the quadrupole anomalously low?



WMAP team

The “SH” initials of Stephen Hawking are shown in the ILC sky map.



WMAP team

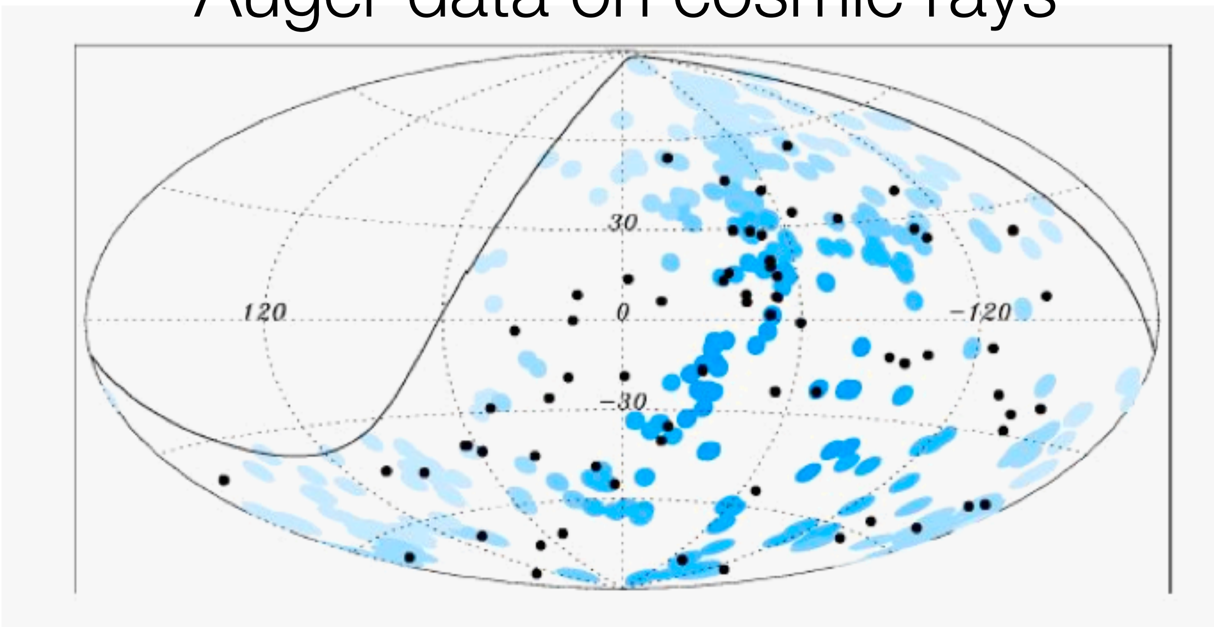
Model selection

Model selection/hypothesis testing

Imperial College
London

- For well-defined problems with “easily” identifiable prior information, Bayesian model selection is the tool of choice.

Auger data on cosmic rays



Paul Sommer’s talk

Tom Loredo's talk:

Odds favouring association of two sources:

Angular error	Odds O $\theta_{12} = 26^\circ$	$\theta_{12} = 0^\circ$
$\sigma_1 = \sigma_2 = 10^\circ$	≈ 1.5	≈ 75
$\sigma_1 = \sigma_2 = 25^\circ$	≈ 7	≈ 12

Advantages:

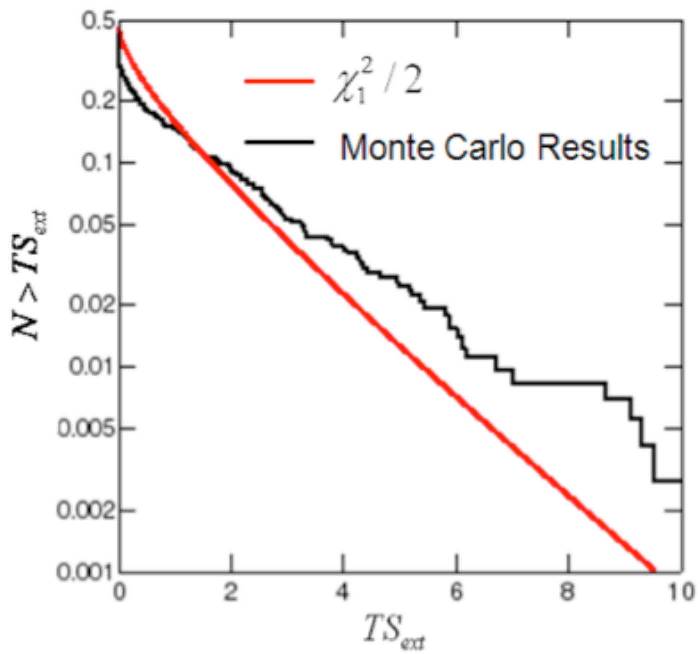
Tuning replaced by averaging

No cuts - the full information is exploited

No fuss with a posteriori statistics

Object detection

- Detection of extended sources in Fermi data falls foul of Wilk's theorem even for toy cases (Elliott Bloom's talk)
- In real life, all other parameters in the fit must also be learnt from the data (background, spectral shape, source location, source extension, amplitude)



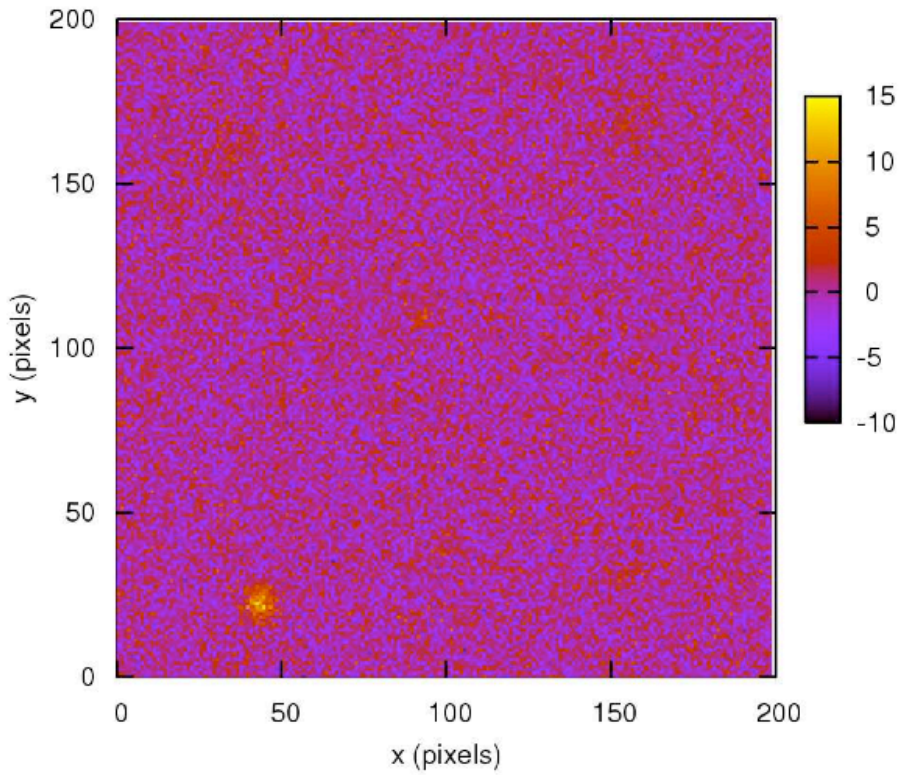
Elliott Bloom's talk

A “simple” example: how many sources?

Imperial College London

Feroz and Hobson (2007)

Signal + Noise

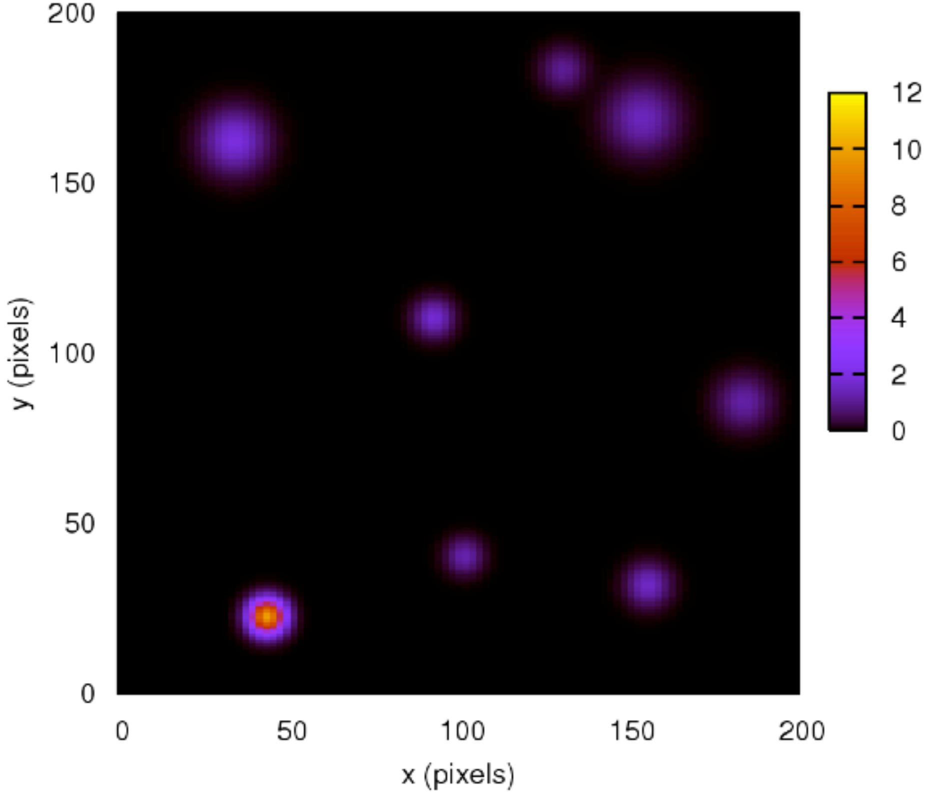


A “simple” example: how many sources?

Imperial College London

Feroz and Hobson (2007)

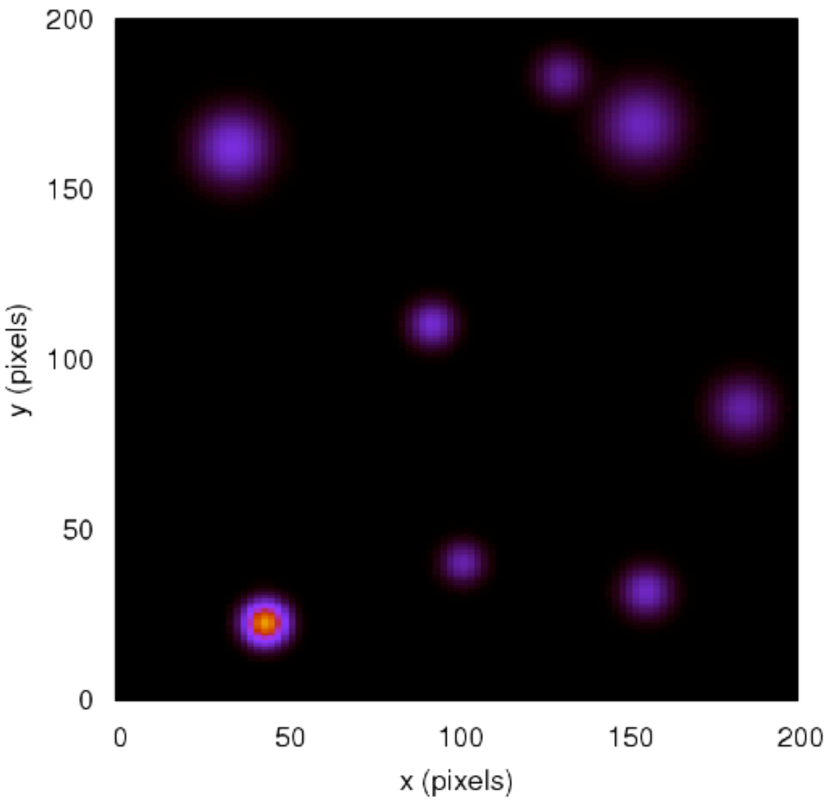
Signal: 8 sources



A “simple” example: how many sources?

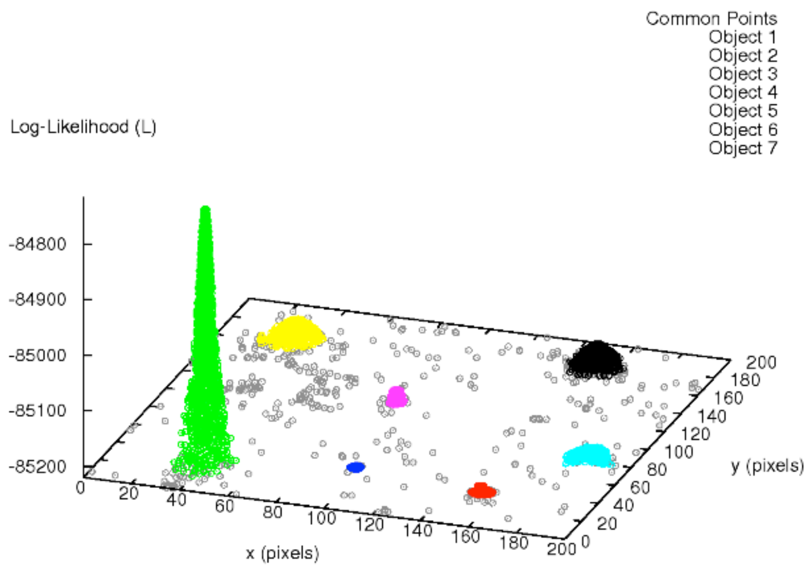
Imperial College London

Feroz and Hobson (2007)

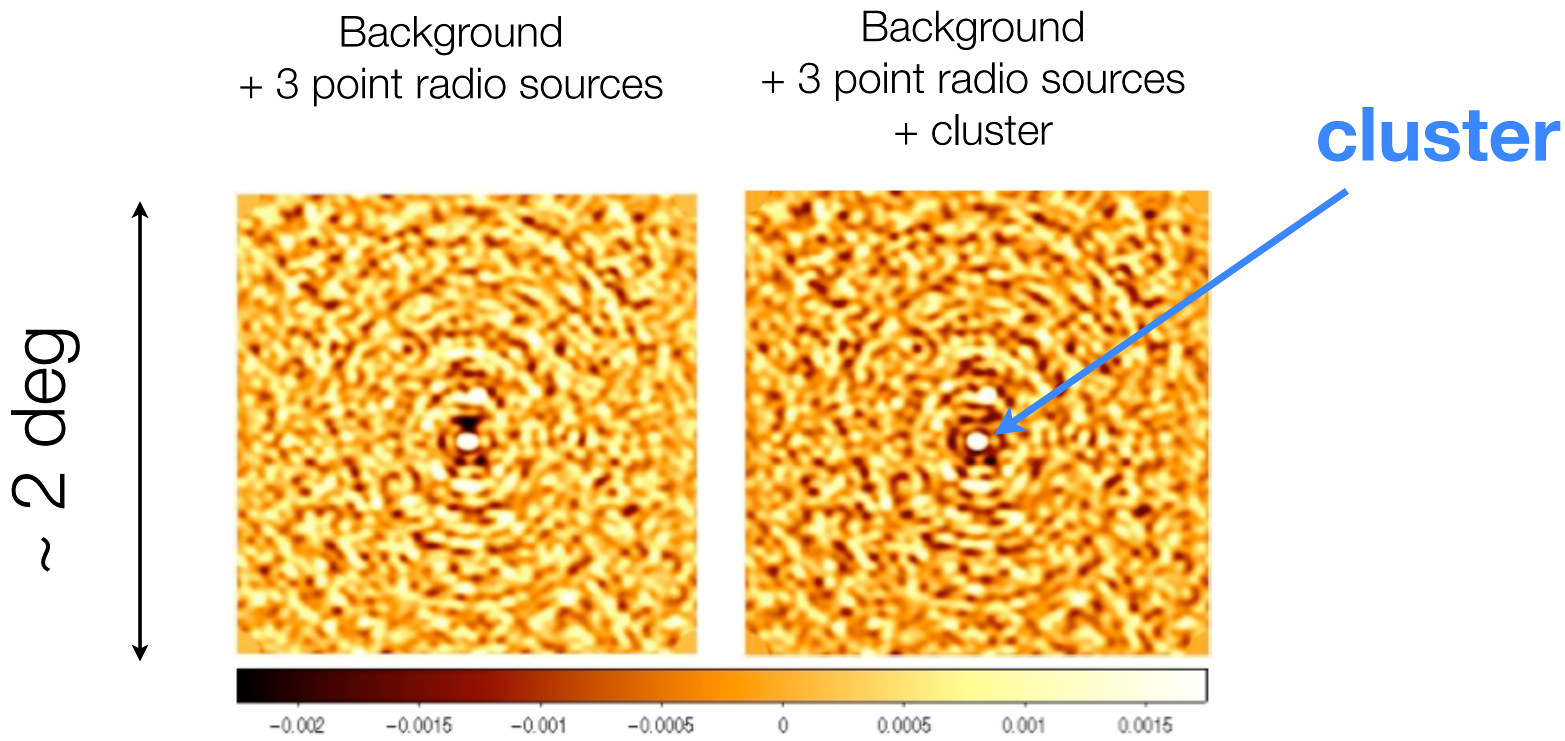


Bayesian reconstruction

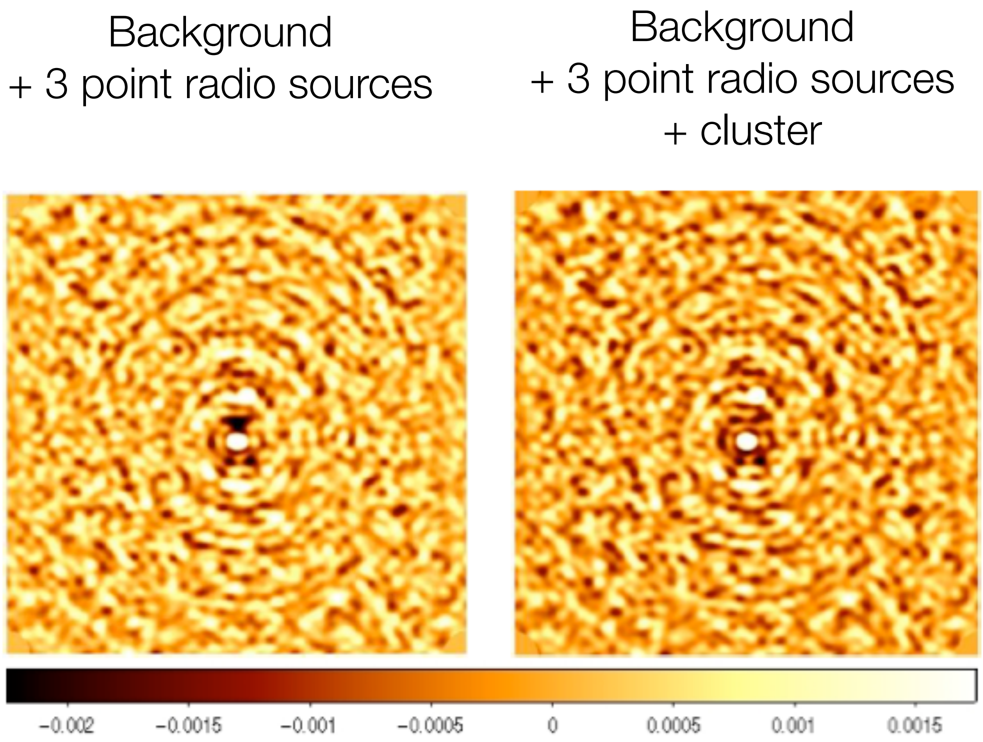
7 out of 8 objects correctly identified.
Mistake happens because 2 objects very close.



Cluster detection from Sunyaev-Zeldovich effect in cosmic microwave background maps



Feroz et al 2009



Bayesian model comparison:

R = P(cluster | data)/P(no cluster | data)

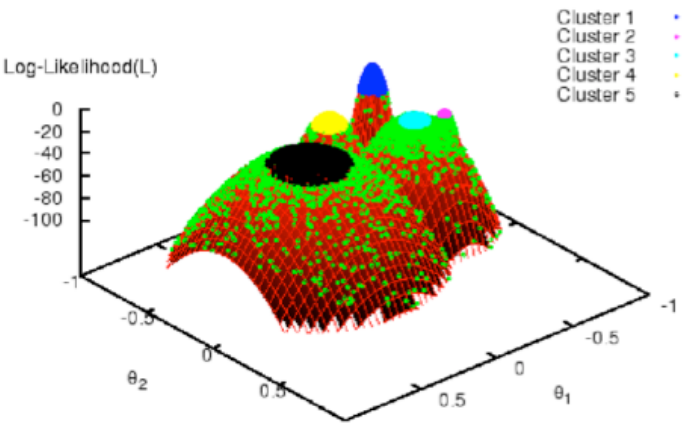
$R = 0.35 \pm 0.05$ $R \sim 10^{33}$

Cluster parameters also recovered (position, temperature, profile, etc)

Computation of the evidence with Multinest

Imperial College London

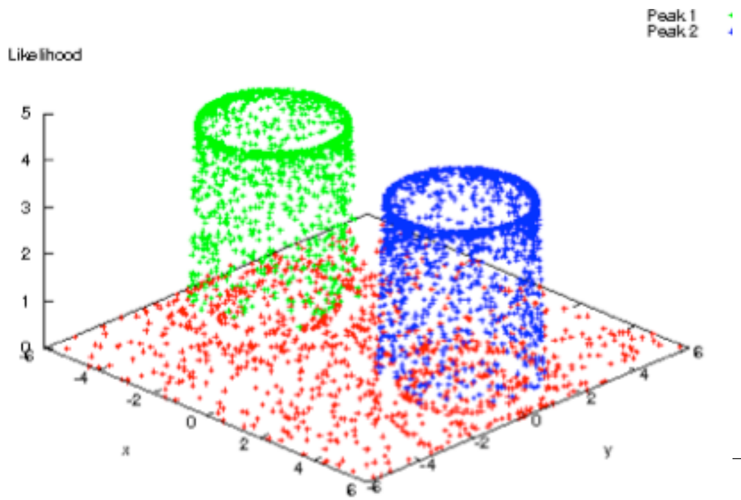
Feroz and Hobson (2007)



Gaussian mixture model:

True evidence: log(E) = -5.27
Multinest:
Reconstruction: log(E) = -5.33 ± 0.11
Likelihood evaluations ~ 10^4
Thermodynamic integration:
Reconstruction: log(E) = -5.24 ± 0.12
Likelihood evaluations ~ 10^6

Courtesy Mike Hobson

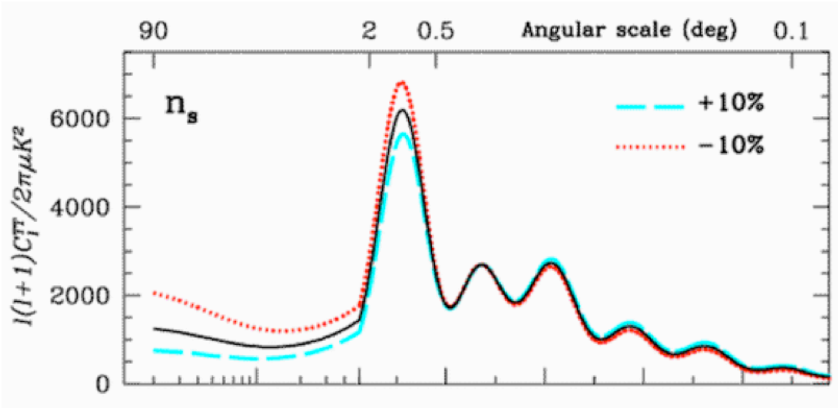


D	N_{like}	efficiency	likes per dimension
2	7000	70%	83
5	18000	51%	7
10	53000	34%	3
20	255000	15%	1.8
30	753000	8%	1.6

Cosmological model selection

Imperial College London

- Is the spectrum of primordial fluctuations scale-invariant (n = 1)?
- Model comparison:
  n = 1 vs n ≠ 1 (with inflation-motivated prior)
- Results:
  n ≠ 1 favoured with odds of 17:1
  (Trotta 2007)
  n ≠ 1 favoured with odds of 15:1
  (Kunz, Trotta & Parkinson 2007)
  n ≠ 1 favoured with odds of 7:1
  (Parkinson 2007 et al 2006)



Where Bayesian model selection can go wrong

Imperial College London

- In cosmology/HEP we have many situations with nested models with extra unknown parameters for the fundamental theory.
- Little or nothing is known about the metric to be imposed on such a parameter space
- “The concept of total ignorance about θ does not have any precise meaning” (Bob Cousins)
- “θ is θ!” (Bob Cousins)

Where Bayesian model selection can go wrong

Imperial College London

- Occam's razor factor may be arbitrary. HOWEVER: if the range of your prior is arbitrary (by many orders of magnitude) then arguably the physics behind it is not strongly predictive...
- In some cases, the upper bound formalism might be useful (Jim Berger and collaborators)
- In the cosmology community, people often use (blindly) Information Criteria (often with silly answers).

Information criteria

Imperial College London

- Several information criteria exist for approximate model comparison
  k = number of fitted parameters, N = number of data points,
  -2 $\ln(L_{max})$ = best-fit chi-squared

- **Akaike Information Criterium (AIC):**

$$\mathrm{AIC} \equiv -2 \ln \mathcal{L}_{\mathrm{max}} + 2k$$

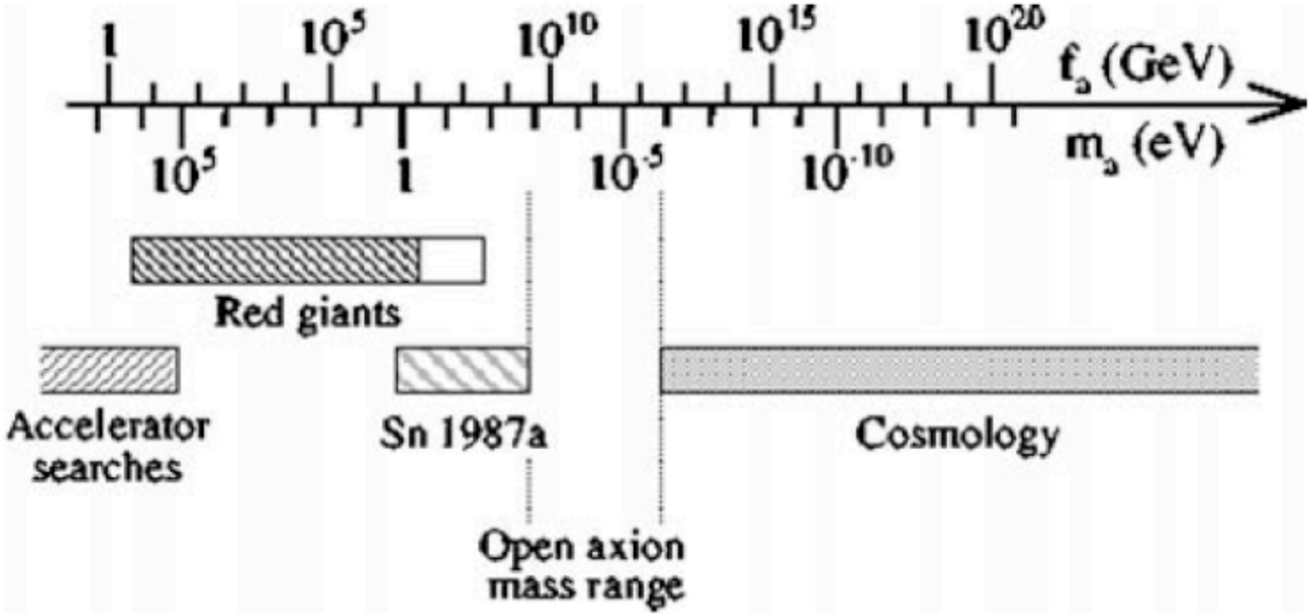
- **Bayesian Information Criterium (BIC):**

$$\mathrm{BIC} \equiv -2 \ln \mathcal{L}_{\mathrm{max}} + k \ln N$$

- **Deviance Information Criterium (DIC):**

$$\mathrm{DIC} \equiv -2 \widehat{D_{\mathrm{KL}}} + 2\mathcal{C}_b.$$

Axion discovery space



For most exploratory experiments I can think of, these metrics just don't exist in a relevant way.

(Bob Cousin's talk)

Nested models

M_0: $\boldsymbol{\theta} = 0$

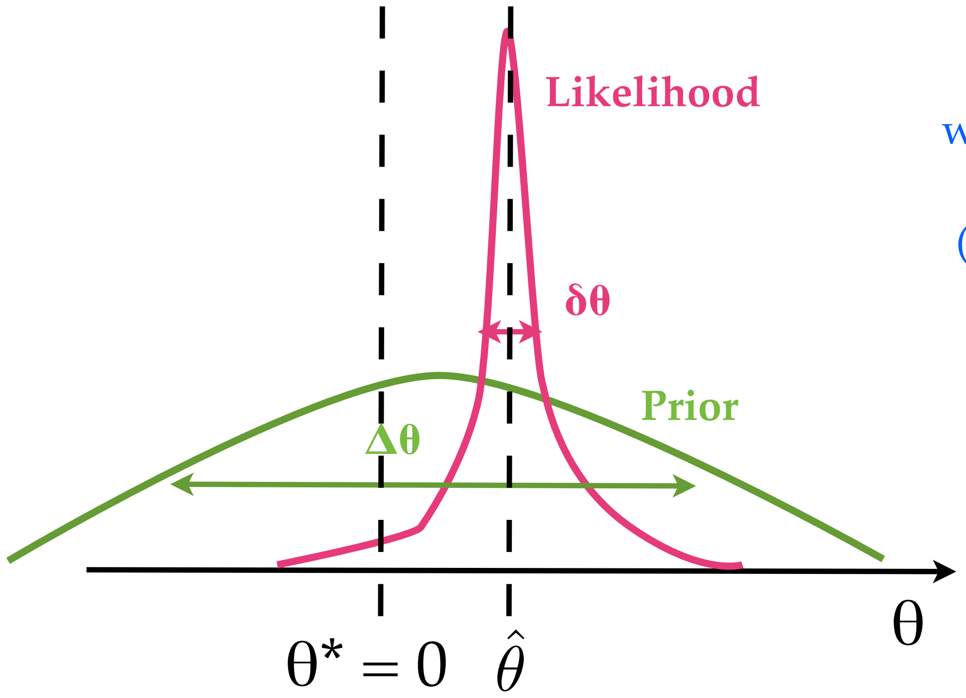
M_1: $\boldsymbol{\theta} \neq 0$ with prior p($\boldsymbol{\theta}$)

$$\lambda \equiv \frac{\hat{\theta} - \theta^\star}{\delta\theta}$$

$$\ln B_{01} \approx \ln \frac{\Delta\theta}{\delta\theta} - \frac{\lambda^2}{2}$$

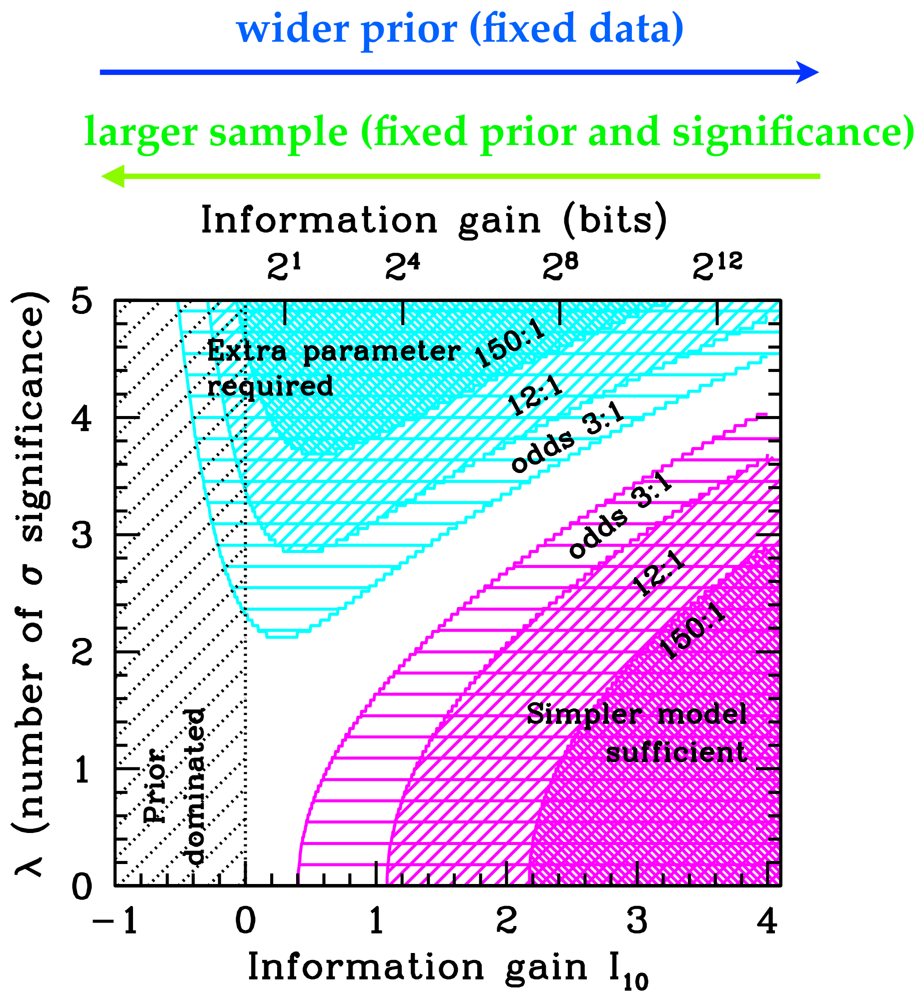
wasted parameter space (favours simpler model)

mismatch of prediction with observed data (favours more complex model)



Model selection for nested models

Imperial College London



Another look at Lindley's paradox (Bob Cousins' talk)

Jim Berger argued that one should look at the scale of the prior and hope that the result is robust for reasonable choices

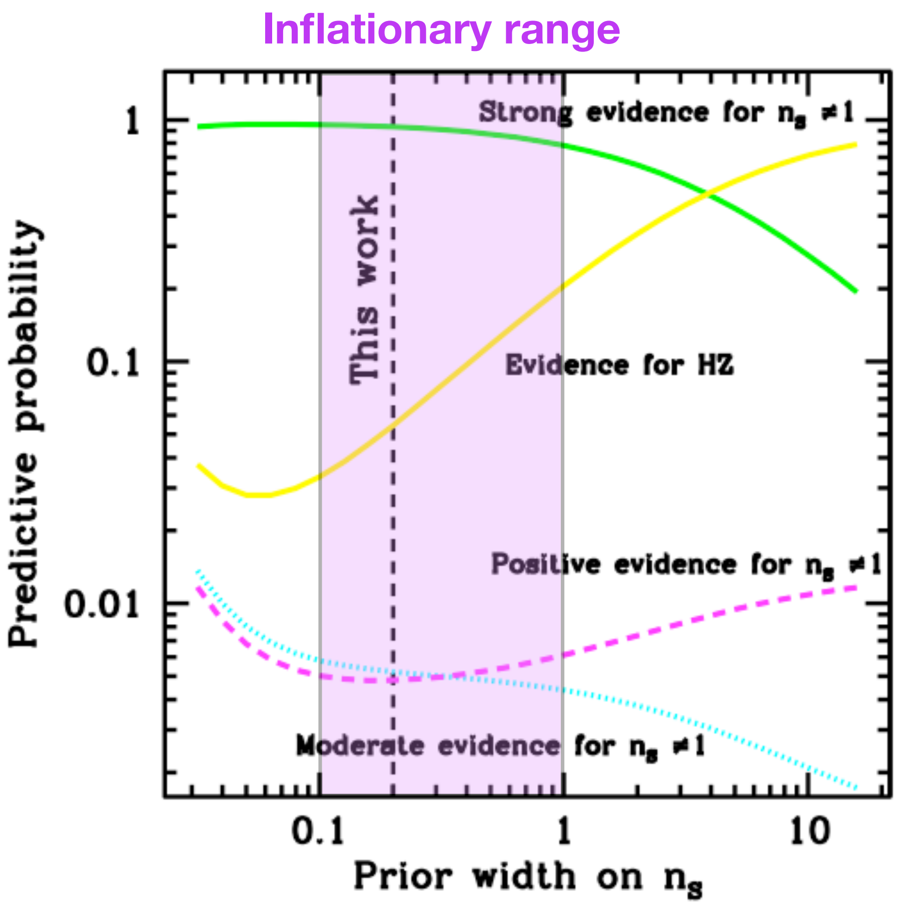
$$I_{10} \equiv \log_{10} \frac{\Delta\theta}{\delta\theta}$$

Trotta (2008)

Example of reasonable sensitivity analysis

Imperial College London

- The answer does not change for physically reasonable changes in the prior width

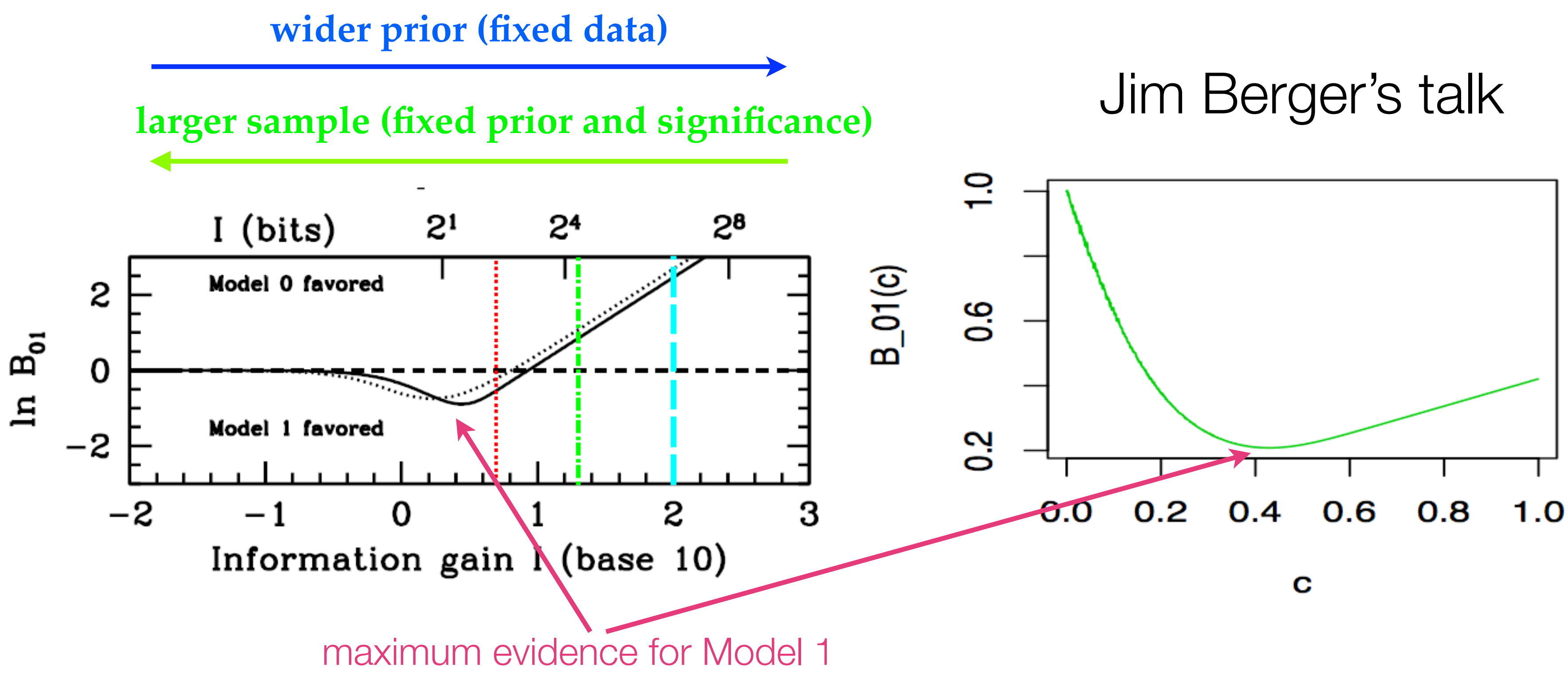


Trotta (2007)

“Prior-free” evidence bounds

Imperial College London

- What if we do not know how to set the prior? For nested models, we can still choose a prior that will maximise the support for the more complex model:



Maximum evidence for a detection

Imperial College London

- **The absolute upper bound:** put all prior mass for the alternative onto the observed maximum likelihood value. Then

$$B < \exp(-\chi^2/2)$$

- **More reasonable class of priors:** symmetric and unimodal around Ψ=0, then (α = significance level)

$$B < \frac{-1}{\exp(1)\alpha \ln \alpha}$$

If the upper bound is small, no other choice of prior will make the extra parameter significant.

Sellke, Bayarri & Berger, *The American Statistician*, 55, 1 (2001)

How to interpret the "number of sigma's"

Imperial College London

p	sigma	Absolute bound on lnB (B)	"Reasonable" bound on lnB (B)
0.05	2.0	2.0 (7:1) weak	0.9 (3:1) undecided
0.003	3.0	4.5 (90:1) moderate	3.0 (21:1) moderate
0.0003	3.6	6.48 (650:1) strong	5.0 (150:1) strong

A conversion table

Imperial College London

p–value	$\bar{B}$	$\ln \bar{B}$	sigma	category
0.05	2.5	0.9	2.0	
0.04	2.9	1.0	2.1	'weak' at best
0.01	8.0	2.1	2.6	
0.006	12	2.5	2.7	'moderate' at best
0.003	21	3.0	3.0	
0.001	53	4.0	3.3	
0.0003	150	5.0	3.6	'strong' at best
6×10^{-7}	43000	11	5.0	

Rule of thumb:

a n-sigma result should be interpreted as a n-1 sigma result

Application: dipole modulation

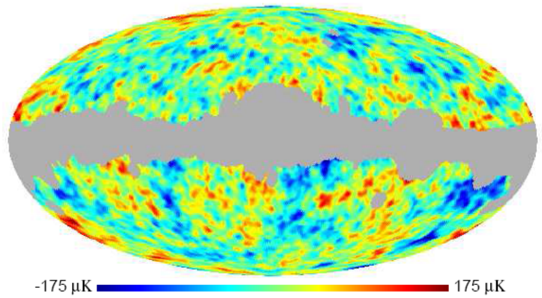
Imperial College London

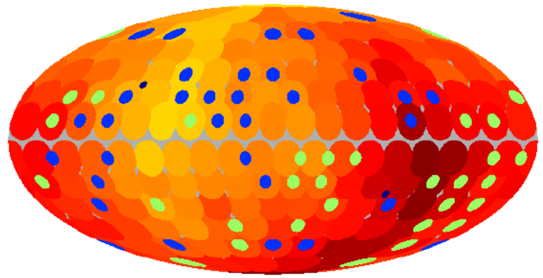
- Eriksen et al (2004) found hints for a dipolar modulation in WMAP1 ILC map

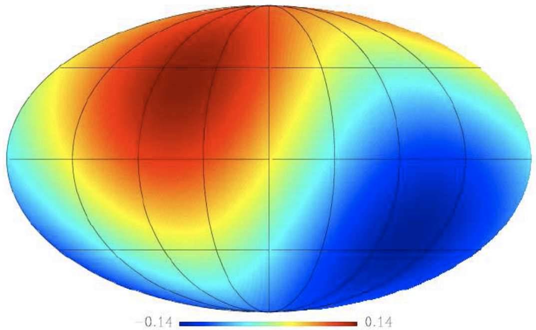
- Adding a phenomenological dipole pattern improves the chi-square by 9 units (for 3 extra parameters)

- Is this significant evidence?

- Not really: upper bound on B is odds of 9:1
  The absolute upper bound is about the same
  (Gordon and Trotta 2007)





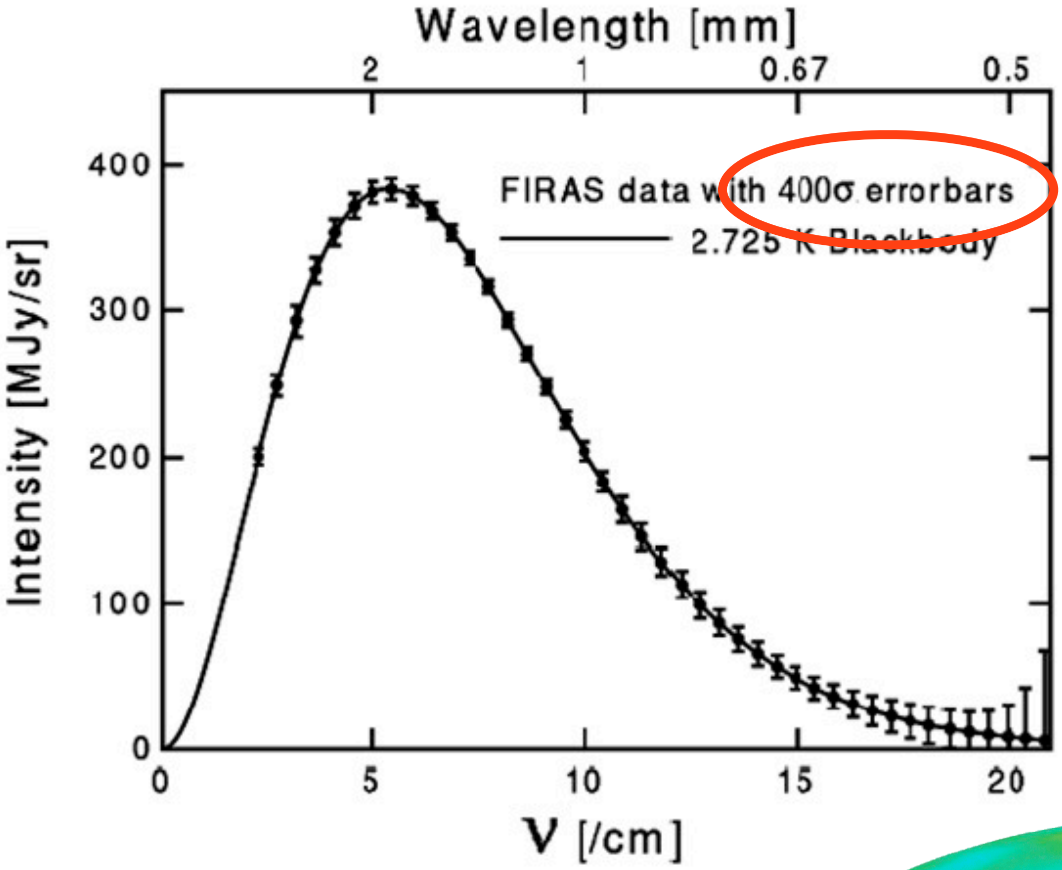


Why 5 sigma?

Relevance of 5 sigma for cosmology

Imperial College London

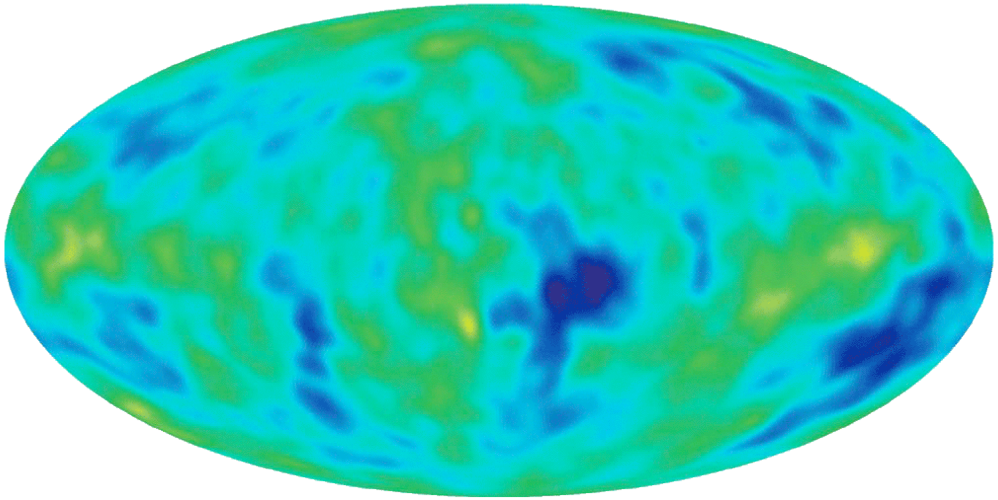
- I cannot think of many examples where this is relevant in cosmology
- **Discovery of the CMB:** Penzias and Wilson (1965) T = 3.5 ± 1.0 K (3.5 sigma)
  NOBEL PRIZE 1978
- **Blackbody nature of the CMB:** this was a slam-dunk discovery
  NOBEL PRIZE 2006 (Mather)
- **COBE measurement of anisotropies in the CMB** (1994)
  Quadrupole measurement = 15.3 +3.8-2.8 μK (~ 5.4 sigma)
  NOBEL PRIZE 2006 (Smoot)



COBE (1994)

$\Delta T/T \sim 10^{-5}$

Angular resolution ~ 10 deg

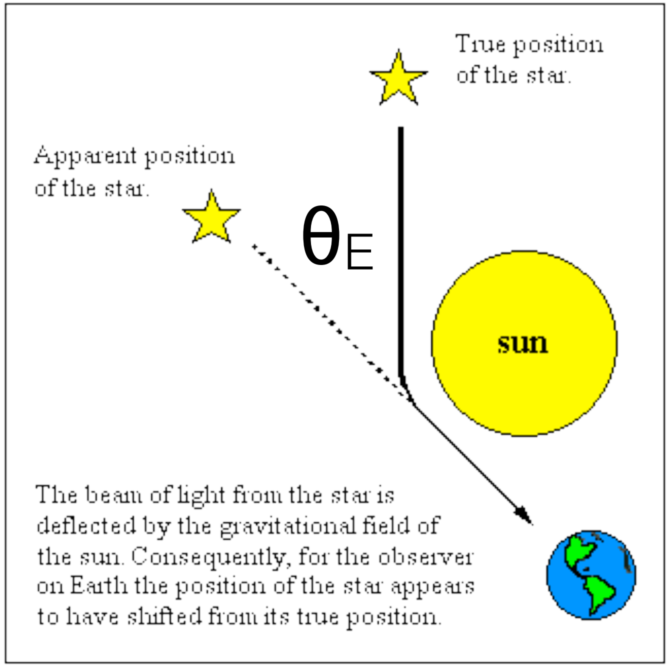


Evidence for Einstein gravity (1919)

Imperial College London

- Einstein's theory of General Relativity made a crucial prediction: the deflection angle around the Sun should be twice what predicted by Newton

$$\theta_E = 2\ \theta_N = 1.75''$$



Evidence for Einstein gravity (1919)

Imperial College London

- Measurements were performed during the solar eclipse of May 29th 1919:

  Eddington: θ = 1.61 ± 0.40 arcsec (based on 5 stars)
  Crommelin: θ = 1.98 ± 0.16 arcsec (based on 7 stars)

	Einstein	**Newton**
Hypothesis	θ =1.75 arcsec	θ = 0.875 arcsec
p-value from Eddington's data	0.72	0.06
Posterior odds for Einstein vs Newton	~ 5 to 1(weak evidence)	

Evidence for varying α?

Imperial College London

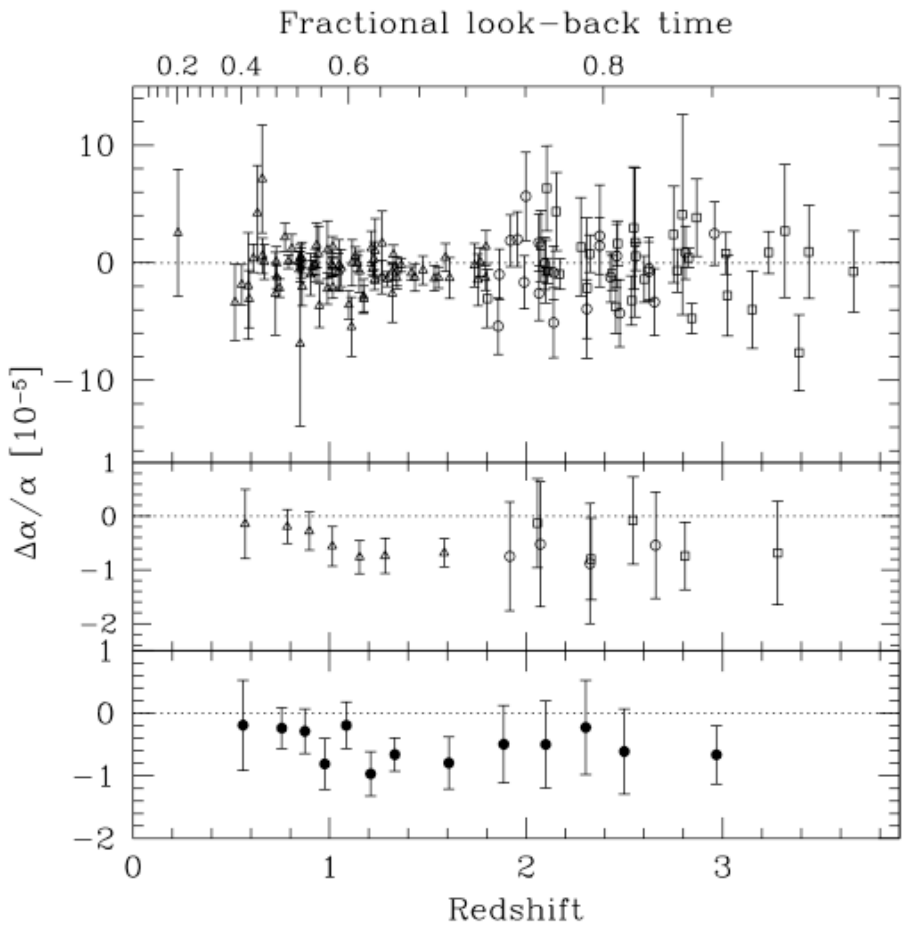
- For several years, Webb and collaborators have claimed a ~ 5 sigma evidence for a time variation in the fine structure constant from analysis of QSO absorption spectra.

random error in 22 high-z systems characterized by transitions with a large dynamic range in apparent optical depth. Increasing the statistical errors on $\Delta\alpha/\alpha$ for these systems gives our fiducial result, a weighted mean $\Delta\alpha/\alpha = (-0.543 \pm 0.116) \times 10^{-5}$, representing $4.7\,\sigma$ evidence for a varying α. Assuming that $\Delta\alpha/\alpha = 0$ at $z_{\rm abs} = 0$, the data marginally prefer a linear increase in α with time rather than a constant offset from the laboratory value: $\dot{\alpha}/\alpha = (6.40 \pm 1.35) \times 10^{-16}\,{\rm yr}^{-1}$. The two-point correlation function for α is consistent with zero over 0.2–13 Gpc comoving scales and the angular distribution of $\Delta\alpha/\alpha$ shows no

Murphy et al (2003)

Evidence for varying α?

Imperial College London



Murphy et al (2003)

My 2 pennies

Imperial College London

- Standards of evidence in physics are not absolute: we have a Bayesian prior in the back of our minds when assessing strength of evidence (systematics, plausibility, scientific experience, appeal of the model, theoretical framework, simplicity, how the model fits within the bigger picture, elegance, etc).
- How those factors could be summarized in P(M) is difficult to imagine.
- Jim Berger argued that priors should be “defendible”, no matter how you got there.
- “Inside every Frequentist there is a Bayesian struggling to get out” (Lindley).
- Bayesian model selection works best in cases where relevant prior information can be objectively specified (e.g., object detection example).

THANK YOU!