José Fonseca
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto
jose.fonseca@astro.up.pt

Fisher forecasts for the MeerKLASS survey

Outline

- Parameter estimation and Bayes Theorem;
- The Fisher matrix;
- Clustering estimators and the Fisher;
- Forecasts for MeerKAT's time proposals.

Prelude

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."

R. A. Fisher (1938)

Prelude

How do we predict the performance of an experiment, or how do we design (or optimise it) with some goals in mind?

Parameter estimation and Bayes Theorem

We have

$\mathbf{X}$ – Data

ϑ – parameters to fit

We want

Probability of the parameters given the data. With it we compute the expectation value and their errors.

$$\mathcal{P}(\vartheta|\mathbf{X}) - \text{posterior}$$

Parameter estimation and Bayes Theorem

We have

$\mathbf{X}$ – Data

ϑ – parameters to fit

We want

Probability of the parameters given the data. With it we compute the expectation value and their errors.

$\mathcal{P}(\vartheta|\mathbf{X})$ – posterior

Bayes Theorem

Conditional probability

$$p(\vartheta|\mathbf{X}) = \frac{\mathbf{p}(\vartheta, \mathbf{X})}{\mathbf{p}(\mathbf{X})} = \frac{\mathbf{p}(\mathbf{X}|\vartheta)\mathbf{p}(\vartheta)}{\mathbf{p}(\mathbf{X})}$$

Conditional probability

Parameter estimation and Bayes Theorem

We have

$$\mathbf{X} - \text{Data}$$

$$\vartheta - \text{parameters to fit}$$

We want

Probability of the parameters given the data. With it we compute the expectation value and their errors.

$$\mathcal{P}(\vartheta|\mathbf{X}) - \text{posterior}$$

Bayes Theorem

Conditional probability

$$p(\vartheta|\mathbf{X}) = \frac{\mathbf{p}(\vartheta, \mathbf{X})}{\mathbf{p}(\mathbf{X})} = \frac{\mathbf{p}(\mathbf{X}|\vartheta)\mathbf{p}(\vartheta)}{\mathbf{p}(\mathbf{X})}$$

Conditional probability

$$p(\mathbf{X}|\vartheta) \equiv \mathcal{L}(\mathbf{X}|\vartheta) - \text{Likelihood}$$

$$p(\vartheta) \equiv \mathbf{\Pi}(\vartheta) - \text{Prior}$$

$$p(\mathbf{X}) - \text{Evidence}$$

Parameter estimation and Bayes Theorem

Bayes Theorem

$$\text{Posterior probability} \propto \text{Likelihood} \times \text{Prior}$$

Average value

$$\langle \vartheta \rangle = \int \mathrm{d}\vartheta \; \mathcal{P}(\vartheta|\mathbf{X}) \; \vartheta$$

Parameter estimation and Bayes Theorem

Error estimation

- Let's assume a gaussian likelihood for the n-dimensional data vector **X** with covariance Sigma. For gaussian random errors it holds!

Basically the Chi-square

$$\mathcal{L}(\mathbf{X}|\vartheta) = \frac{1}{\left((2\pi)^n \det \Sigma\right)^{1/2}} \, e^{-\frac{1}{2}(\mathbf{X}-\bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X}-\bar{\mathbf{X}})}$$

- For a moment let us take the posterior of the m parameters theta to be gaussian. It doesn't have to be. **F** is then the covariance of the parameters;

$$\mathcal{P}(\vartheta|\mathbf{X}) = \frac{1}{\left((2\pi)^m \det(\mathbf{F}^{-1})\right)^{m/2}} \, e^{-\frac{1}{2}(\vartheta-\bar{\vartheta})^T \mathbf{F} (\vartheta-\bar{\vartheta})}$$

If the posterior is not a multivariate gaussian the expected value may not be the maximum of the posterior.

If the prior is flat (or if we don't care about it) the maximum likelihood estimator (or minimising the Chi-square) is the same as getting the maximum posterior

Parameter estimation and Bayes Theorem

Error estimation

If we Taylor expand the posterior at the maximum

$$\nabla_\vartheta \; \mathcal{P}(\vartheta|\mathbf{X})\Big|_{\vartheta=\bar{\vartheta}} = 0$$

Parameter estimation and Bayes Theorem

Error estimation

If we Taylor expand the posterior at the maximum $\nabla_\vartheta \; \mathcal{P}(\vartheta|\mathbf{X})\Big|_{\vartheta=\bar{\vartheta}} = 0$

$$\ln \mathcal{P}(\vartheta|\mathbf{X}) = \ln \mathcal{P}(\bar{\vartheta}|\mathbf{X}) + \frac{1}{2}\left(\vartheta_\alpha - \bar{\vartheta}_\alpha\right) \frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta} \left(\vartheta_\beta - \bar{\vartheta}_\beta\right)\Bigg|_{\vartheta=\bar{\vartheta}} + \ldots$$

Parameter estimation and Bayes Theorem

Error estimation

If we Taylor expand the posterior at the maximum

$$\nabla_\vartheta \; \mathcal{P}(\vartheta|\mathbf{X})\Big|_{\vartheta=\bar{\vartheta}} = 0$$

$$\ln \mathcal{P}(\vartheta|\mathbf{X}) = \ln \mathcal{P}(\bar{\vartheta}|\mathbf{X}) + \frac{1}{2}\left(\vartheta_\alpha - \bar{\vartheta}_\alpha\right) \frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta} \left(\vartheta_\beta - \bar{\vartheta}_\beta\right)\Bigg|_{\vartheta=\bar{\vartheta}} + \dots$$

The errors are given by the concavity of the posterior around the maximum

$$\mathbf{F}_{\vartheta_\alpha\vartheta_\beta} \equiv -\frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta}\Bigg|_{\vartheta=\bar{\vartheta}}$$

Parameter estimation and Bayes Theorem

Error estimation

If we Taylor expand the posterior at the maximum $\nabla_\vartheta\ \mathcal{P}(\vartheta|\mathbf{X})\Big|_{\vartheta=\bar{\vartheta}} = 0$

$$\ln \mathcal{P}(\vartheta|\mathbf{X}) = \ln \mathcal{P}(\bar{\vartheta}|\mathbf{X}) + \frac{1}{2}\left(\vartheta_\alpha - \bar{\vartheta}_\alpha\right) \frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta} \left(\vartheta_\beta - \bar{\vartheta}_\beta\right)\Big|_{\vartheta=\bar{\vartheta}} + \dots$$

The errors are given by the concavity of the posterior around the maximum

$$\mathbf{F}_{\vartheta_\alpha\vartheta_\beta} \equiv - \frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta}\Big|_{\vartheta=\bar{\vartheta}}$$

Marginal Error $\sigma_\vartheta = \sqrt{(\mathbf{F}^{-1})_{\vartheta\vartheta}}$

Conditional Error $\sigma_\vartheta^{\text{conditional}} = 1/\sqrt{\mathbf{F}_{\vartheta\vartheta}}$

The Fisher Matrix

How do we estimate future errors?

Estimate the Fisher matrix with fiducial data

I will abuse of notation and use the same notation as before

Fiducial Likelihood

Prior information

$$\mathbf{F}_{\vartheta_\alpha \vartheta_\beta} \equiv -\left.\frac{\partial^2 \ln \mathcal{P}(\vartheta|\mathbf{X})}{\partial\vartheta_\alpha \partial\vartheta_\beta}\right|_{\vartheta=\bar{\vartheta}} = -\left.\frac{\partial^2 \ln \mathcal{L}(\mathbf{X}|\vartheta)}{\partial\vartheta_\alpha \partial\vartheta_\beta}\right|_{\vartheta=\bar{\vartheta}} - \left.\frac{\partial^2 \ln \Pi(\vartheta)}{\partial\vartheta_\alpha \partial\vartheta_\beta}\right|_{\vartheta=\bar{\vartheta}}$$

Let's neglect it for now

For a Maximum Likelihood estimator the theorems are true

- For any unbiased estimator the Cramér-Rao inequality holds; $$\Delta\vartheta = \left(\langle\vartheta^2\rangle - \langle\vartheta\rangle^2\right)^{1/2} \geq 1/\sqrt{\mathbf{F}_{\vartheta\vartheta}}$$
- If an unbiased estimator attaining ("saturating") the Cramér-Rao bound exists, it is the ML estimator (or a function thereof);
- The ML-estimator is asymptotically a Best Unbiased Estimator.

The Fisher Matrix

The gaussian case

$$\mathcal{L}(\mathbf{X}|\vartheta) = \frac{1}{\left((2\pi)^n \det \Sigma\right)^{1/2}} \, e^{-\frac{1}{2}(\mathbf{X}-\bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X}-\bar{\mathbf{X}})}$$

Dropping the conditionals...

$$-2 \ln \mathcal{L} = \ln \det \Sigma + (\mathbf{X} - \bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X} - \bar{\mathbf{X}}) + n \ln 2\pi$$

The Fisher Matrix

The gaussian case

$$\mathcal{L}(\mathbf{X}|\vartheta) = \frac{1}{((2\pi)^n \det \Sigma)^{1/2}} \; e^{-\frac{1}{2}(\mathbf{X}-\bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X}-\bar{\mathbf{X}})}$$

Dropping the conditionals...

$$-2 \ln \mathcal{L} = \ln \det \Sigma + (\mathbf{X} - \bar{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X} - \bar{\mathbf{X}}) + n \ln 2\pi$$

$$\mathbf{F}_{\vartheta_i \vartheta_j} = - \left. \frac{\partial^2 \ln \mathcal{L}(\mathbf{X}|\vartheta)}{\partial \vartheta_i \partial \vartheta_j} \right|_{\vartheta_i = \bar{\vartheta}_i, \vartheta_j = \bar{\vartheta}_j} \longrightarrow \boxed{\mathbf{F}_{\vartheta_i \vartheta_j} = \frac{1}{2} \mathrm{Tr} \left[\Sigma^{-1} \Sigma_{,\vartheta_i} \Sigma^{-1} \Sigma_{,\vartheta_j} + \Sigma^{-1} M_{\vartheta_i \vartheta_j} \right]}$$

$$M_{\vartheta_i \vartheta_j} = \bar{\mathbf{X}}^T_{,\vartheta_j} \bar{\mathbf{X}}_{,\vartheta_i} + \bar{\mathbf{X}}^T_{,\vartheta_i} \bar{\mathbf{X}}_{,\vartheta_j}$$

Full derivation

Heavans (arXiv:0906.0664)

Clustering Estimators and the Fisher

Observable

$$\mathbf{F}_{\vartheta_i\vartheta_j} = \frac{1}{2}\mathrm{Tr}\left[\Sigma^{-1}\Sigma_{,\vartheta_i}\Sigma^{-1}\Sigma_{,\vartheta_j} + \Sigma^{-1}M_{\vartheta_i\vartheta_j}\right]$$

$$M_{\vartheta_i\vartheta_j} = \bar{\mathbf{X}}^T_{,\vartheta_j}\bar{\mathbf{X}}_{,\vartheta_i} + \bar{\mathbf{X}}^T_{,\vartheta_i}\bar{\mathbf{X}}_{,\vartheta_j}$$

$$a_{\ell m}$$

$$\bar{\mathbf{X}} = \langle a_{\ell m}\rangle = 0$$

$$\Sigma = \langle a_{\ell m}a^*_{\ell m}\rangle = C_\ell + \mathcal{N}_\ell$$

Number of m for each ell

$$\mathbf{F}_{\vartheta_i\vartheta_j} = \sum_{\ell_{\min}}^{\ell_{\max}} \frac{(2\ell+1)}{2} f_{\mathrm{sky}}\, \mathrm{Tr}\Big[\left(\partial_{\vartheta_i}\mathbf{C}_\ell\right)\boldsymbol{\Sigma}_\ell^{-1}\left(\partial_{\vartheta_j}\mathbf{C}_\ell\right)\boldsymbol{\Sigma}_\ell^{-1}\Big]$$

Incomplete m summation

Tegmark et al., ApJ (1997)

Clustering Estimators and the Fisher

Observable

$$\mathbf{F}_{\vartheta_i \vartheta_j} = \frac{1}{2} \mathrm{Tr}\left[\Sigma^{-1}\Sigma_{,\vartheta_i}\Sigma^{-1}\Sigma_{,\vartheta_j} + \Sigma^{-1}M_{\vartheta_i\vartheta_j}\right]$$

$$M_{\vartheta_i\vartheta_j} = \bar{\mathbf{X}}^T_{,\vartheta_j}\bar{\mathbf{X}}_{,\vartheta_i} + \bar{\mathbf{X}}^T_{,\vartheta_i}\bar{\mathbf{X}}_{,\vartheta_j}$$

Becomes a 4-point function which we neglect

C_ℓ

$$\bar{\mathbf{X}} = C_\ell$$

$$\begin{aligned}(\Sigma_\ell)_{ij,pq} &= Cov\left[\hat{C}^{\mathcal{M}}_{\ell,ij}, \hat{C}^{\mathcal{M}}_{\ell',pq}\right] \\ &= \frac{1}{(2\ell+1)\, f_{\mathrm{sky}}}\left(C^{\mathcal{G}}_{\ell,ip}C^{\mathcal{G}}_{\ell,jq} + C^{\mathcal{G}}_{\ell,iq}C^{\mathcal{G}}_{\ell,jp}\right)\end{aligned}$$

$$C^{\mathcal{G}}_{\ell,ij} \equiv C^{\mathcal{S}}_{\ell,ij} + C^{\mathcal{N}}_{\ell,ij}$$

$$\mathbf{F}_{\vartheta_\alpha\vartheta_\beta} = \sum_\ell \frac{\partial C^{\mathcal{M}}_{\ell,ij}}{\partial\vartheta_\alpha}\Sigma^{-1}_{\ell,ij,mn}\frac{\partial C^{\mathcal{M}}_{\ell,mn}}{\partial\vartheta_\beta}$$

Clustering Estimators and the Fisher

Observable

$$\mathbf{F}_{\vartheta_i\vartheta_j} = \frac{1}{2}\mathrm{Tr}\left[\Sigma^{-1}\Sigma_{,\vartheta_i}\Sigma^{-1}\Sigma_{,\vartheta_j} + \Sigma^{-1}M_{\vartheta_i\vartheta_j}\right]$$

$$M_{\vartheta_i\vartheta_j} = \bar{\mathbf{X}}^T_{,\vartheta_j}\bar{\mathbf{X}}_{,\vartheta_i} + \bar{\mathbf{X}}^T_{,\vartheta_i}\bar{\mathbf{X}}_{,\vartheta_j}$$

δ

$$\bar{\mathbf{X}} = \langle\delta\rangle = 0$$

$$\mathbf{F}_{\vartheta_i\vartheta_j} = \frac{1}{2}\frac{V_{\mathrm{survey}}}{4\pi^2}\int_{-1}^{1}\mathrm{d}\mu\int_{k_{\mathrm{min}}}^{k_{\mathrm{max}}}\mathrm{d}k\,\frac{\partial P(k,\mu)}{\partial\vartheta_i}\frac{\partial P(k,\mu)}{\partial\vartheta_j}\left(\frac{1}{P(k,\mu)+N}\right)^2$$

$$\Sigma = \langle\delta_k\delta^*_{k'}\rangle = (2\pi)^3\delta^3_D(\vec{k}-\vec{k}')(P(k,\mu)+N)$$

Clustering Estimators and the Fisher

Example: Cross Correlation Power spectrum

$$\mathbf{F}_{\vartheta_i \vartheta_j} = \frac{1}{2}\mathrm{Tr}\left[\Sigma^{-1}\frac{\partial \mathbf{P}}{\partial \vartheta_i}\Sigma^{-1}\frac{\partial \mathbf{P}}{\partial \vartheta_j}\right]$$

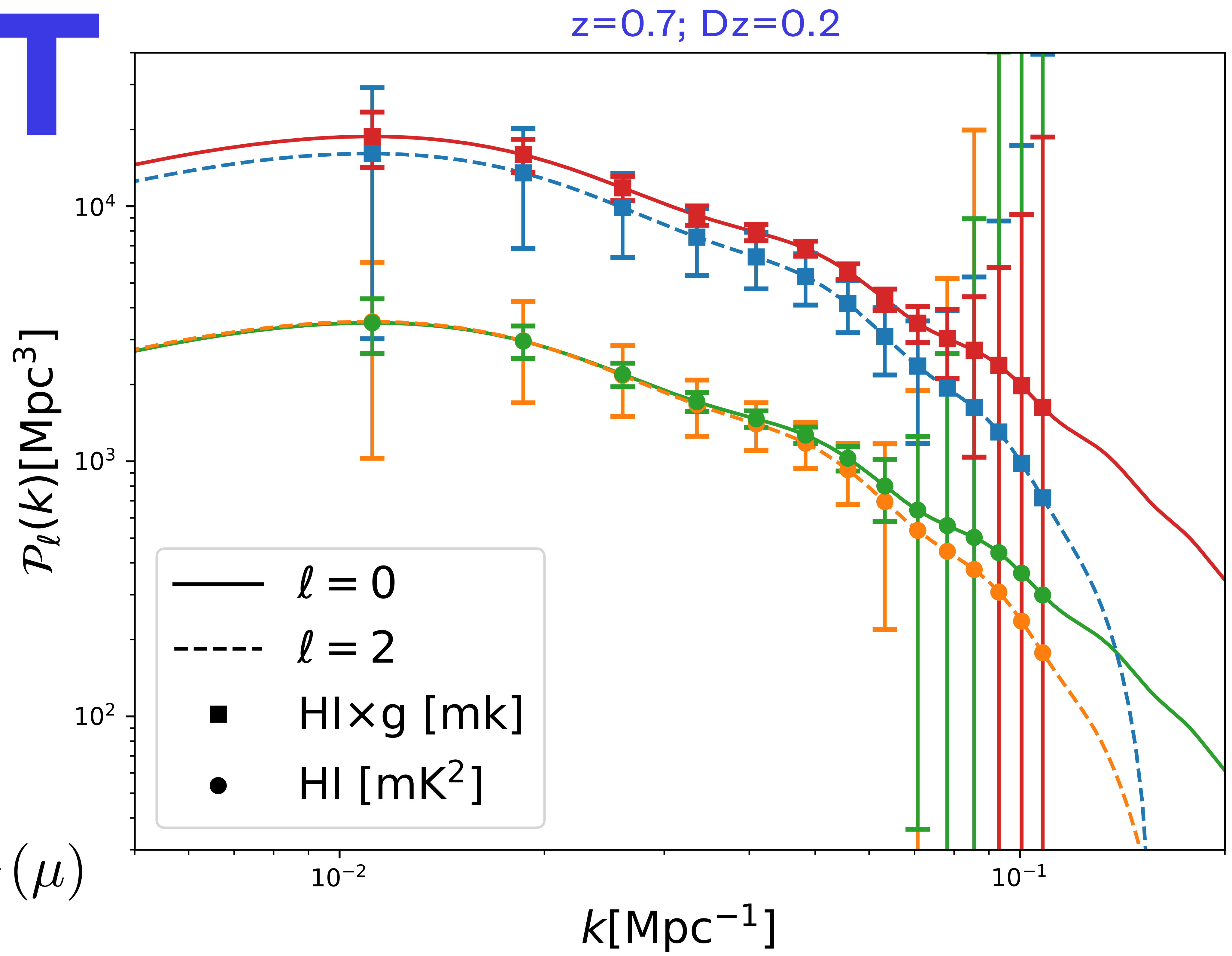
Let us simplify to one k mode

$$\vartheta = \{P_1, P_2, P_{12}\} \qquad \mathbf{P} = \begin{bmatrix} P_1 & P_{12} \\ P_{12} & P_2 \end{bmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} P_1 + N_1 & P_{12} \\ P_{12} & P_2 + N_2 \end{bmatrix}$$

$$\mathbf{F} = \frac{1}{2}\frac{N_k}{(P_1+N_1)(P_2+N_2) - P_{12}^2}\begin{bmatrix} (P_2+N_2)^2 & P_{12}^2 & -2P_{12}(P_2+N_2) \\ P_{12}^2 & (P_1+N_1)^2 & -2P_{12}(P_1+N_1) \\ -2P_{12}(P_2+N_2) & -2P_{12}(P_1+N_1) & 2((P_1+N_1)(P_2+N_2)+P_{12}^2) \end{bmatrix}$$

$$\mathbf{F}^{-1} = \frac{2}{N_k}\begin{bmatrix} (P_1+N_1)^2 & P_{12}^2 & P_{12}(P_1+N_1) \\ P_{12}^2 & (P_2+N_2)^2 & P_{12}(P_2+N_2) \\ P_{12}(P_1+N_1) & P_{12}(P_2+N_2) & ((P_1+N_1)(P_2+N_2)+P_{12}^2)/2 \end{bmatrix}$$

Forecasts for MeerKAT



Using multipoles of the power spectrum

$$\mathcal{P}_\ell(z,k) = \frac{2\ell+1}{2}\int_{-1}^{1} \mathrm{d}\mu\; P(z,k,\mu)\mathcal{L}_\ell(\mu)$$

$$\Sigma_\ell^{\mathrm{auto}}(z,k) = \frac{(2\ell+1)(2\ell'+2)}{2N_k}\int_{-1}^{1} \mathrm{d}\mu\;\left(P(z,k,\mu)+N(z)\right)^2 \mathcal{L}_\ell(\mu)\mathcal{L}_{\ell'}(\mu)$$

$$\Sigma_\ell^{\mathrm{X}}(z,k) = \frac{(2\ell+1)(2\ell'+2)}{4N_k}\int_{-1}^{1} \mathrm{d}\mu\Big[\left(P_1(z,k,\mu)+N_1(z)\right)\left(P_2(z,k,\mu)+N_2(z)\right)+P_X(z,k,\mu)\Big]\mathcal{L}_\ell(\mu)\mathcal{L}_{\ell'}(\mu)$$

Rubiola et al., (2021)

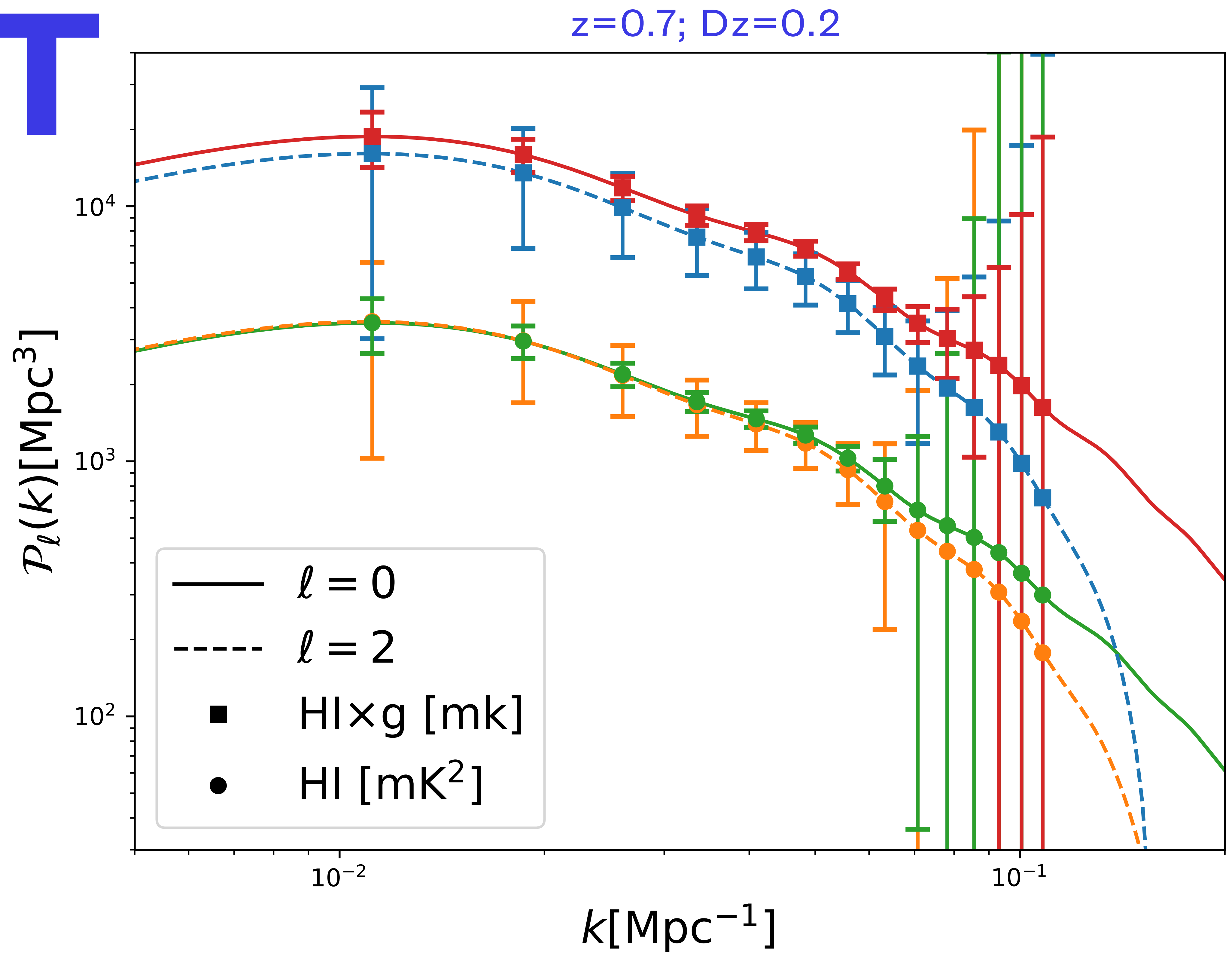
Forecasts for MeerKAT

Using multipoles of the power spectrum

Specs $$P_{\rm N} = \frac{S_{\rm A} T_{\rm sys}^2}{2 N_D t_{\rm tot}} \frac{c(1+z)^2 \chi^2(z)}{\nu_{\rm HI} H(z)} \times \frac{1}{W^2}$$

$$W(k,\mu,z) = e^{-k^2(1-\mu^2)\chi^2\theta_B^2/16\ln 2}$$

- For HI: bias and Temperature from Red Book, UHF system temperature with eta=0.72 (calibrated from L-band), 60 dishes, 60 hr and 500 deg2;
- For the galaxies a generic n=4e-5 Mpc^-3, bias sqrt(1+z);
- k_min given by the volume and k_max given by the beam;
- Exclude RFI dominated regions from 0.46-0.52.



$$SNR_\ell = \sqrt{\sum_k \left(\frac{\mathcal{P}_\ell(k)}{\Delta\mathcal{P}_\ell(k)}\right)^2}$$

PHI_0:	27.4
PHI_2:	9.8
PHIg_0:	32.2
PHIg_2:	9.8

Summary

- Summary of where the Fisher Matrix comes from;
- Forecasts for MeerKAT's time proposals.