Statistical techniques for data analysis in Cosmology

arXiv:0712.3028; arXiv:0911.3105

Numerical recipes (the “bible”)

Licia Verde

ICREA & ICC UB-IEEC

http://icc.ub.edu/~liciaverde

outline

- Lecture 1: Introduction Bayes vs Frequentists, priors, the importance of being Gaussian, modeling and statistical inference, some useful tools. Monte Carlo methods.
- Lecture 2: Different type of errors. Going beyond parameter fitting. forecasting: Fisher matrix approach. Introduction to model selection. Real world effects

  Conclusions.

recap

$$\mathcal{P}(H|D) = \frac{\mathcal{P}(H)\mathcal{P}(D|H)}{\mathcal{P}(D)}$$

Bayes theorem

Likelihood

Prior,posterior

Statistical inference

But we have lost something…

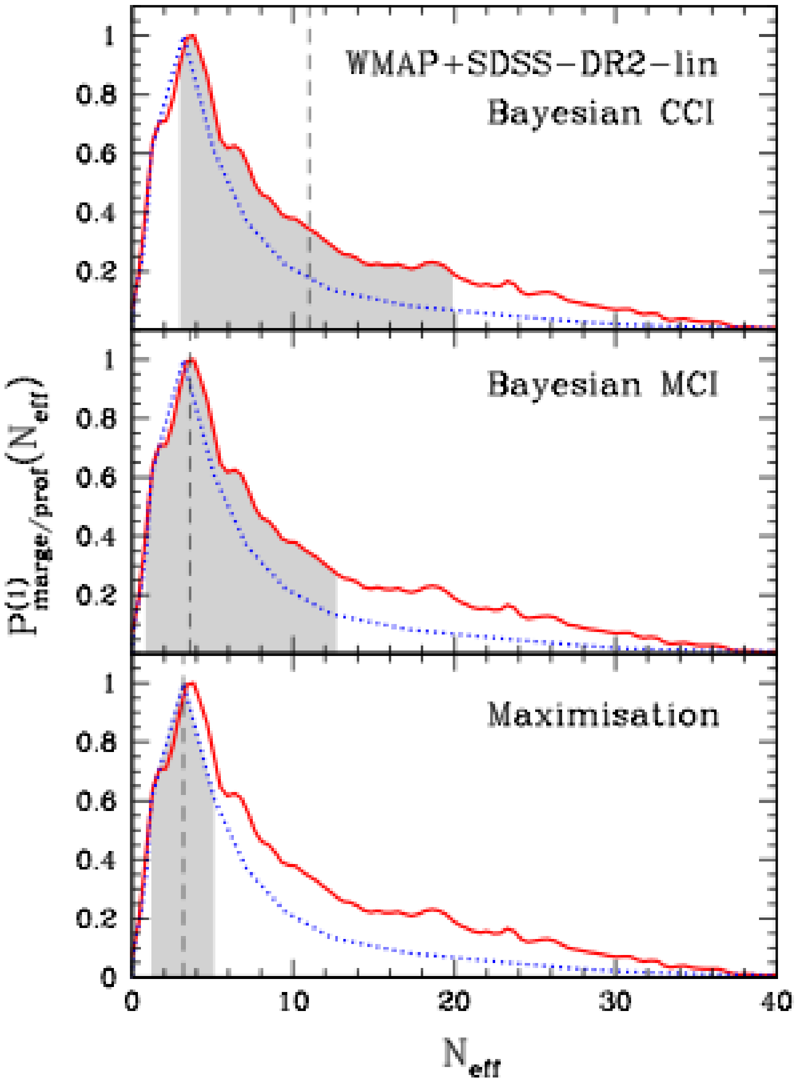
Best fit: max (likelihood/posterior)

(which maximum?)

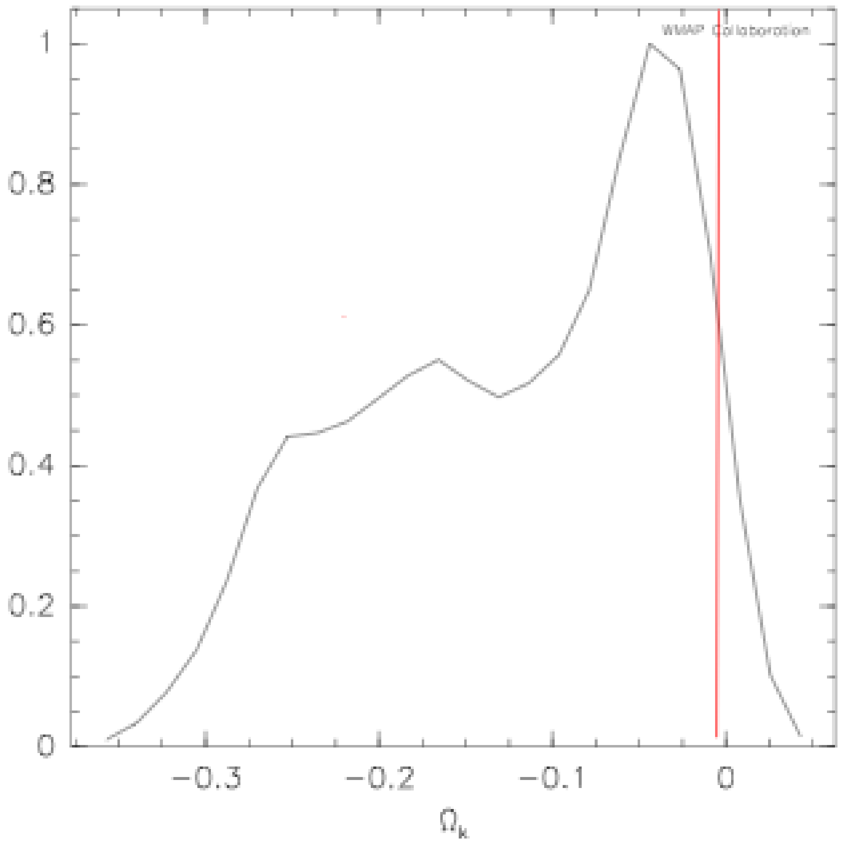
Approximate but fast ways to sample the posterior

Errors or confidence intervals

Errors, what errors?



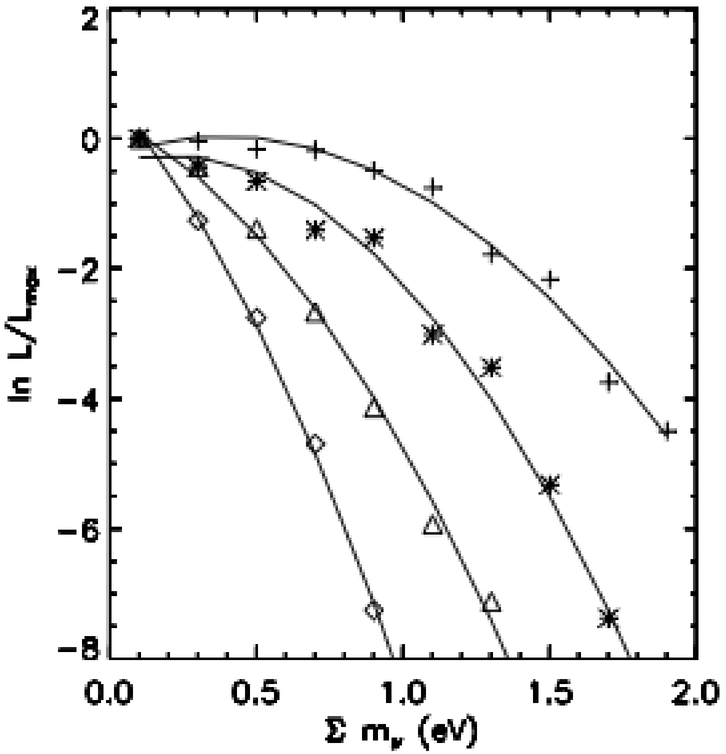
Hamann et al. arXiv:0705.0440


WMAP Collaboration
1
0.8
0.6
0.4
0.2
0
−0.3
−0.2
−0.1
0
Ω_k

Prior-independence?

Once you and an MCMC output what you can do is to look at the likelihood value not the weight.

Say you have n uninteresting parameters and one that you are interested in e.g. $m\nu$. For each value of $m\nu$ find the maximum likelihood L_m regardless of the values assumed by the other parameters. Then consider L_m/L_{max} as a function of $m\nu$.

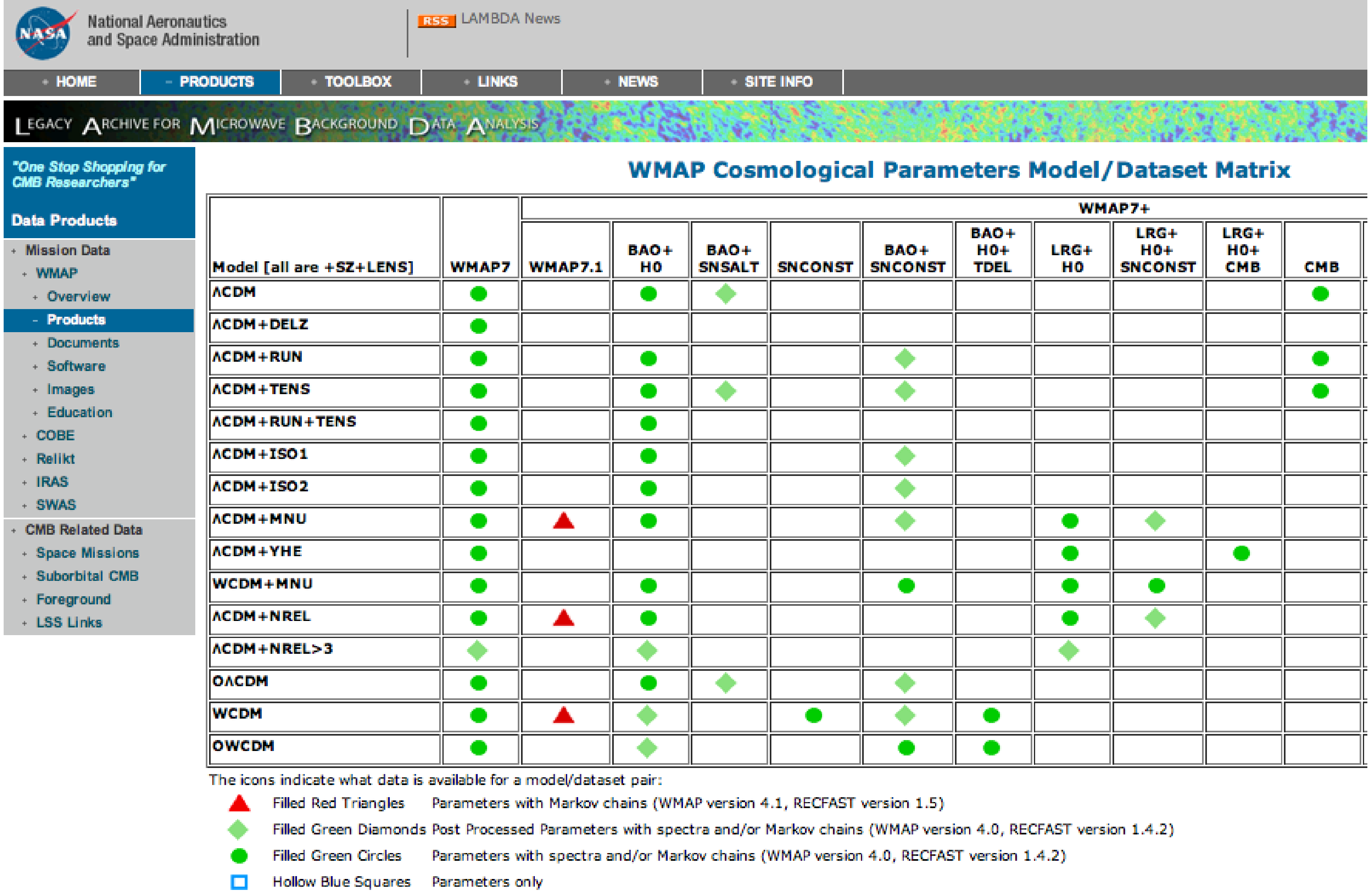


Reid et al ‘10

Figure 4. The profile likelihood defined in Section 2.3 in bins of $\Delta(\sum m_\nu) = 0.2$ eV for the ΛCDM model with WMAP5 only (crosses), WMAP5+maxBCG (stars), WMAP5+H_0 (triangles), and WMAP5+maxBCG+H_0 (diamonds). The black curves overlay a quadratic fit to these points, illustrating that a Gaussian curve provides a good fit to this one-dimensional distribution.

You can do it yourself! http://lambda.gsfc.nasa.gov

In particular:
http://lambda.gsfc.nasa.gov/product/map/drX/parameters.cfm



WMAP Cosmological Parameters Model/Dataset Matrix

Model [all are +SZ+LENS]	WMAP7	WMAP7+									
		WMAP7.1	BAO+ H0	BAO+ SNSALT	SNCONST	BAO+ SNCONST	BAO+ H0+ TDEL	LRG+ H0	LRG+ H0+ SNCONST	LRG+ H0+ CMB	CMB
ΛCDM	●		●	◆							●
ΛCDM+DELZ	●										
ΛCDM+RUN	●		●			◆					●
ΛCDM+TENS	●		●	◆		◆					●
ΛCDM+RUN+TENS	●		●								
ΛCDM+ISO1	●		●			◆					
ΛCDM+ISO2	●		●			◆					
ΛCDM+MNU	●	▲	●			◆		●	◆		
ΛCDM+YHE	●							●		●	
WCDM+MNU	●		●			●		●	●		
ΛCDM+NREL	●	▲	●					●	◆		
ΛCDM+NREL>3	◆		◆					◆			
OΛCDM	●		●	◆		◆					
WCDM	●	▲	◆		●	◆	●				
OWCDM	●		◆			●	●				

The icons indicate what data is available for a model/dataset pair:

▲	Filled Red Triangles	Parameters with Markov chains (WMAP version 4.1, RECFAST version 1.5)
◆	Filled Green Diamonds	Post Processed Parameters with spectra and/or Markov chains (WMAP version 4.0, RECFAST version 1.4.2)
●	Filled Green Circles	Parameters with spectra and/or Markov chains (WMAP version 4.0, RECFAST version 1.4.2)
☐	Hollow Blue Squares	Parameters only

Beyond parameter fitting: model testing

Akaike Information criterion (Akaike 1974; Liddle 04)

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

k=Number of parameters

Bayesian Information criterion (Schwarz 78, Liddle 04)

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

N=Number of data points

Bayesian Evidence

$$E = \int \mathcal{L}(\theta) \mathrm{Pr}(\theta) d\theta \,.$$

it does not focus on the best-fitting parameters of the model, but rather asks “of all the parameter values you thought were viable before the data came along, how well on average did they fit the data?”

Computationally expensive! (there are packages to help out there e.g. cosmonest)

$$\mathrm{P}(H \mid D) = \frac{\mathrm{P}(D \mid H)\mathrm{P}(H)}{\mathrm{P}(D)}$$

Bayes

$$\mathrm{P}(\boldsymbol{\theta} \mid D, H) = \frac{\mathrm{P}(D \mid \boldsymbol{\theta}, H)\mathrm{P}(\boldsymbol{\theta} | H)}{\mathrm{P}(D \mid H)}$$

Bayes, for parameter fitting

$$D_{KL} \equiv \int p(\Theta|D) \ln \frac{p(\Theta|D)}{p(\Theta)} d\Theta.$$

$$\mathcal{E} \equiv \mathrm{P}(D \mid H) = \int \mathrm{d}^n \boldsymbol{\theta} \mathrm{P}(D \mid \boldsymbol{\theta}, H)\mathrm{P}(\boldsymbol{\theta} \mid H)$$

Bayes for the MODEL itself

Use RATIOS!

Suggested exercises

- Go and download the H(z) data from table 2 of the link from http://icc.ub.edu/~liciaverde/clocks.html make a plot in the Ω_m -$\Omega_\Lambda$plane marginalizing over Ho. You can do that using a grid or using a MCMC approach.
- Add a prior given by the measurement of Ho of Riess et al. http://arxiv.org/pdf/0905.0695 without re-running!
- Download one of the WMAP chains, plot confidence limits for a few parameters and for an example of couple of parameters.
- Importance-sample it to add information from e.g. the Ho measurement or the H(z) measurements.
- Or try to compute profile likelihood for one of the parameters and compare the results with the standard MCMC error.

If you are familiar with numerical integrals you can try the SNeIA sample

e.g., http://supernova.lbl.gov/Union/

If you are a wizard with computers you can try to install cosmomc and run chains.

Beyond parameter fitting

Going minimally parametric

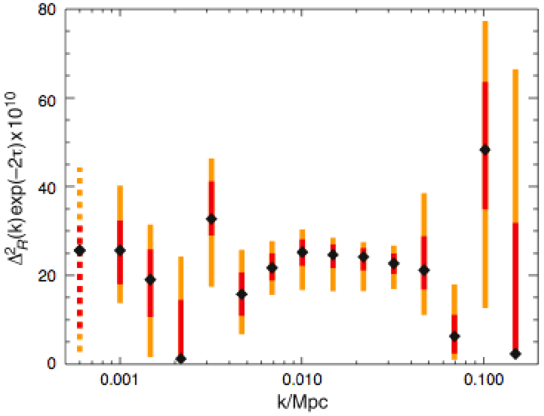
Working example: the shape of the primordial P(k)

Parameter fitting e.g., : $P(k)= A\ (k/k0)^{n-1}$

Other popular options are:

Instead of fitting a function to the data, use a basis function (wavelets, principal components etc…)

Use bins



Piecewise linear

Spergel et al 07

How do you know you are not “fitting the noise”?

How do you know the model (e.g. power law, running) is OK?

Minimally parametric technique
Based on smoothing splines JUST AN EXAMPLE!
(Gaussian processes are fashonable these days`)

Splines: Piecewise polynomial (usually cubic) fit. Describe P(k) with splines

Smoothing: Suitable for looking for smooth deviations from power laws

Knots: Discrete values of k,ki. P(ki) will be “free” parameters.
Do spline for the knots

Sealfon et al (2005); Verde, Peris (2008); Peiris, Verde (2010); Bird et al (2010)

How do you know you are not “fitting the noise”?

How do you know the model (e.g. power law, running) is OK?

Minimally parametric technique (in 3 “easy” steps):

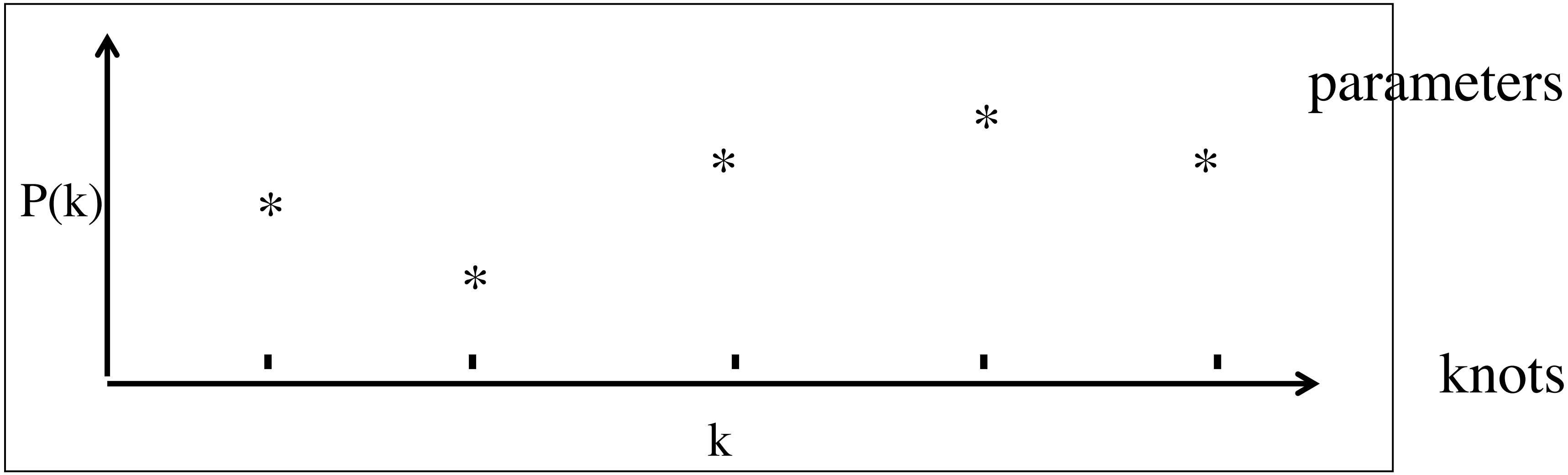
1)Select # knots and use a piecewise cubic spline

2)Penalize the likelihood for the “wiggliness”

3)Use CROSS VALIDATION to chose optimal penalty

full analysis is computationally expensive!

1)



2)

$$\log \mathcal{L} = \log \mathcal{L}(\mathrm{Data}|\alpha, P(k)) + \lambda \int_k (P(k)'')^2 dk$$

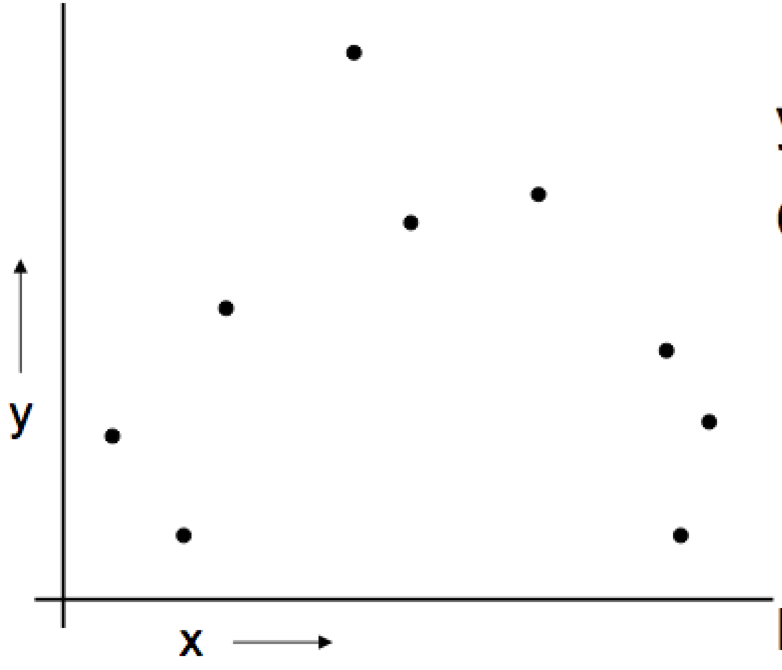
HOW TO SELECT THE BEST PENALTY?

3) Beware of overfitting:

Cross Validation is a powerful technique to make sure one is not fitting the noise

Cross Validation is a powerful technique to make sure one is not fitting the noise

A Regression Problem

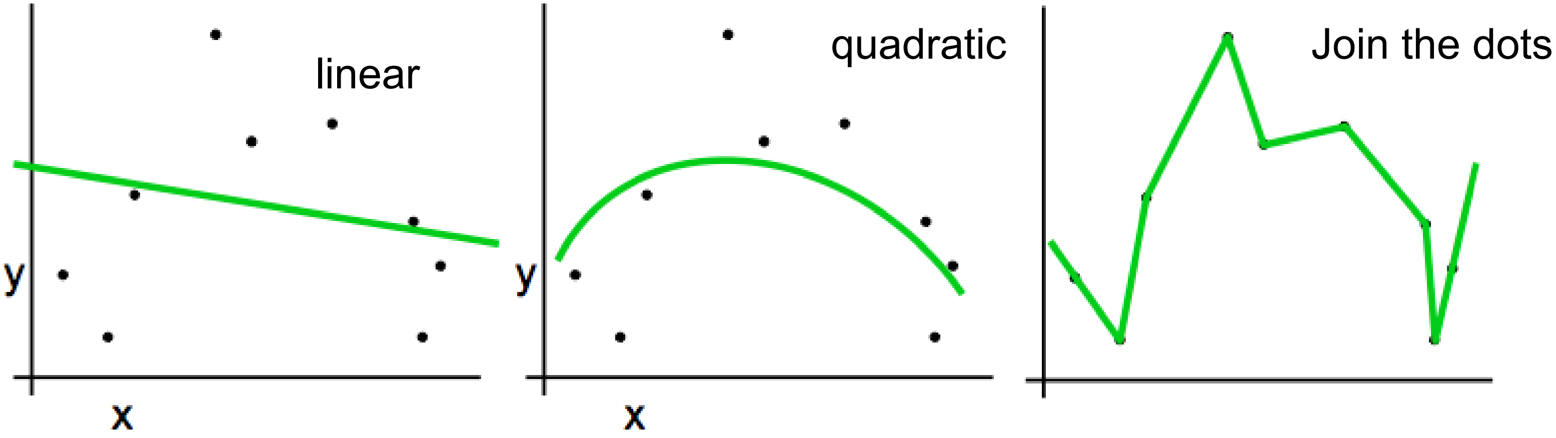


y = f(x) + noise

Can we learn f from this data?

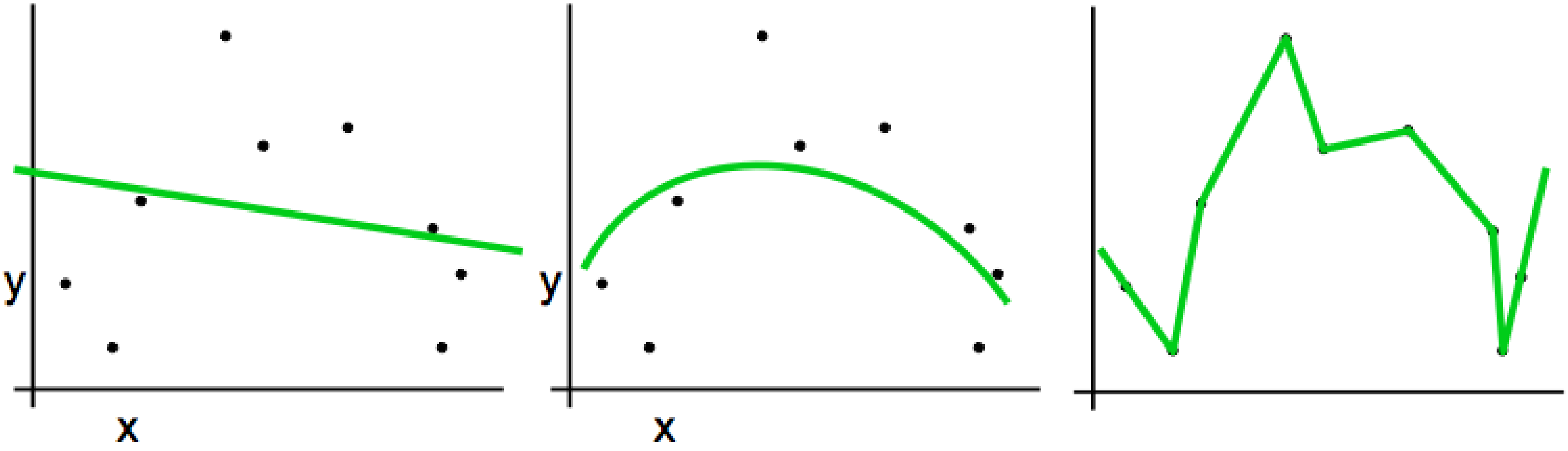
Let's consider three methods…

Which is best?



Why not choose the method with the best fit to the data?

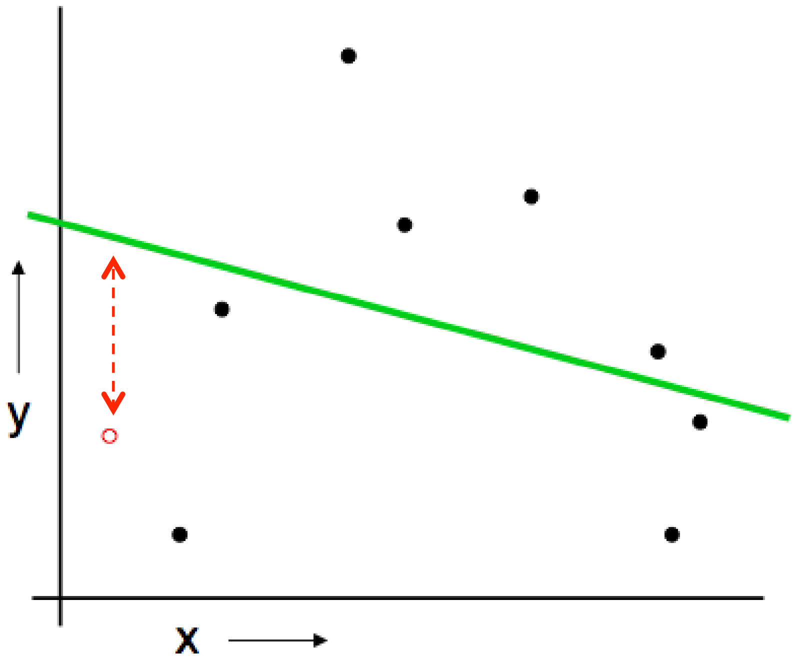
What do we really want?



Why not choose the method with the best fit to the data?

"How well are you going to predict future data drawn from the same distribution?"

Leave one out cross validation:

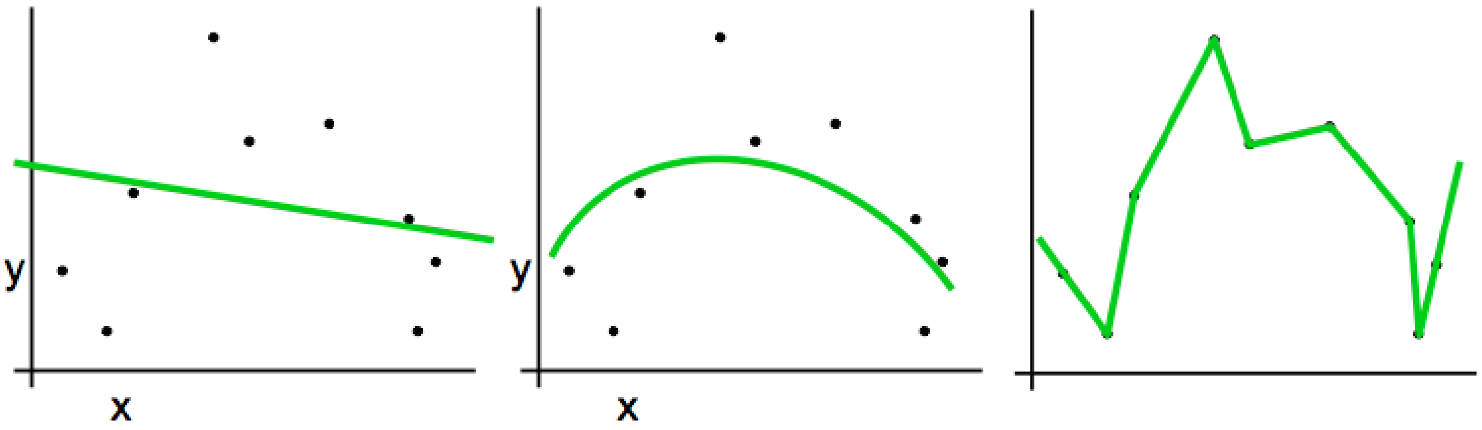


For k=1 to R

1. Let (x_k,y_k) be the k^{th} record
2. Temporarily remove (x_k,y_k) from the dataset
3. Train on the remaining R-1 datapoints (example shown for linear model)
4. Note your error (x_k,y_k)

When you've done all points, report the mean error(CV score)

In this example:



MSE_{LOOCV} = 2.12

MSE_{LOOCV} =0.962

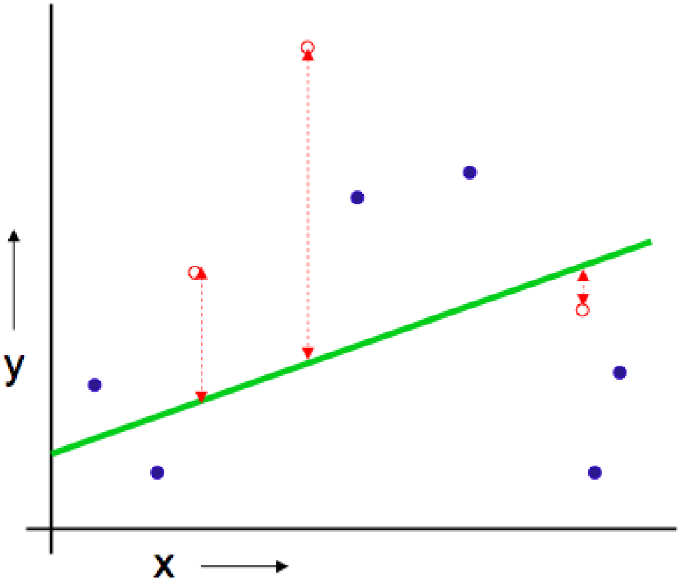
MSE_{LOOCV} =3.33

CV score

Leave one out CV is the ideal: does not waste much data but it is very expensive

The training/test set approach is similar to leave one out CV.

Train on subset of the data



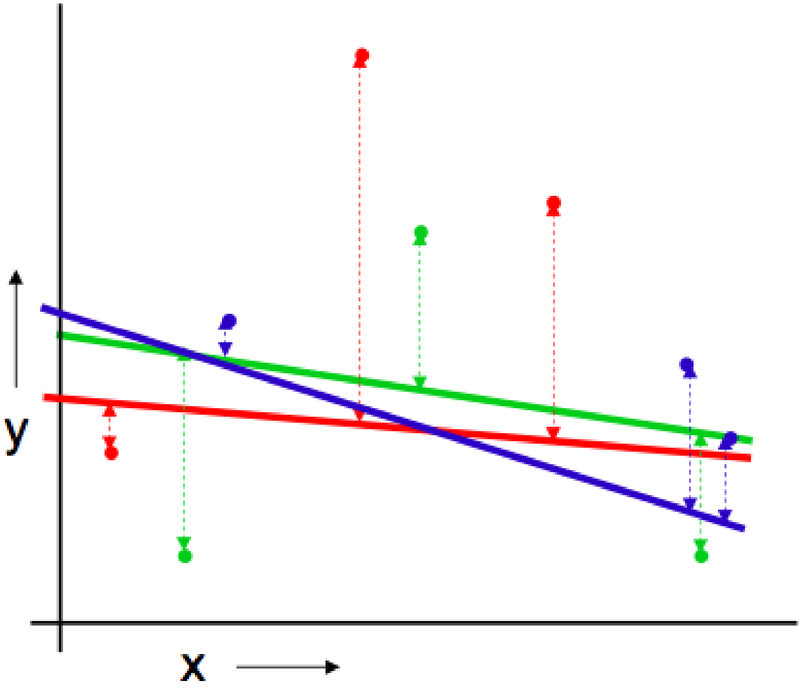
(Linear regression example)

1. Randomly choose 30% of the data to be in a test set
2. The remainder is a training set
3. Perform your regression on the training set
4. Estimate your future performance with the test set

(this may remind you of training sets for photo-z)

Statisticians prefer:

k-fold Cross Validation



Linear Regression

Randomly break the dataset into k partitions (in our example we'll have k=3 partitions colored Red Green and Blue)

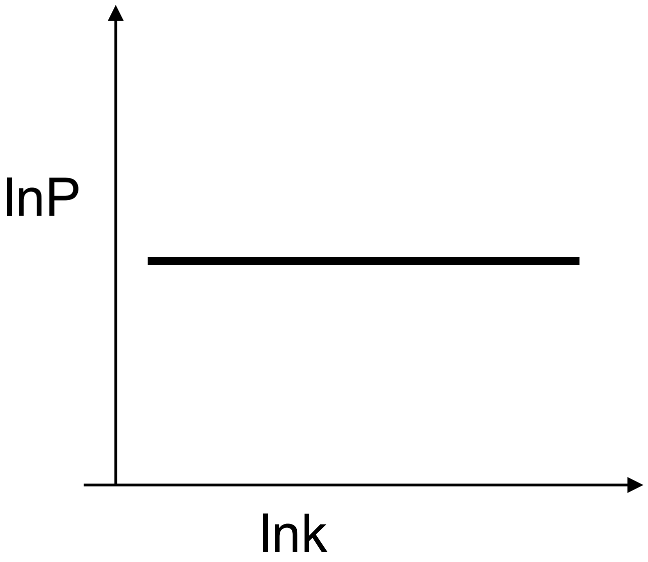
For the red partition: Train on all the points not in the red partition. Find the test-set sum of errors on the red points.

For the green partition: Train on all the points not in the green partition. Find the test-set sum of errors on the green points.

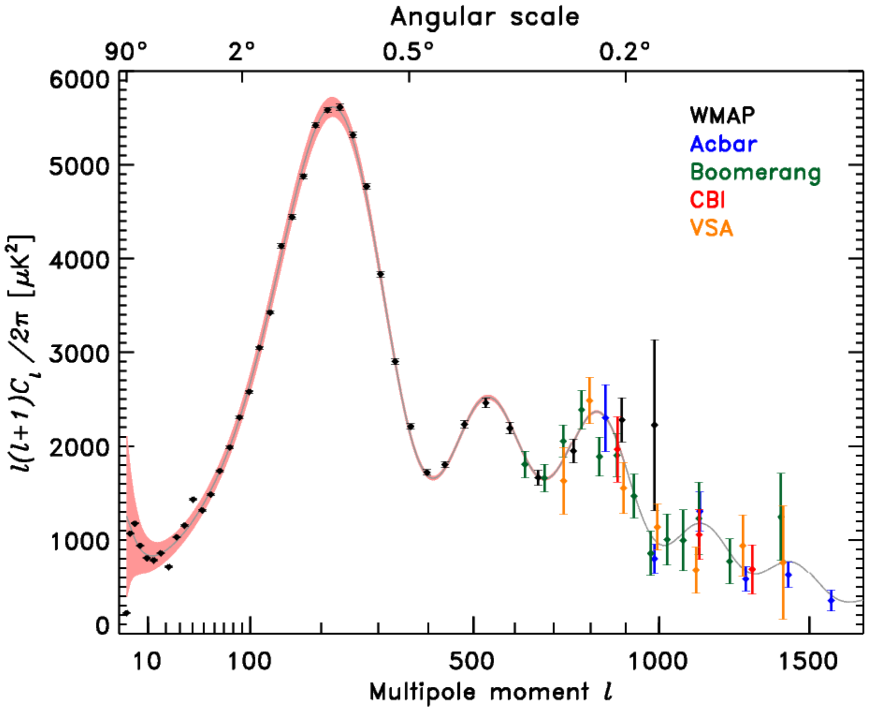
For the blue partition: Train on all the points not in the blue partition. Find the test-set sum of errors on the blue points.

Then report the mean error
& compare different models

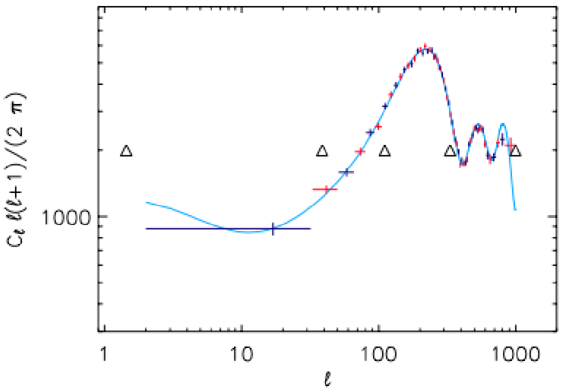
While “leave 1 out” CV would be ideal, it is too computationally intensive; we do 2-fold CV.

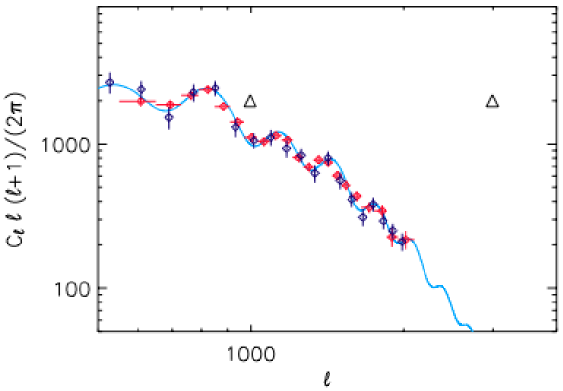


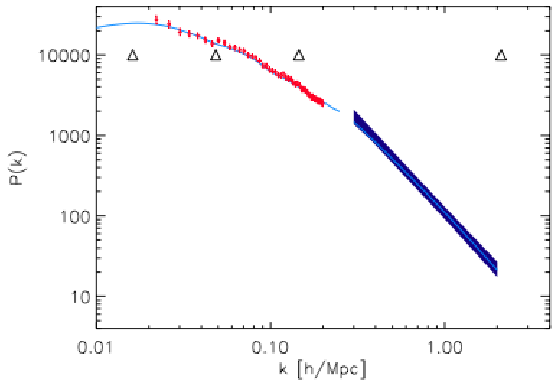
Non-analytic Transfer function

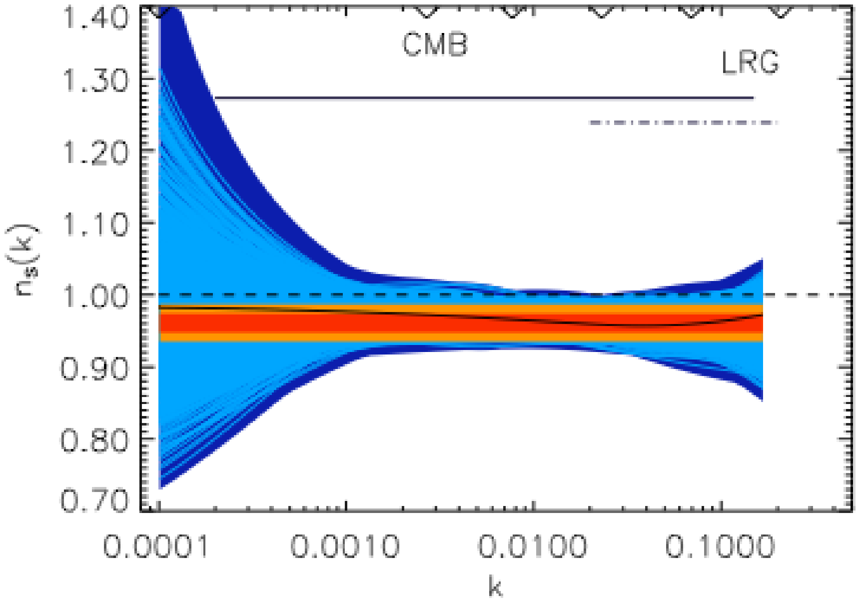


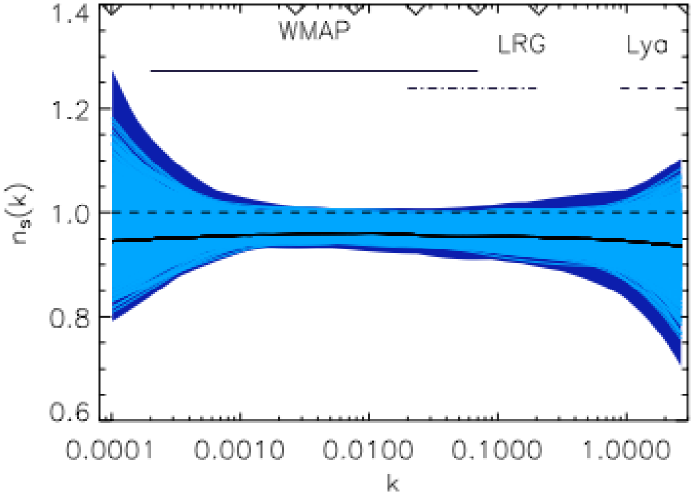
Split the data in 2 samples (CV1, CV2)
for each penalty value do a MCMC.
Compute the likelihood for the best fit model from CV1 and data of CV2 and viceversa.
The sum of these two log likelihoods give the CV score.
The optimal penalty is the one that minimizes the CV score.











Peiris, Verde 2010

Statistical vs systematic errors

Statistics can tell you how to deal with statistical errors

As a data set grows, the statistical errors shrink; systematic errors do not shrink

You’ve got a problem.

Rumsfeld can help:

There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.

Donald Rumsfeld

Jokes aside: some interesting literature has appeared in the past 2 yr…

Introduction to Fisher

Cosmological examples of
<u>hypothesis testing</u>:

Are CMB data consistent with the hypothesis of Gaussian initial fluctuations?
Are CMB+LSS data consistent with the hypothesis of a spatially flat Universe?

Parameter estimation:
What is the value of the matter density parameter (in the LCDM model)?
And what it the value of the Hubble parameter today?

Model selection:
Is there evidence for a non-flat Universe?
Is there evidence for a non-constant dark energy?

Back to likelihoods

X data vector (random variable)

e.g., T value in pixels of CMB map,
Fourier coefficients of density of survey
etc.

$\Theta = (\theta_1, \theta_2, ..., \theta_m).$ Vector of model parameters

$L(x; \Theta)$; Probability distribution of x

Since x is a random variable also Θ will be so … ideally:

$\langle \Theta \rangle = \Theta_0$; UNBIASED

$\Delta\theta_i \equiv \left(\langle \theta_i^2 \rangle - \langle \theta_i \rangle^2 \right)^{1/2}$ Minimize this i.e. errors

i.e. we want the best unbiased estimator

Fisher information matrix

$$F_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle \qquad \mathcal{L} \equiv -\ln L.$$

Fisher 1935

The maximum likelihood estimator is Θ_{ML} that maximizes $L(x;\Theta)$

a number of powerful theorems apply (e.g., Kendall & Stuart 1969):

For any unbiased estimator:

$\Delta\theta_i \geq (F^{-1})_{ii}^{1/2}$ or $\Delta\theta_i \geq 1/\sqrt{F_{ii}}$. Cramer-Rao inequality

If there is the best unbiased estimator it is the ML or function of thereof

The ML estimator is asymptotically the best unbiased estimator

Fisher matrix approach

(Fisher 1935)

How well can a future experiment do?

(quick and easy but not always accurate)

Fisher information matrix $F_{ij} = \left\langle \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} \right\rangle$ $L = -\ln \mathcal{L}$

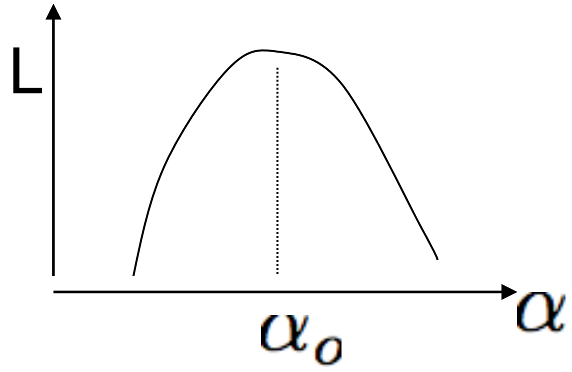
e.g.

$$1/2(data - theory(\vec{\alpha}))C^{-1}(data - theory(\vec{\alpha})) \simeq$$
$$1/2(fiducial - theory(\vec{\alpha}))C^{-1}(fiducial - theory(\vec{\alpha}))$$

To develop intuition, one parameter case, Gaussian likelihood.

$$L = 1/2(\alpha - \alpha_0)^2/\sigma_\alpha^2$$

Second deriv. w.r.t. $\alpha \longrightarrow 1/\sigma_\alpha^2$



In general

Expand in Taylor series around α_0

$$\Delta L = \frac{1}{2}\frac{d^2L}{d\alpha^2}(\alpha - \alpha_0)^2$$

First deriv 0 by construction

Should remind you of χ^2

All it is: Quadratic expansion around the max

$1/\sqrt{d^2L/d\alpha^2}$ is the 1 sigma displacement of α from α_0

Like a measure for the width of the peak…..

Multi dimensional case…

$$\sigma^2_{\alpha_i,\alpha_j} \geq (\mathbf{F}^{-1})_{ij}$$

Parameters covariance

$$\sigma_{\alpha_i} \geq \sqrt{\frac{1}{F_{ii}}}$$

If all other parameters are fixed

$$\sigma_{\alpha} = (\mathbf{F}^{-1})_{ii}^{1/2}$$

Marginalized errors

Matrix inversion performed

Conditional and marginal errors

Minumum error on α_i if all other parameters are known

$$\sigma_{\alpha_i} \geq \sqrt{\frac{1}{F_{ii}}}$$

ALMOST NEVER USED

The marginal distribution of α_i :integrate over other parameters

$$p(\alpha_1) = \int d\alpha_2..d\alpha_N p(\alpha)$$

$$\sigma_\alpha \geq (\mathbf{F}^{-1})_{ii}^{1/2}$$

What are we really saying?

$$\langle p(\mathbf{x}|\boldsymbol{\theta}, M)\rangle = L_0 \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)_\alpha \mathsf{F}_{\alpha\beta}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)_\beta\right]$$

parameters

$$\propto \frac{1}{\sqrt{detC}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t\right]$$

data

This is sometimes called: Laplace approximation

Explicit calculation

$$2\mathcal{L} = \ln\det \boldsymbol{C} + (\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t$$

Dropped irrelevant constant
Assumed Gaussianity

$$\boldsymbol{C} = \langle(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t\rangle$$

DATA covariance (can depend on the parameters)

You can show that:

$$\boxed{F_{\alpha\beta} = \langle \mathcal{L}_{,\alpha\beta} \rangle = \frac{1}{2}\mathrm{Tr}[C^{-1}C_{,\alpha}C^{-1}C_{,\beta} + C^{-1}M_{\alpha\beta}],}$$

This simplifies
in specific cases

where $M_{\alpha\beta} \equiv \langle D_{,\alpha\beta} \rangle = \boldsymbol{\mu}_{,\alpha}\,\boldsymbol{\mu}_{,\beta}^T + \boldsymbol{\mu}_{,\beta}\,\boldsymbol{\mu}_{,\alpha}^T$ and $C_{,\alpha} \equiv \frac{\partial}{\partial\theta_\alpha}C.$

REQUIRES NO DATA!

Other option:

Compute explicitly:

For the data

$$-\frac{\partial^2 \ln L}{\partial\theta_\alpha \partial\theta_\beta}$$

Taking the data to be = a fiducial model

Numerical second derivative: beware!

ALWAYS TEST STABILITY OF DERIVATIVES!!!!

You can compute the Fisher matrix
BEFORE you do the experiment.

You can then use it as a tool
to design or optimize experiments

Within the assumptions made, now you know everything!

Say you have worked with 5 parameters but now you want to keep parameters 1 and 3 fixed at fiducial model….

Take the submatrix (1,3) of F_{ij}

Say you have a 5 parameters Fisher and you want to plot the joint 2D forecasted constraint for parameters 2 and 4 marginalized over the other parameters

Invert Fij, take the submatrix (2,4) invert this back. Call this Q.
Q describes a Gaussian 2D likelihood i.e.

$$\tilde{\chi}^2 = \sum_{kq} (\alpha_k - \alpha_k^{fid.}) Q_{jq} (\alpha_q - \alpha_q^{fid})$$

$$\Delta = \delta\vec{\alpha} Q \delta\vec{\alpha}$$

And look up Δ in the "famous" table

You can also draw the ellipses!

What if you want to reparameterize?

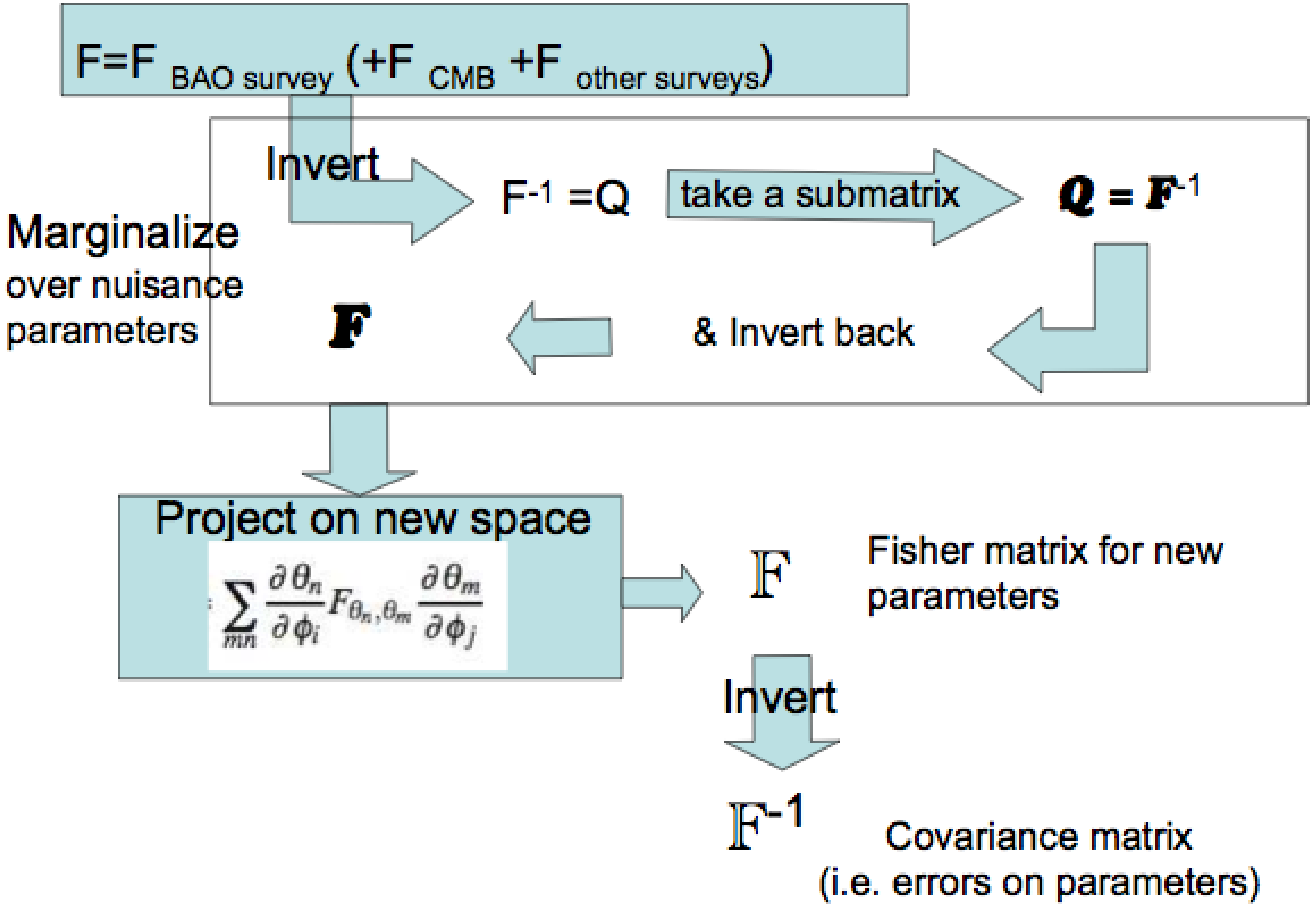
Typical example
CMB: parameters

Now you want to combine with BAO constraints

BAO parameters H(z), Da(z)

$$F_{\varphi_i,\phi_j} = \sum_{mn} \frac{\partial \theta_n}{\partial \phi_i} F_{\theta_n,\theta_m} \frac{\partial \theta_m}{\partial \phi_j}$$

This is what you do….



$F=F_{\text{BAO survey}}\ (+F_{\text{CMB}}\ +F_{\text{other surveys}})$

Marginalize
over nuisance
parameters

Invert

$F^{-1}=Q$

take a submatrix

$\boldsymbol{Q}=\boldsymbol{F}^{-1}$

& Invert back

$\boldsymbol{F}$

Project on new space

$$\sum_{mn} \frac{\partial \theta_n}{\partial \phi_i} F_{\theta_n,\theta_m} \frac{\partial \theta_m}{\partial \phi_j}$$

$\mathbb{F}$ Fisher matrix for new parameters

Invert

$\mathbb{F}^{-1}$ Covariance matrix
(i.e. errors on parameters)

Practical tools: icosmo

iCosmo

http://www.icosmo.org/Initiative_Web/Initiative.html

Notes from a tutorial course on icosmo link from:
http://icc.ub.edu/~liciaverde/ERCtraining.html

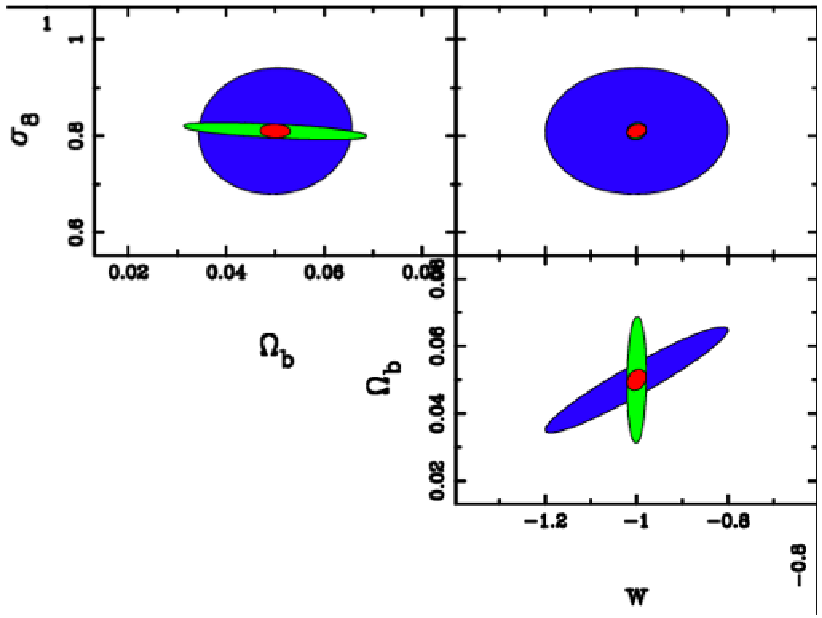


Fig. 1. The expected error ellipses for cosmological parameters (σ_8, baryon density parameter Ω_b, and dark energy equation of state $w \equiv p/\rho c^2$) from a 3D weak lensing survey of 1000 square degrees, with a median redshift of 1 and a photometric redshift error of 0.15. Probabilities are marginalised over all other parameters, except that $n = 1$ and a flat Universe are assumed. Dark ellipses represent a prior from WMAP, pale represents the 3D lensing survey alone, and the central ellipses show the combination (from Kitching T., priv. comm.).

aside

Popular CMB Fisher matrix

$$F_{ij} = \sum_{\ell} \frac{(2\ell+1)}{2} \frac{\frac{\partial C_\ell}{\partial \alpha_j} \frac{\partial C_\ell}{\partial \alpha_j}}{(C_\ell + \mathcal{N} e^{\sigma^2 \ell^2})^2}$$

Approximations?

Covariance?

Applicability: Fisher vs non Fisher

Fisher and systematic errors

Can the Fisher approach account for systematic errors?

In general NO

But there’s an exception

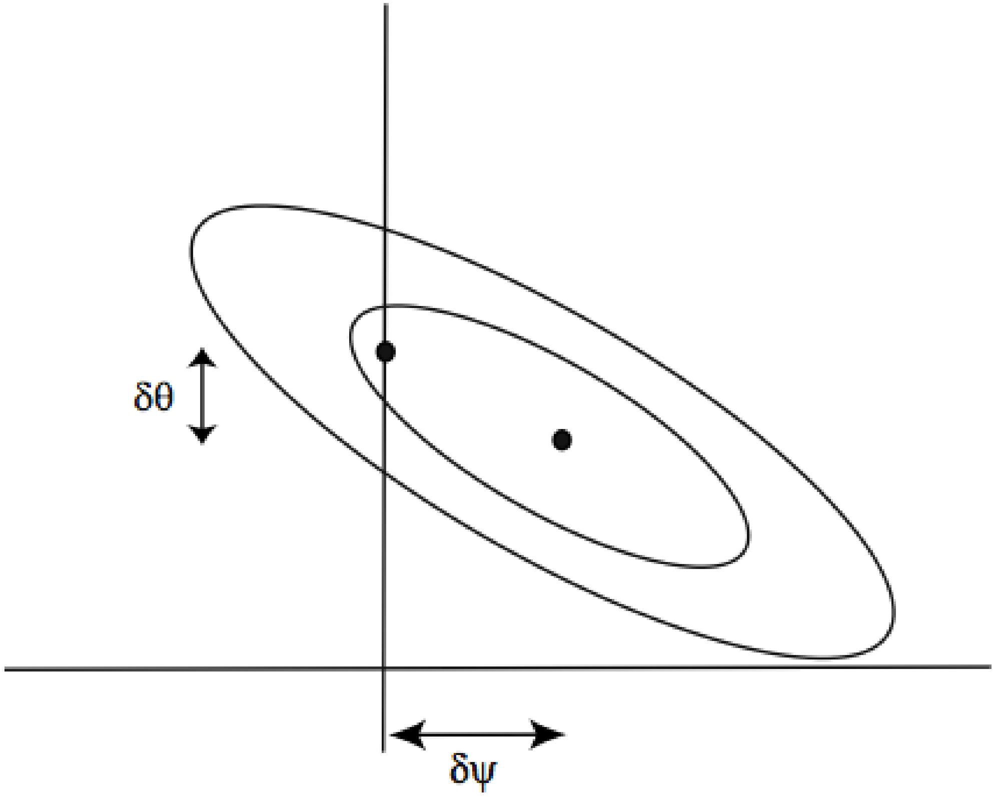
Imagine you have two competing models M and M’: M with n parameters and M’ with n’, where n’<n. Say also that the two models are NESTED, i.e. M’ is a particular case of M

If the true underlying model is M and you instead fit the data with M’, the maximum expected likelihood will not be at the correct values of the parameters: if n-n’=p, the n’ parametrs shift from their value to compensate for the fact that p parameters are kept fixed at “wrong” values.
If the p parameters differ by $\delta\psi_\alpha$ fro their true values, the other parameters are shifted by:

$$\delta\boldsymbol{\theta}'_\alpha = -(\mathsf{F}'^{-1})_{\alpha\beta}\mathsf{G}_{\beta\zeta}\delta\psi_\zeta \qquad \alpha,\beta = 1\ldots n', \zeta = 1\ldots p$$

$$\mathsf{G}_{\beta\zeta} = \frac{1}{2}\mathrm{Tr}\left[\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}\mathsf{C}_{,\zeta} + \mathsf{C}^{-1}(\mu_{,\zeta}\mu^T_{,\beta} + \mu_{,\beta}\mu^T_{,\zeta})\right],$$

This is what’s going on:



This is very useful: e.g., isocurvatures, delayed recombination, neutrinos…

And there's more

Remember the evidence?

$$p(\mathbf{x}|M) = \int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}M)p(\boldsymbol{\theta}|M)$$

Back to models M and M':

$$\frac{p(M'|\mathbf{x})}{p(M|\mathbf{x})} = \frac{p(M')}{p(M)} \boxed{\frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}'M')p(\boldsymbol{\theta}'|M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}M)p(\boldsymbol{\theta}|M)}.}$$

Bayes factor B

For non-commital priors

$$p(M') = p(M).$$

What’ s the Bayes factor?

$$B \equiv \frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}' M')p(\boldsymbol{\theta}'|M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta} M)p(\boldsymbol{\theta}|M)}$$

M will have higher likelihood (or as high) but the evidence will favour the simpler model if the fit is nearly as good, through the smaller prior volume.

For uniform separable priors:

$$B = \frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}', M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}, M)} \frac{\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n}{\Delta\boldsymbol{\theta}'_1 \ldots \Delta\boldsymbol{\theta}'_{n'}}.$$

If prior is wide enough to encompass the”support” of the likelihood

$$\frac{\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n}{\Delta\boldsymbol{\theta}'_1 \ldots \Delta\boldsymbol{\theta}'_{n'}} = \Delta\boldsymbol{\theta}_{n'+1} \ldots \Delta\boldsymbol{\theta}_{n'+p}$$

Requires a painful multi-dimensional integration, but….

Laplace approximation and Fisher to the rescue!

$$\langle B\rangle = (2\pi)^{-p/2}\frac{\sqrt{\det \mathsf{F}}}{\sqrt{\det \mathsf{F}'}}\frac{L_0'}{L_0}\Delta\boldsymbol{\theta}_{n'+1}\ldots\Delta\boldsymbol{\theta}_{n'+p}.$$

Where you already know how to compute L’o and Lo

And you’ll see that this simplifies to:

$$\langle B\rangle = (2\pi)^{-p/2}\frac{\sqrt{\det \mathsf{F}}}{\sqrt{\det \mathsf{F}'}}\exp\left(-\frac{1}{2}\delta\boldsymbol{\theta}_\alpha \mathsf{F}_{\alpha\beta}\delta\boldsymbol{\theta}_\beta\right)\prod_{q=1}^{p}\Delta\boldsymbol{\theta}_{n'+q}.$$

$$\delta\boldsymbol{\theta}_\alpha = \delta\boldsymbol{\theta}'_\alpha \text{ for } \alpha \le n' \qquad\qquad \delta\boldsymbol{\theta}_\alpha = \delta\psi_{\alpha-n'} \qquad \alpha > n'.$$

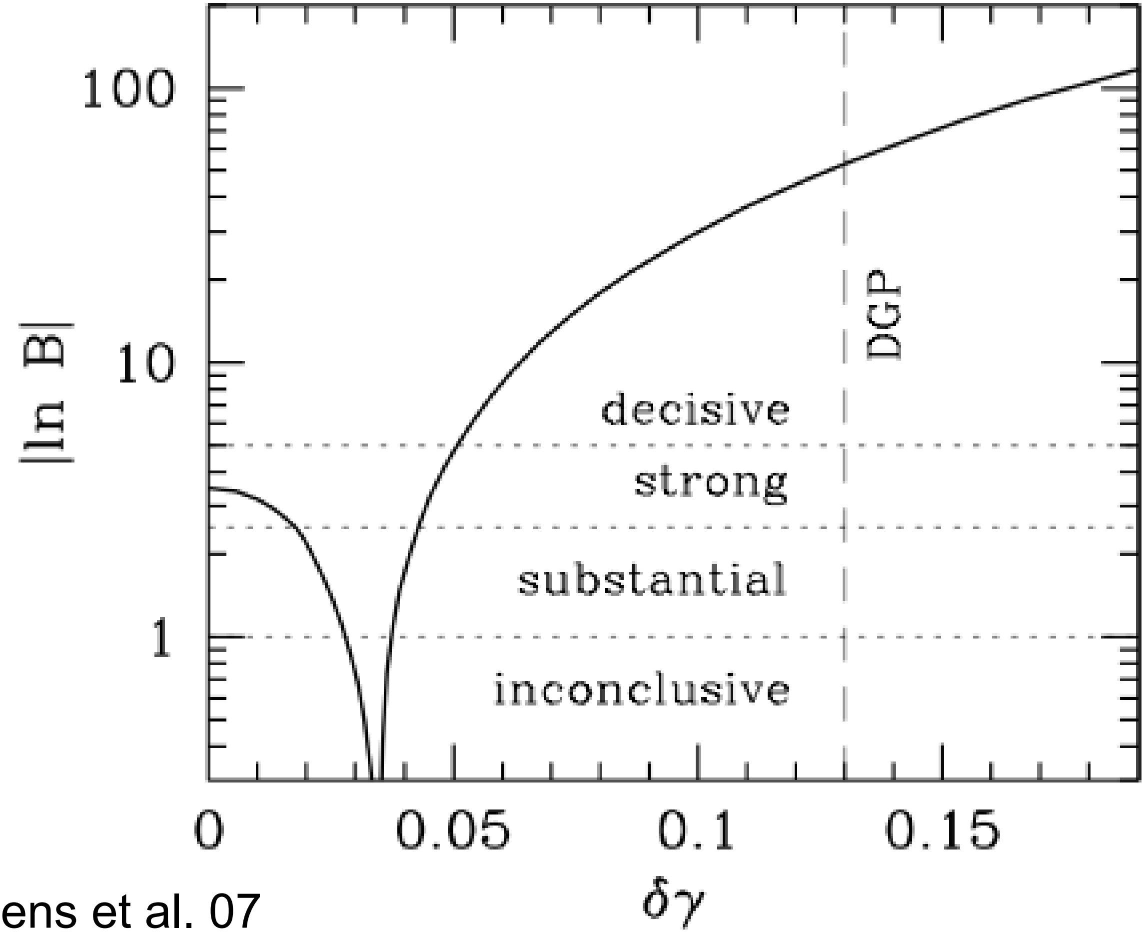
Example:

Is gravity described by General relativity?

$d \ln \delta / d \ln \Omega_m = \gamma$, where $\gamma = 0.55$ for GR

e.g., DGP:

$\gamma = 0.68$



Euclid+Planck

Heavens et al. 07

Real World Issues: CMB

How is the information extracted?

For gaussian initial conditions the power spectrum completely characterizes the statistical properties of the CMB temperature fluctuations. Therefore the information enclosed in the mega-pixel CMB maps is *compressed* into a CMB angular power spectrum

Higher orders are also important !!!

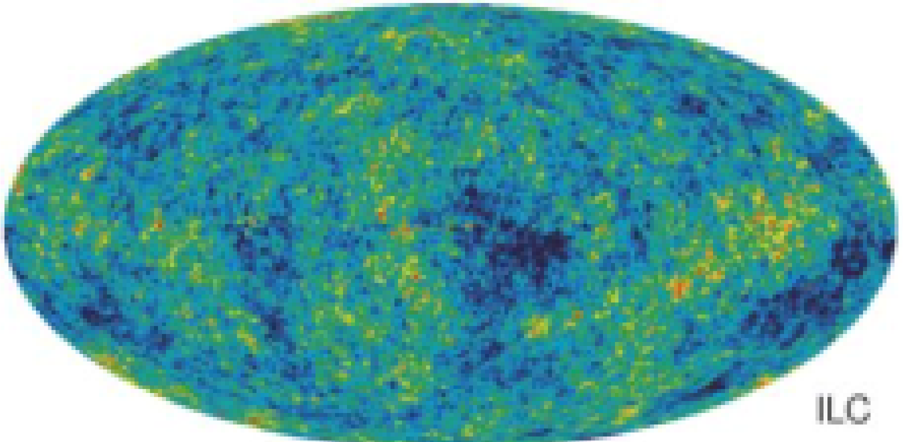
$$\Delta T(\hat{n}) = \sum_{\ell > 0} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}).$$

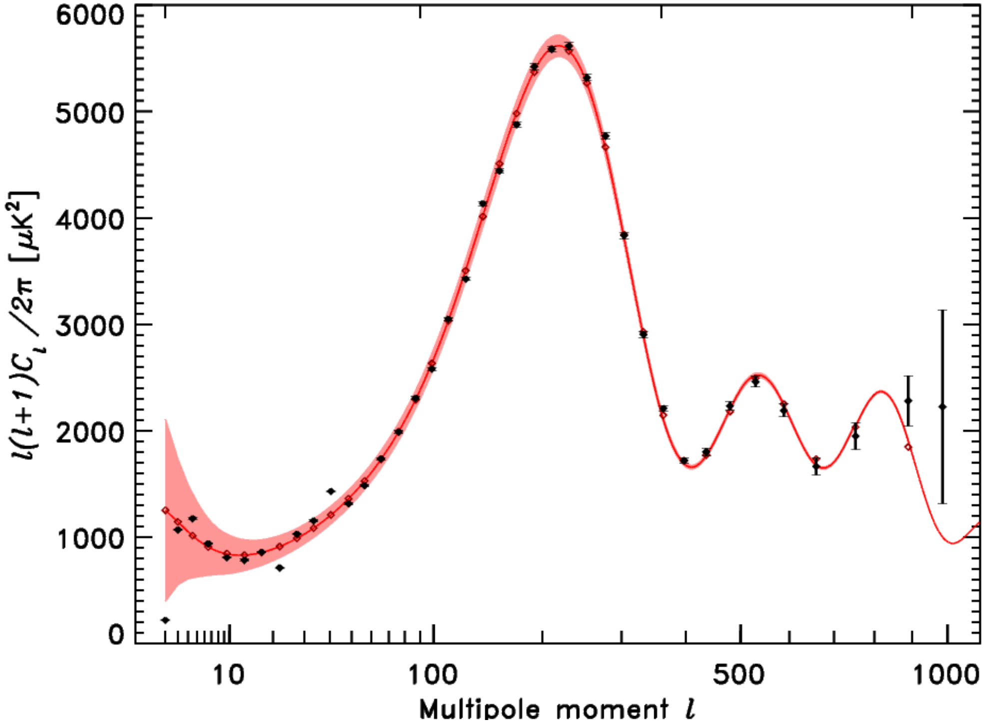
$$a_{\ell m} = \int d\Omega_n \Delta T(\hat{n}) Y^*_{\ell m}(\hat{n}).$$

$$< |a_{\ell m}|^2 >= \langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell$$

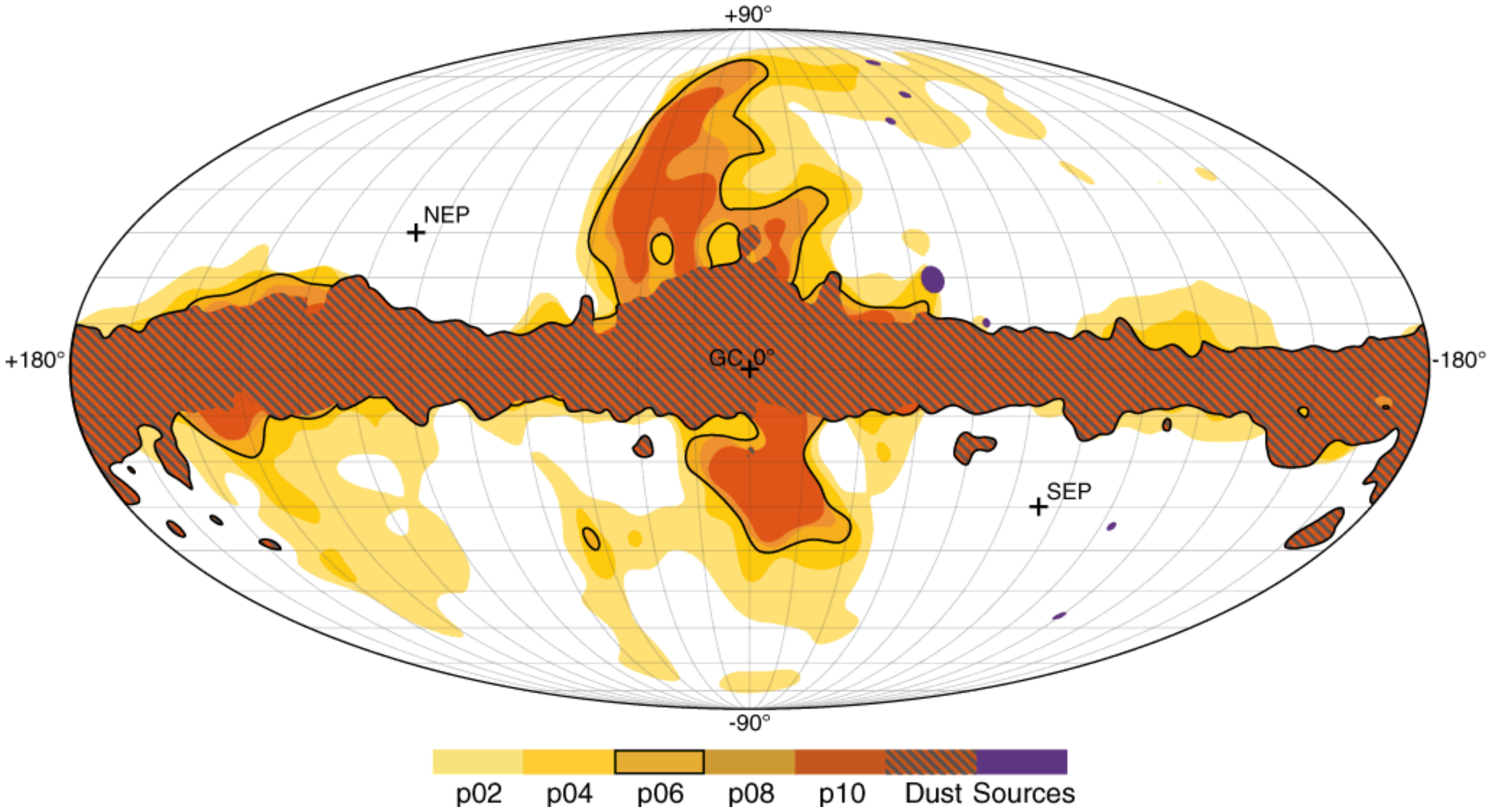
$$C_\ell = \frac{1}{(2\ell + 1)} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$$

Of course there is also polarization: TT, TE, EE, BB
and cross correlations…. And lensing ….
But let’s start from the basics





In principle it is also possible to extract cosmological information directly from CMB maps



Instrumental noise, finite resolution, foregrounds, sky cut…

Real world issues

$$\widetilde{a}_{\ell m} = \int d\Omega_n \Delta T(\hat{n}) W(\hat{n}) Y^*_{\ell m}(\hat{n})$$

Sky cut

$$\widetilde{a}_{\ell m} = \Omega_p \sum_p \Delta T(p) W(p) Y^*_{\ell m}(p)$$

pixelization

Solid angle subtended by pixel

Warning: this can get nasty…

Pseudo-Cl (Hivon et al 2002)

$$\widetilde{C}_\ell = \frac{1}{(2\ell+1)} \sum_{m=-\ell}^{\ell} |\widetilde{a}_{\ell m}|^2$$

Clearly $\widetilde{C}_\ell \neq C_\ell$ but

$$\langle \widetilde{C}_\ell \rangle = \sum_{\ell} G_{\ell\ell'} \langle C_{\ell'} \rangle$$

Mode coupling

$$G_{\ell_1 \ell_2} = \frac{2\ell_2+1}{4\pi} \sum_{\ell_3} (2\ell_3+1) W_{\ell_3} \binom{l_1 l_2 l_3}{0\ 0\ 0}^2$$

$$W_\ell = \frac{1}{2\ell+1} \sum_m |W_{\ell m}|^2 \ \text{ and } \ W_{\ell m} = \int d\Omega_n W(\hat{n}) Y^*_{\ell m}(\hat{n})$$

N.B. The window does not need to be only 0 or 1

$$\langle \tilde{C}_\ell \rangle = \sum_{\ell} G_{\ell\ell'} \langle C_{\ell'} \rangle$$

$$P^{meas}(k) = P(k) * W(k)$$

Large scale structure people stop here,
while CMB people, sometimes….

If you are good enough to invert G and identify <Cl> with Cl

$$C_\ell = \sum_{\ell'} G^{-1}_{\ell\ell'} \tilde{C}_{\ell'}$$

Unfortunately this operation is not always possible/doable http://arxiv.org/pdf/astro-ph/9504054.pdf

Noise

$a_{lm} \longrightarrow a_{\ell m}^{signal} + a_{\ell m}^{noise}$

While $\langle a_{\ell m}^{noise} \rangle = 0$,

$$C_\ell^{measured} = C_\ell^{signal} + C_\ell^{noise}$$

$$C_\ell^{noise} = \ell(2\ell + 1) \sum_m |a_{\ell m}^{noise}|^2$$

Biased estimator

Non-zero

Trick!

Use the cross power spectrum….

Does the noise disappear also from the error on the Cl?

noise:

For an experiment with a detector sensitivity of s (usually expressed in $\mu k\sqrt{s}$), the rms per sky (or map) pixel is given by $\sigma_{\text{pix}} = s/\sqrt{t_{pix}}$ where t_{pix} is the observing time spent on each pixel.

Note that for detecting a polarized signal if the instrument need to ”split the photons”, the sensitivity s is at least a factor $\sqrt{2}$ worst than for T (all other characteristics being equal).

For an experiment with negligible beam smearing (i.e. beam smearing much smaller than the pixel size) the noise spectrum per multipole becomes $w = (\sigma^2_{pix}\,\Omega_{pix})^{-1}$ where the pixel solid angle: $\Omega pix = 4\pi\, f_{sky}/N_{pix}$ Thus $C^{noise} = w^{-1}$.

http://arxiv.org/pdf/astro-ph/9504054.pdf beams

Point spread function in optical…

$$T_i = \int d\Omega'_n T(\hat{n}) b(|\hat{n} - \hat{n'}|)$$

convolution

$$C_\ell^{measured} = C_\ell^{sky} e^{-\ell^2 \sigma_b^2}$$

For gaussian beams

$$\sigma_b = 0.425 FWHM.$$

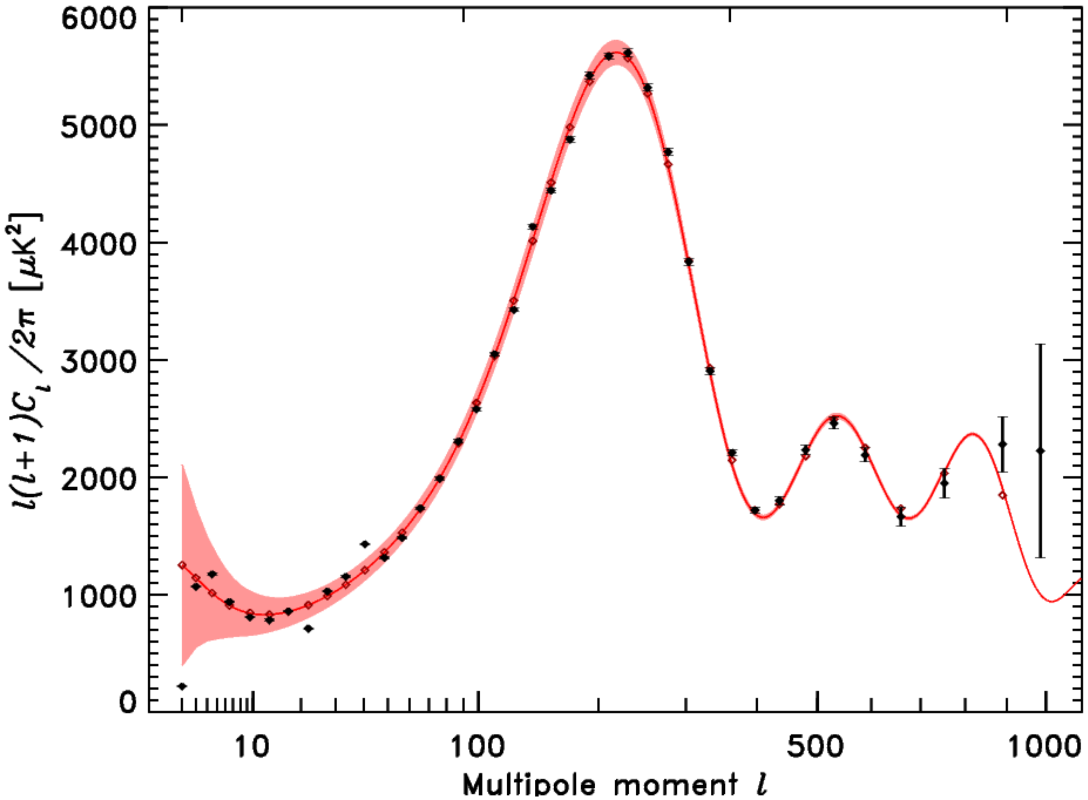
and now with noise

$$C_\ell^{measured} = C_\ell^{sky} e^{-\ell^2 \sigma_b^2} + C_\ell^{noise}$$

Deconvolve for beams

$$C_\ell^{measured'} = C_\ell^{sky} + C_\ell^{noise} e^{\ell^2 \sigma_b^2} .$$

Noise blows up



It is important to know well the beams

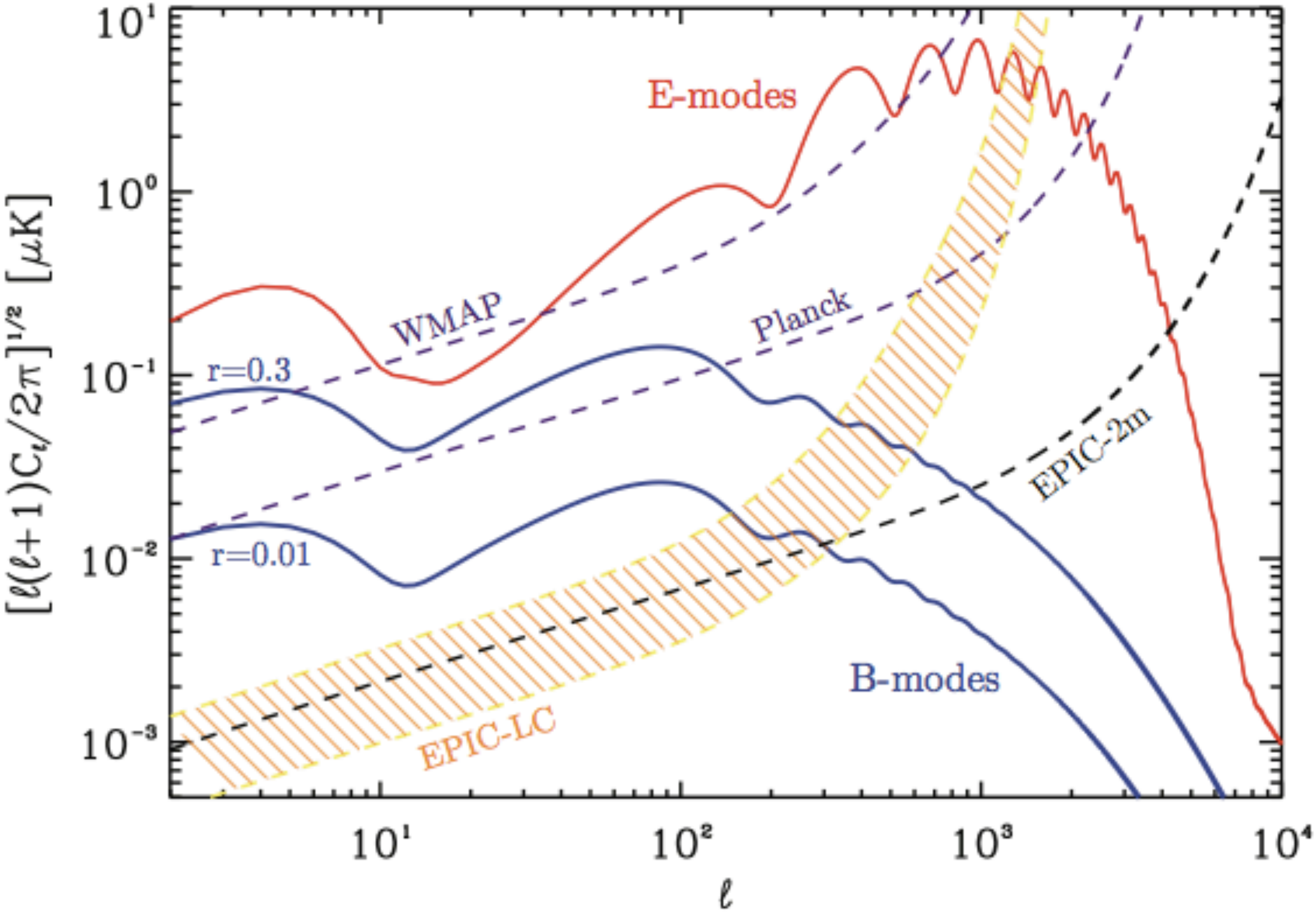


Fig. 1.7 E- and B-mode power spectra for a tensor-to-scalar ratio saturating current bounds, r = 0.3, and for r = 0.01. Shown are also the experimental sensitivities for WMAP, Planck and two different realizations of CMBPol (EPIC-LC and EPIC-2m). (Figure from[Baumann et al.(2009)]).

Exercise: Compute the expression for C_ℓ^{noise} given:

t = observing time

s = detector sensitivity (in $\mu K/\sqrt{s}$)

n = number of detectors

N = number of pixels

f_{sky} = fraction of the sky observed

Assume uniform noise and observing time uniformly distributed.

…..Back to likelihoods

CMB ΔT distribution is close to gaussian,

So the Cl’s are NOT (and at low l CLT does not hold)

$$\mathcal{L}(T|C_\ell^{th}) \propto \frac{\exp[-(TS^{-1}T)/2]}{\sqrt{det(S)}}$$

$$S_{ij} = \sum_\ell \frac{(2\ell+1)}{4\pi} C_\ell^{th} P_\ell(\hat{n}_i \cdot \hat{n}_j)$$

Signal covariance

Given by a
theoretical model

Legendre
polynomials

Now in spherical harmonics

$$\mathcal{L}(T|C_\ell^{th}) \propto \frac{\exp[-1/2|a_{\ell m}|^2/C_\ell^{th}]}{\sqrt{C_\ell^{th}}}$$

Isotropy means that we can sum over $m's$

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\ln\left(\frac{C_\ell^{th}}{C_\ell^{data}}\right) + \left(\frac{C_\ell^{data}}{C_\ell^{th}}\right) - 1\right]$$

$$C_\ell^{data} = \sum_m |a_{\ell m}|^2/(2\ell+1).$$

Exercise:
show that

With noise	$C_\ell^{th} \longrightarrow C_\ell^{th} + \mathcal{N}_\ell$
Partial sky	$\ln\mathcal{L} \overset{\sim}{\longrightarrow} f_{sky}\ln\mathcal{L}$

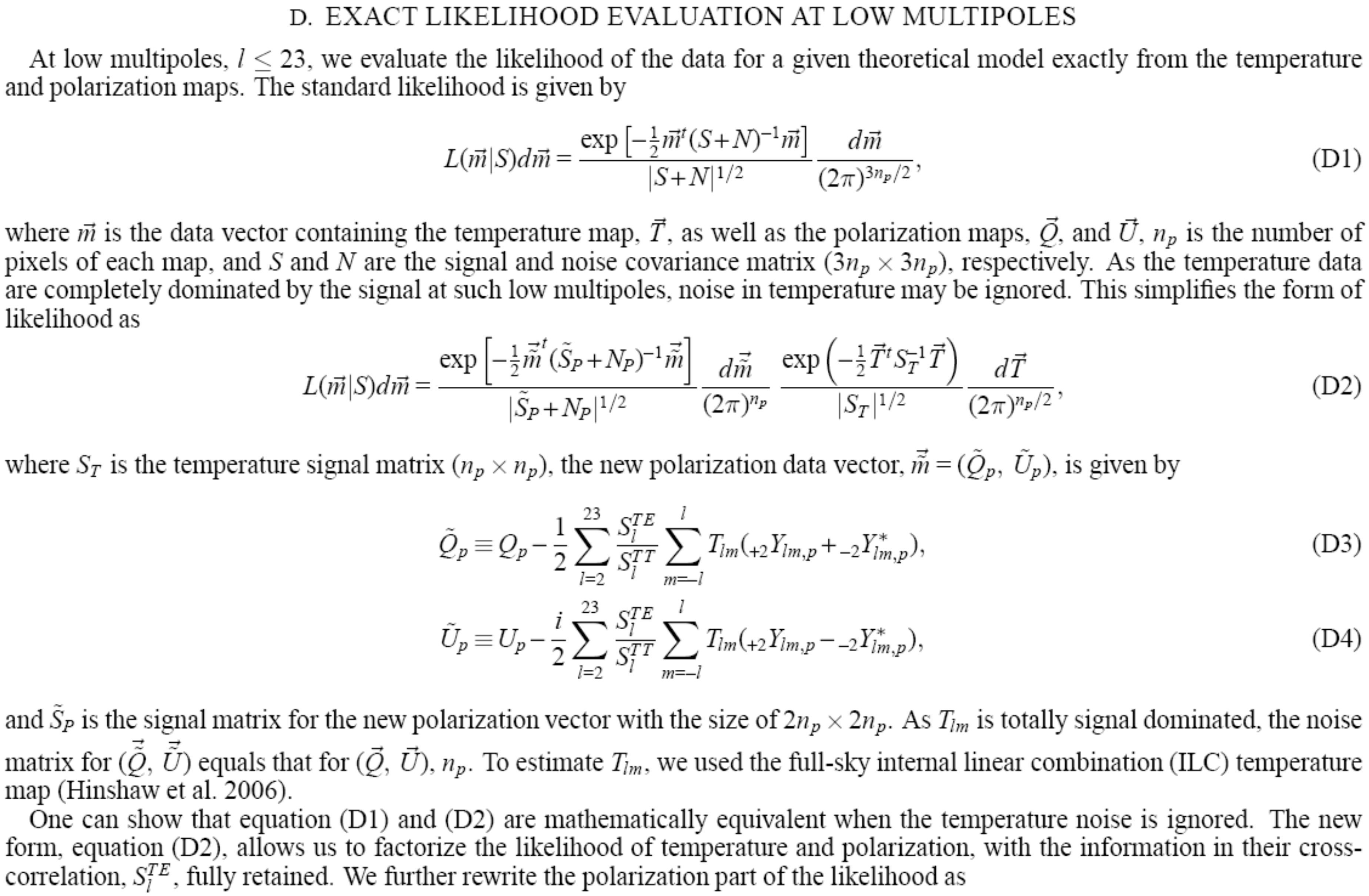
CMB light is polarized!

You can easily generalize the above to

$$
\begin{aligned}
-2\ln\mathcal{L} = & \sum_\ell (2\ell+1)\left\{\ln\left(\frac{C_\ell^{BB}}{\hat{C}_\ell^{BB}}\right) + \ln\left(\frac{C_\ell^{TT}C_\ell^{EE}-(C_\ell^{TE})^2}{\hat{C}_\ell^{TT}\hat{C}_\ell^{EE}-(\hat{C}_\ell^{TE})^2}\right)\right. \\
& \left. + \frac{\hat{C}_\ell^{TT}C_\ell^{EE}+C_\ell^{TT}\hat{C}_\ell^{EE}-2\hat{C}_\ell^{TE}C_\ell^{TE}}{C_\ell^{TT}C_\ell^{EE}-(C_\ell^{TE})^2} + \frac{\hat{C}_\ell^{BB}}{C_\ell^{BB}} - 3\right\},
\end{aligned}
$$

C_ℓ denotes C_ℓ^{th} and $\hat{C}_\ell$ denotes C_ℓ^{data}.

Exact TE,EE,BB Likelihood



D. EXACT LIKELIHOOD EVALUATION AT LOW MULTIPOLES

At low multipoles, $l \leq 23$, we evaluate the likelihood of the data for a given theoretical model exactly from the temperature and polarization maps. The standard likelihood is given by

$$L(\vec{m}|S)d\vec{m} = \frac{\exp\left[-\frac{1}{2}\vec{m}^t(S+N)^{-1}\vec{m}\right]}{|S+N|^{1/2}} \frac{d\vec{m}}{(2\pi)^{3n_p/2}}, \tag{D1}$$

where $\vec{m}$ is the data vector containing the temperature map, $\vec{T}$, as well as the polarization maps, $\vec{Q}$, and $\vec{U}$, n_p is the number of pixels of each map, and S and N are the signal and noise covariance matrix ($3n_p \times 3n_p$), respectively. As the temperature data are completely dominated by the signal at such low multipoles, noise in temperature may be ignored. This simplifies the form of likelihood as

$$L(\vec{m}|S)d\vec{m} = \frac{\exp\left[-\frac{1}{2}\vec{\tilde{m}}^t(\tilde{S}_P+N_P)^{-1}\vec{\tilde{m}}\right]}{|\tilde{S}_P+N_P|^{1/2}} \frac{d\vec{\tilde{m}}}{(2\pi)^{n_p}} \frac{\exp\left(-\frac{1}{2}\vec{T}^t S_T^{-1}\vec{T}\right)}{|S_T|^{1/2}} \frac{d\vec{T}}{(2\pi)^{n_p/2}}, \tag{D2}$$

where S_T is the temperature signal matrix ($n_p \times n_p$), the new polarization data vector, $\vec{\tilde{m}} = (\tilde{Q}_p,\ \tilde{U}_p)$, is given by

$$\tilde{Q}_p \equiv Q_p - \frac{1}{2}\sum_{l=2}^{23}\frac{S_l^{TE}}{S_l^{TT}}\sum_{m=-l}^{l} T_{lm}({}_{+2}Y_{lm,p} + {}_{-2}Y^*_{lm,p}), \tag{D3}$$

$$\tilde{U}_p \equiv U_p - \frac{i}{2}\sum_{l=2}^{23}\frac{S_l^{TE}}{S_l^{TT}}\sum_{m=-l}^{l} T_{lm}({}_{+2}Y_{lm,p} - {}_{-2}Y^*_{lm,p}), \tag{D4}$$

and $\tilde{S}_P$ is the signal matrix for the new polarization vector with the size of $2n_p \times 2n_p$. As T_{lm} is totally signal dominated, the noise matrix for $(\vec{\tilde{Q}},\ \vec{\tilde{U}})$ equals that for $(\vec{Q},\ \vec{U})$, n_p. To estimate T_{lm}, we used the full-sky internal linear combination (ILC) temperature map (Hinshaw et al. 2006).

One can show that equation (D1) and (D2) are mathematically equivalent when the temperature noise is ignored. The new form, equation (D2), allows us to factorize the likelihood of temperature and polarization, with the information in their cross-correlation, S_l^{TE}, fully retained. We further rewrite the polarization part of the likelihood as

Approximations

$$-2\ln\mathcal{L} = \sum_{\ell}(2\ell+1)\left[\ln\left(\frac{\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell}{\tilde{\mathcal{C}}_\ell}\right)+\frac{\tilde{\mathcal{C}}_\ell}{\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell}-1\right]$$

$$\ln\mathcal{L}_{\rm Gauss} \propto -\frac{1}{2}\sum_{\ell\ell'}(\mathcal{C}_\ell^{\rm th}-\widehat{\mathcal{C}}_\ell)Q_{\ell\ell'}(\mathcal{C}_{\ell'}^{\rm th}-\widehat{\mathcal{C}}_{\ell'})$$

$$-2\ln\mathcal{L}_{\rm LN} = \sum_{\ell\ell'}(z_\ell^{\rm th}-\widehat{z}_\ell)\mathcal{Q}_{\ell\ell'}(z_{\ell'}^{\rm th}-\widehat{z}_{\ell'}) \qquad \left\{\begin{array}{l} z_\ell^{\rm th} = \ln(\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell),\ \widehat{z}_\ell = \ln(\widehat{\mathcal{C}}_\ell+\mathcal{N}_\ell) \\ \mathcal{Q}_{\ell\ell'} = (\widehat{\mathcal{C}}_\ell+\mathcal{N}_\ell)Q_{\ell\ell'}(\widehat{\mathcal{C}}_{\ell'}+\mathcal{N}_{\ell'}). \end{array}\right.$$

Expand around the max $;\ \widehat{\mathcal{C}}_\ell = \mathcal{C}_\ell^{\rm th}(1+\epsilon).$

$$-2\ln\mathcal{L}_\ell = (2\ell+1)[\epsilon-\ln(1+\epsilon)] \simeq (2\ell+1)\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}+\mathcal{O}(\epsilon^4)\right)$$

Oh, look! $\ln\mathcal{L} = \frac{1}{3}\ln\mathcal{L}_{\rm Gauss}+\frac{2}{3}\ln\mathcal{L}'_{\rm LN}$ $\qquad \mathcal{Q}_{\ell\ell'} = (\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell)Q_{\ell\ell'}(\mathcal{C}_{\ell'}^{\rm th}+\mathcal{N}_{\ell'}).$

For Fisher

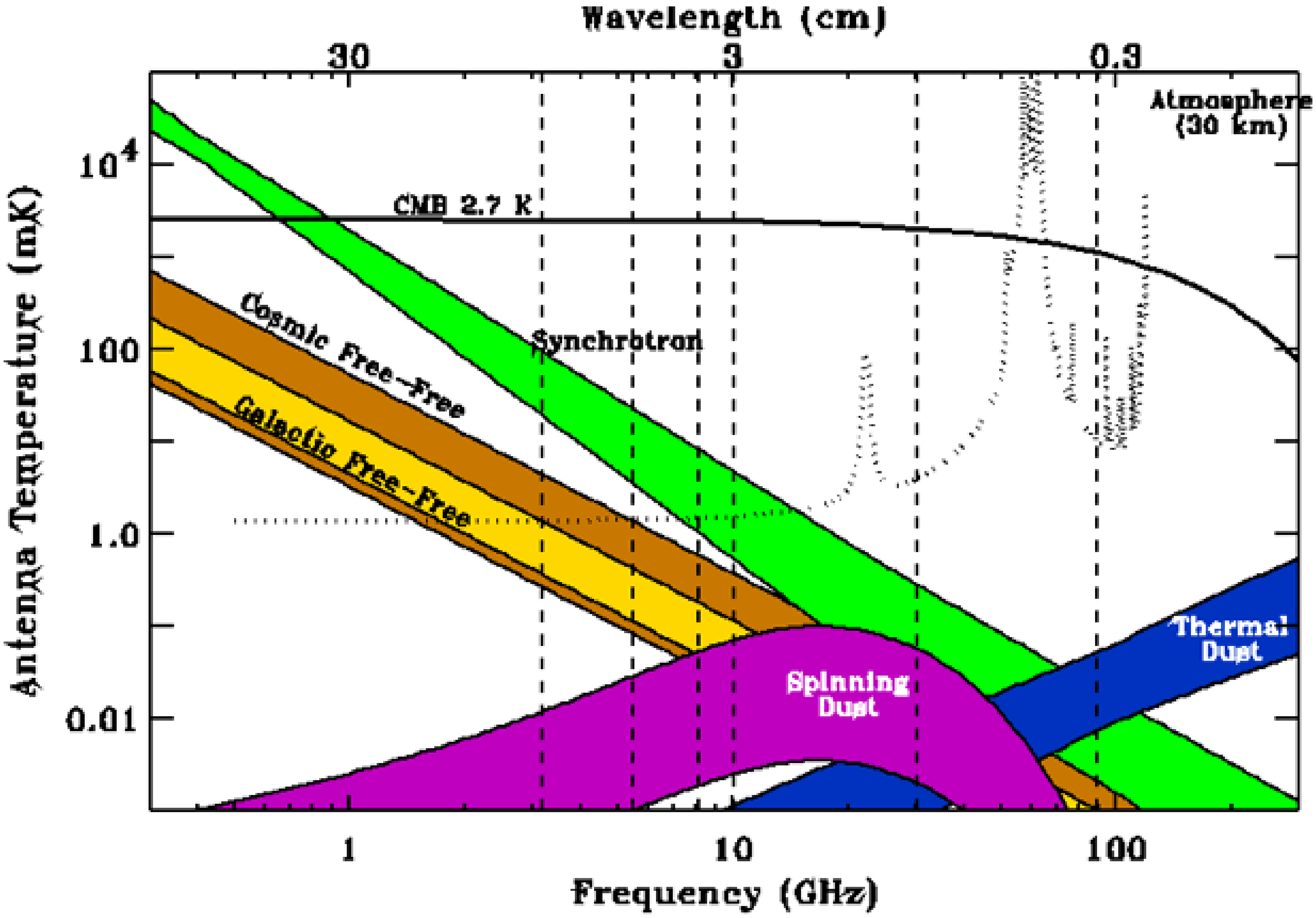
$$F_{ij}^{CMB} = \sum_{XY} \sum_{\ell} \frac{\partial C_\ell^X}{\partial \theta_i} \left(\mathscr{C}_\ell^{XY}\right)^{-1} \frac{\partial C_\ell^Y}{\partial \theta_j}$$

X,Y=TT, TE, EE, BB etc.,

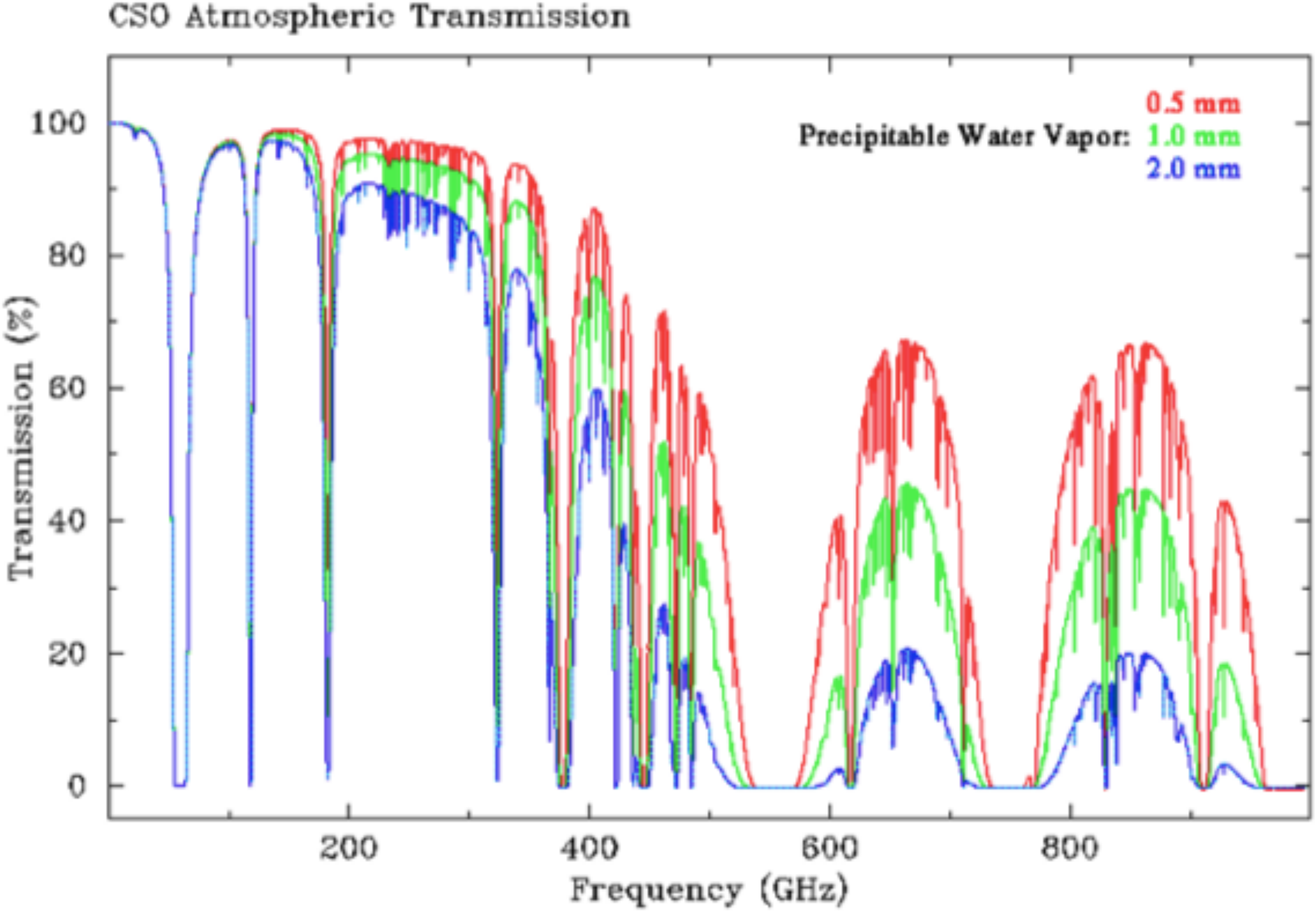
$$\mathscr{C}_\ell = \frac{2}{2\ell+1} \begin{pmatrix} (C_\ell^{TT})^2 & (C_\ell^{TE})^2 & C_\ell^{TT} C_\ell^{TE} & 0 \\ (C_\ell^{TE})^2 & (C_\ell^{EE})^2 & C_\ell^{EE} C_\ell^{TE} & 0 \\ C_\ell^{TT} C_\ell^{TE} & C_\ell^{EE} C_\ell^{TE} & 1/2[(C_\ell^{TE})^2 + C_\ell^{TT} C_\ell^{EE}] & 0 \\ 0 & 0 & 0 & (C_\ell^{BB})^2 \end{pmatrix}$$

Proof by intimidation: “just do it”

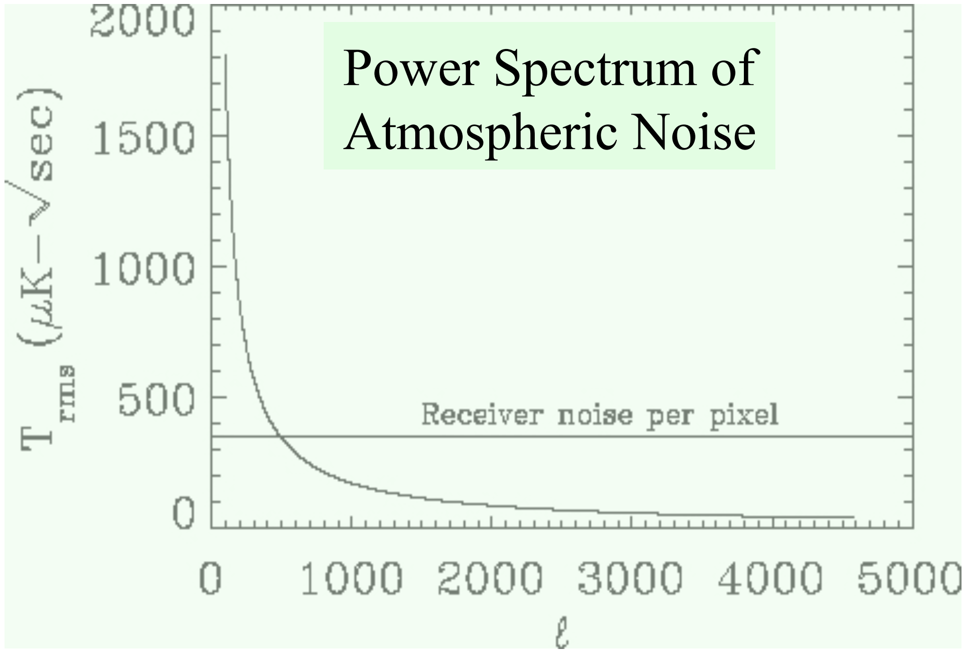
Back to the basics: How is the information extracted for the sky?



Water vapor and atmospheric transmission



ATMOSPHERE



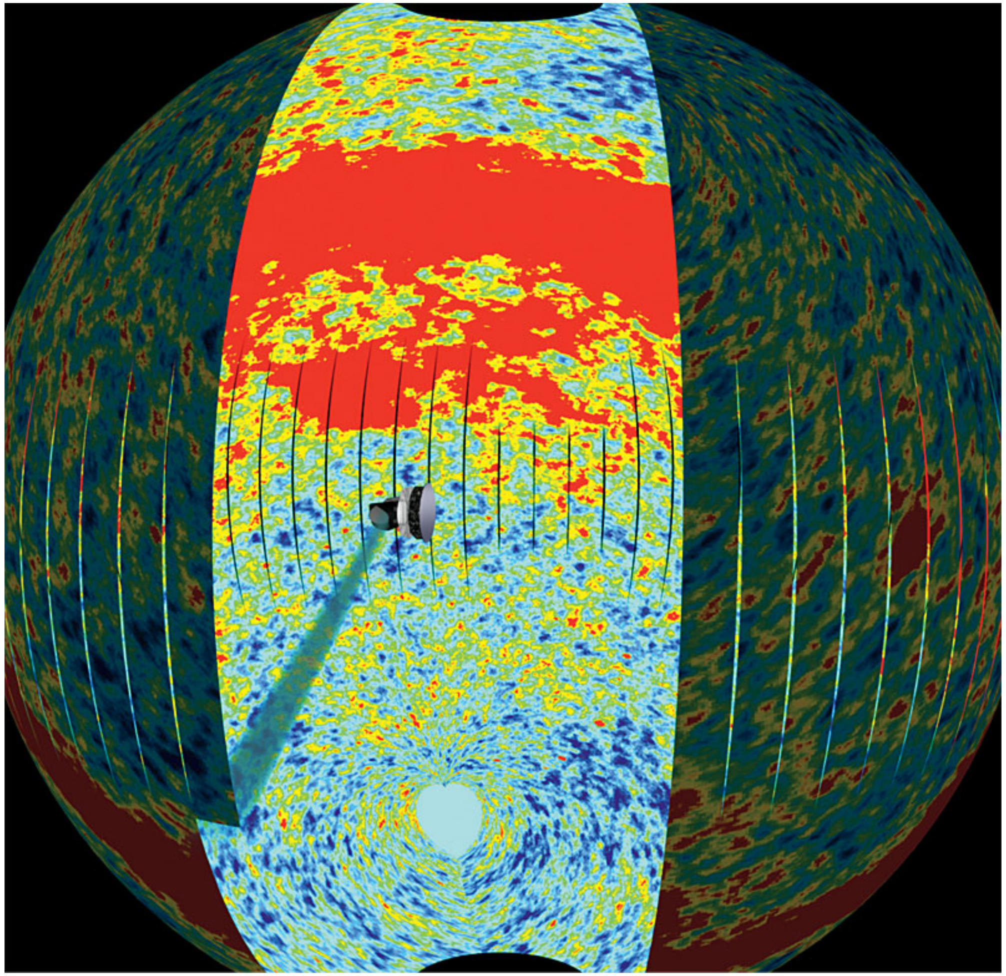
(based on Lay & Halverson)

The atmosphere is essentially featureless for $l > 1000$.

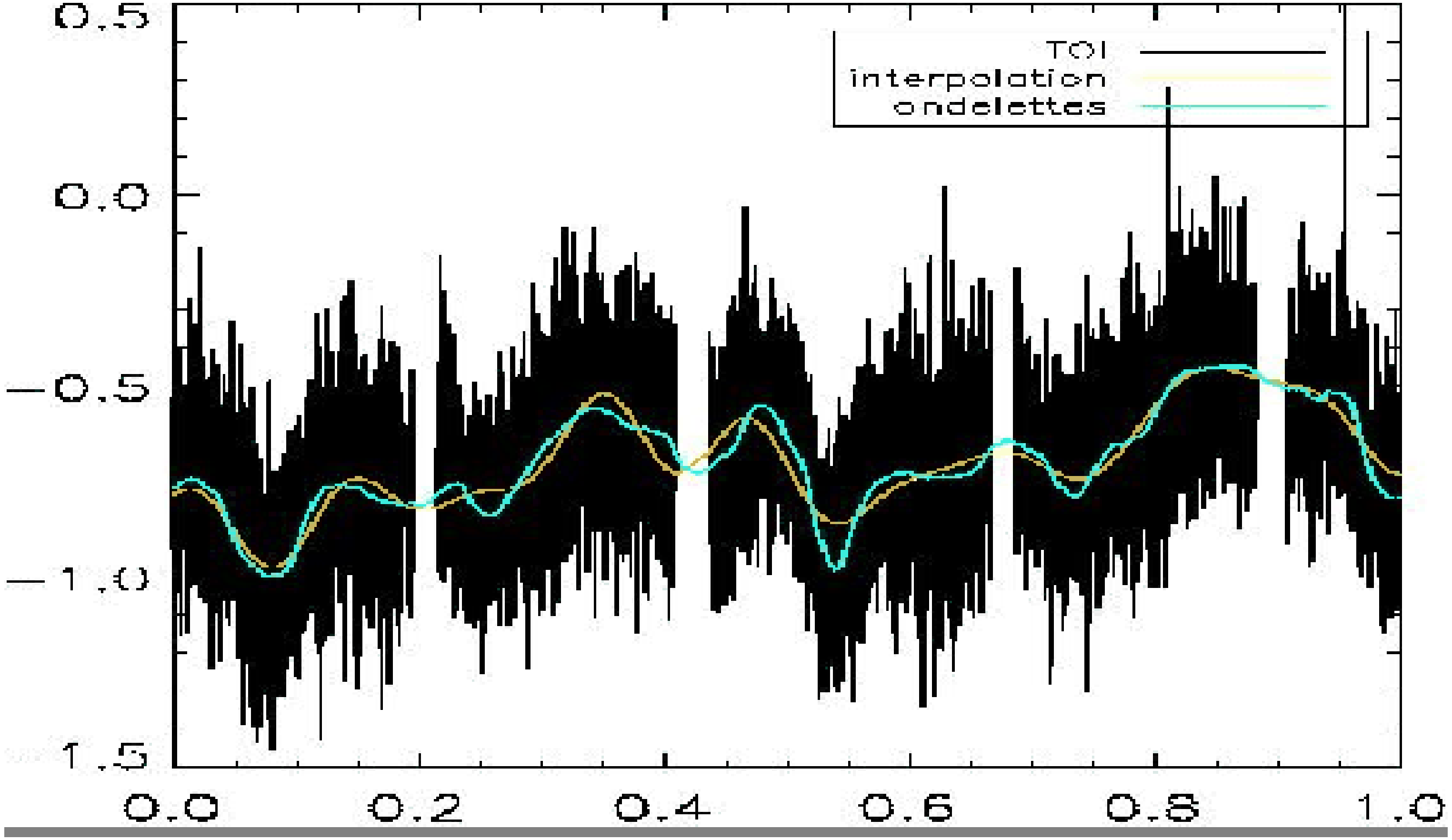
For *l* < 1000 solve for atmosphere with swept, over-sampled, filled array.

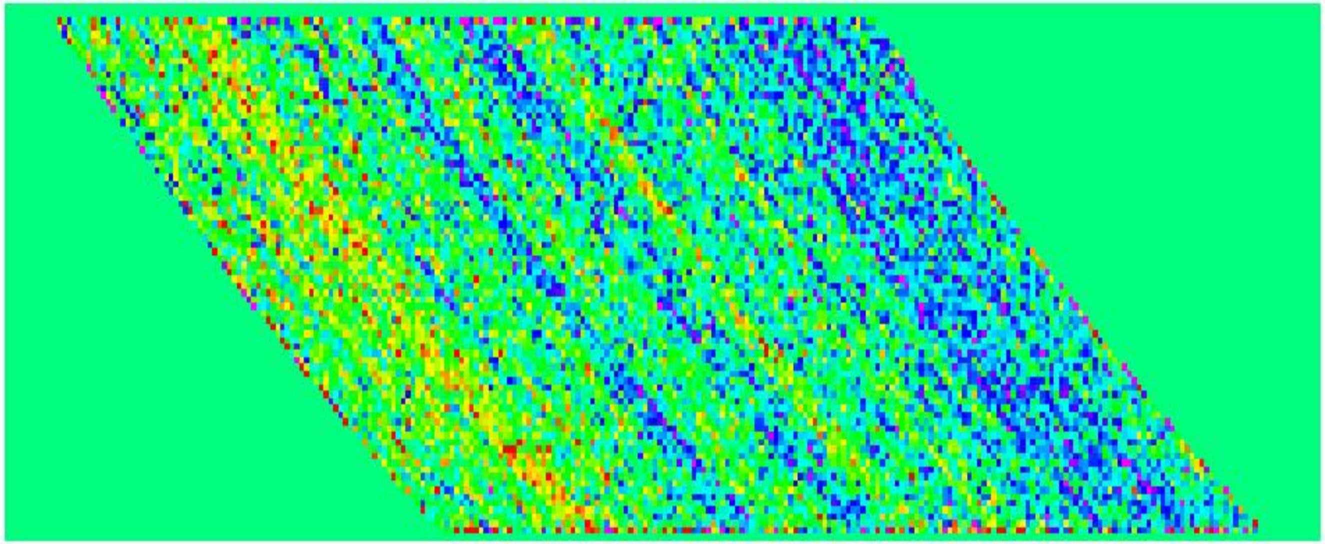
..but… How do you make a map in the first place?

The beam scans the sky with time, following a “scanning strategy”

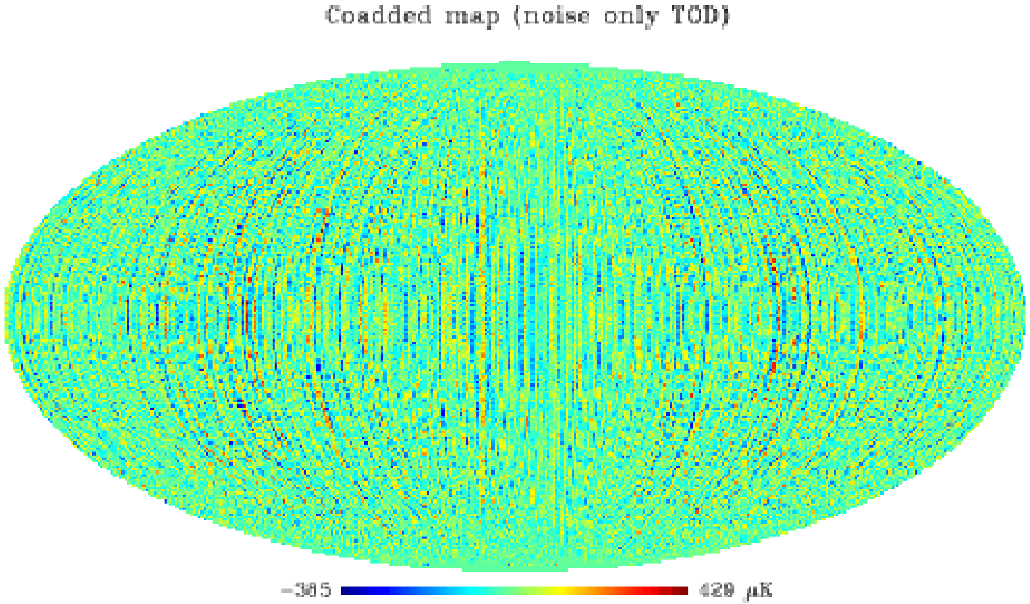


Time Ordered Data TOD Example from Archeops





Simulated ACT map



Simulated Planck
(noise only)

MAPMAKING

The problem can be recast in terms of operation of matrices on vectors.

$$d = g[M(T + T_{fg})] + g_{det}n + c$$

d is the raw TOD vector, can have extra info associated not just T(t) Elements are separated by a fraction of a second, for an experiment of years.. Imagine the size of the vector!

g denotes the gain, which is expected to vary but more slowly than T(t). There are several contribution to it: detector (det), receiver etc…

c is the baseline vector that depends on the details of the instrument

T is the CMB map vector, here I have explicitely separated the foregrounds contribution

M is the pointing matrix

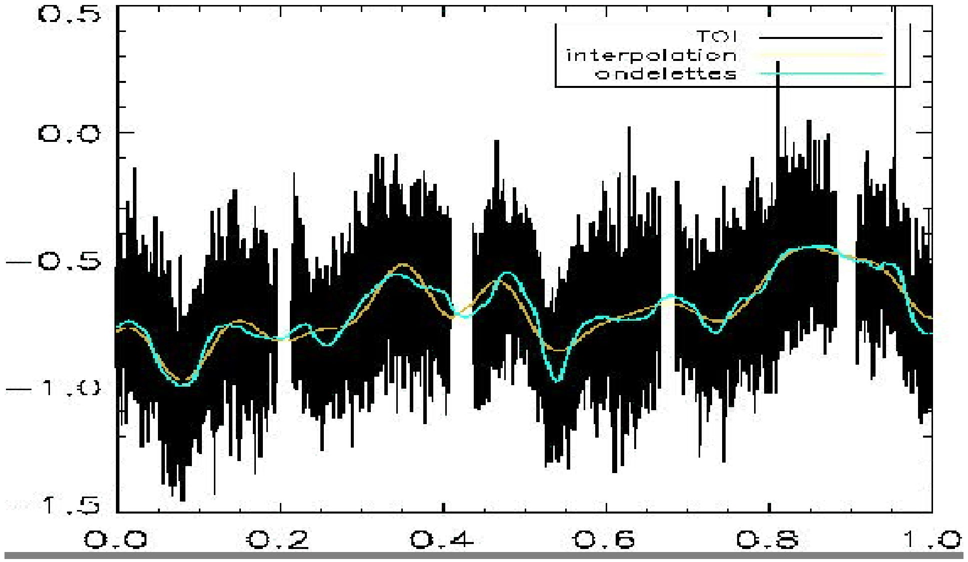
n is the noise, <n>=0,
but $<nn^T>$=N is not

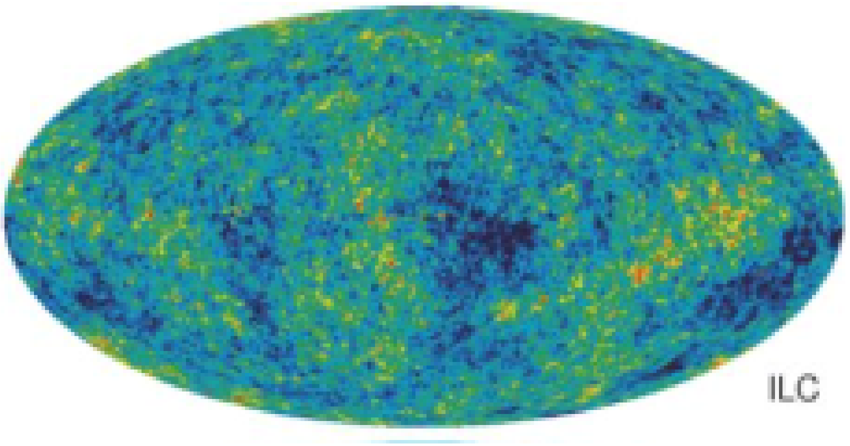
MAPMAKING

$$d = g[M(T + T_{fg})] + g_{det} n + c$$

Have this

Want this





It is good to have models for the various elements,
Use extra information (what varies slowly and what varies fast,
Frequency dependence etc.)

Note that if you had instead: d=MT+n
you could use a maximum likelihood estimator

$$\hat{T} = (M^T N^{-1} M)^{-1} (M^T N^{-1} d)$$

can be computed directly once *N* is characterized

Triky, ex. conjugate gradient with pre-conditioner

Simplified case:

Say: $T_0=M^TN^{-1}d$ then $T_0=(\underbrace{M^TN^{-1} M)}\hat{T}$

Call this Σ^{-1}

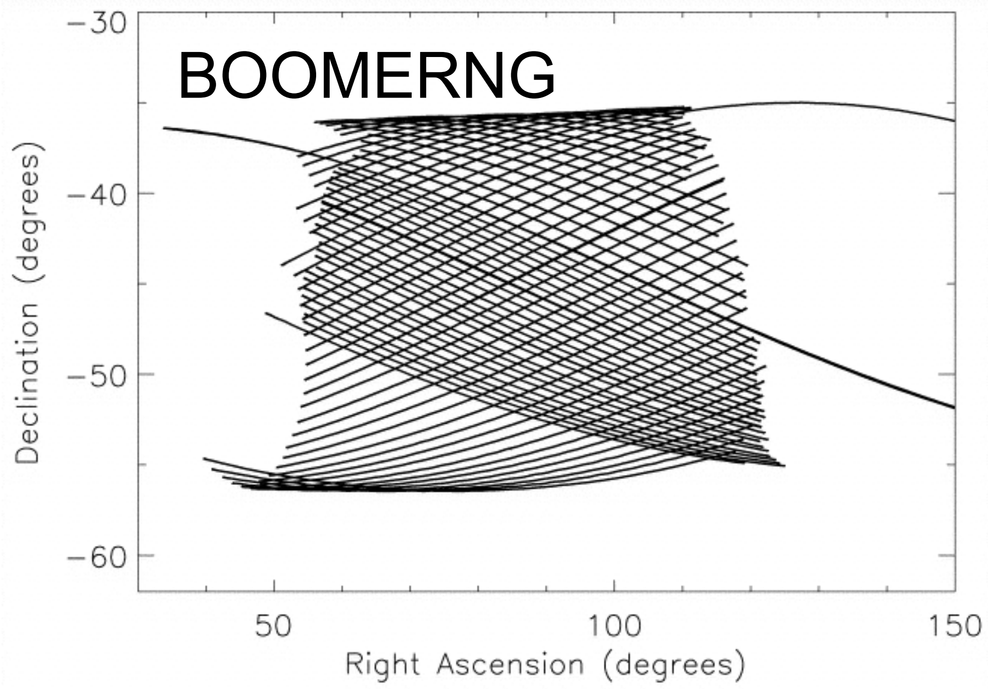
Σ is the pixel-to-pixel noise correlation matrix
Sove iteratively

To speed up the process use a preconditioner:
imagine exist a matrix S such that $S\Sigma^{-1}$ is diagonal

Then you solve for $S\Sigma^{-1}\hat{T} = ST_o$ which is easier/faster!

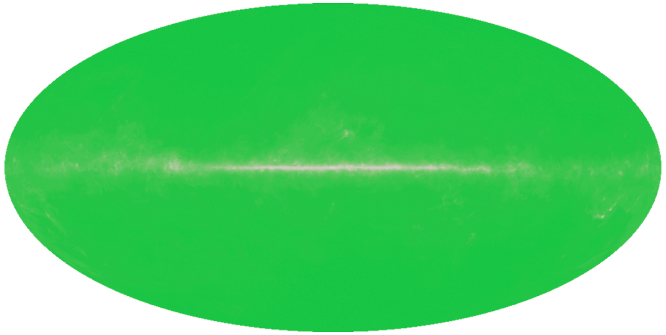
Unfortunately one does not have this simpler case:
Real map making becomes an iterative process

Take advantage of:
CROSS LINKING



<u>A word about calibration</u>:

The CMB fluctuations are 1 part in 10^5



The response (gain) of the detector needs to be much much better than that!

Calibration on CMB dipole or on point sources, planets etc….

AMAZING ACHIEVEMENT!

“If tortured sufficiently, data will confess to almost anything”

Fred Menger

Treat your data with respect

(Licia Verde)

Thank you!

references

http://arxiv.org/pdf/0712.3028

http://arxiv.org/pdf/0911.3105

http://arxiv.org/pdf/0906.0664

---

http://arxiv.org/pdf/astro-ph/9603021 (only sec 1 & 2)

http://arxiv.org/pdf/astro-ph/0703191