Status of the QML estimator

Alessandro Gruppuso
(INAF-OAS BO)

with M.Migliaccio and P.Natoli

Euclid CMBXC meeting
Ferrara - 4 October 2018

Outline

- Intro to the QML algebra
- Description of the implementation
- Validation
- Binning option
- Uncertainties for TG from Fisher and their dependence on the fiducial spectra
- example at nside=64
- Summary and next steps

QML method

$$\hat{C}_\ell = F^{-1}_{\ell\ell'}\left[x^t E_{\ell'} x - Tr[N E_{\ell'}]\right]$$

Sum on ℓ' is understood

total covariance, in pixel space

$$C = S(\{C_\ell^{fid}\}) + N$$

$$E_\ell = \frac{1}{2} C^{-1} P_\ell C^{-1}$$

matrix E, for each multipole
pixel space

$$P_\ell = \frac{\partial C}{\partial C_\ell^{fid}}$$

matrix P in pixel space,
related to the Legendre polynomials

$$F_{\ell\ell'} = \frac{1}{2} Tr\left[C^{-1} P_\ell C^{-1} P_{\ell'}\right]$$

Fisher matrix, in harmonic space

x map, in pixel space

QML method

$y_{\ell'}$

vector, in harmonic space

$$\hat{C}_\ell = F^{-1}_{\ell\ell'} \left[x^t E_{\ell'} x - Tr[N E_{\ell'}] \right]$$

Sum on ℓ' is understood

total covariance, in pixel space

$$E_\ell = \frac{1}{2} C^{-1} P_\ell C^{-1}$$

matrix E, for each multipole
pixel space

$$C = S(\{C_\ell^{fid}\}) + N$$

$$P_\ell = \frac{\partial C}{\partial C_\ell^{fid}}$$

matrix P in pixel space,
related to the Legendre polynomials

$$F_{\ell\ell'} = \frac{1}{2} Tr \left[C^{-1} P_\ell C^{-1} P_{\ell'} \right]$$

Fisher matrix, in harmonic space

x map, in pixel space

QML method

$y_{\ell'}$

vector, in harmonic space

$$\hat{C}_\ell = F^{-1}_{\ell\ell'}\left[x^t E_{\ell'} x - Tr[N E_{\ell'}]\right]$$

Sum on ℓ' is understood

This estimator is **optimal** since it is "unbiased and minimum variance". They are minimum variance because they saturate the Fisher-Cramer-Rao inequality (under very general assumptions it is possible to show that the variance of a given estimator has a lower bound).

$$\langle \hat{C}_\ell \rangle = C_\ell$$

unbiased

$$\langle (\hat{C}_\ell - C_\ell)(\hat{C}_{\ell'} - C_{\ell'}) \rangle = F^{-1}_{\ell\ell'}$$

minimum variance

QML properties

$$\hat{C}_\ell = F^{-1}_{\ell\ell'} y_{\ell'}$$

- Is computationally heavy for situations of practical interest and limited by the number of pixels. The number of operations are roughly driven by the building of matrix E and the Fisher matrix. Its implementation needs to be parallel.

- it can be shown (*even algebraically*) that the fiducial spectrum used to build the E or Fisher matrix does not impact on the estimates. In other words, the method is unbiased as long as E and F are built starting from the same objects (same fiducial spectra and same noise covariance matrix). This makes the method very robust.

This feature might be used to make the computation lighter in principle (see later).

Implementation currently available

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$

QML algebra

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG_1} \\ \hat{C}_\ell^{G_1G_1} \\ \hat{C}_\ell^{G_2G_2} \\ \hat{C}_\ell^{TG_2} \\ \hat{C}_\ell^{G_1G_2} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG_1} \\ y_{\ell'}^{G_1G_1} \\ y_{\ell'}^{G_2G_2} \\ y_{\ell'}^{TG_2} \\ y_{\ell'}^{G_1G_2} \end{pmatrix}$$

explicitly written the elements of the vectors to make clear the dimensionality of the problem

Implementation currently available

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} TT,TT & TT,TG_1 & TT,G_1G_1 & TT,G_2G_2 & TT,TG_2 & TT,G_1G_2 \\ & TG_1,TG_1 & TG_1,G_1G_1 & TG_1,G_2G_2 & TG_1,TG_2 & TG_1,G_1G_2 \\ & & G_1G_1,G_1G_1 & G_1G_1,G_2G_2 & G_1G_1,TG_2 & G_1G_1,G_1G_2 \\ & & & G_2G_2,G_2G_2 & G_2G_2,TG_2 & G_2G_2,G_1G_2 \\ & \text{symmetric} & & & TG_2,TG_2 & TG_2,G_1G_2 \\ & & & & & G_1G_2,G_1G_2 \end{pmatrix}$$

where each entry of the matrix stands for

$$F^{XY}_{\ell\ell'} = \frac{1}{2} Tr \left[C^{-1} P^X_\ell C^{-1} P^Y_{\ell'} \right]$$

$$X, Y = TT, TG_1, G_1G_1, G_2G_2, TG_2, G_1G_2$$

Implementation currently available

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$

and the Fisher matrix is

To make you aware of the heaviness of the method

$$F_{\ell\ell'} = \begin{pmatrix} TT,TT & TT,TG_1 & TT,G_1G_1 & TT,G_2G_2 & TT,TG_2 & TT,G_1G_2 \\ & TG_1,TG_1 & TG_1,G_1G_1 & TG_1,G_2G_2 & TG_1,TG_2 & TG_1,G_1G_2 \\ & & G_1G_1,G_1G_1 & G_1G_1,G_2G_2 & G_1G_1,TG_2 & G_1G_1,G_1G_2 \\ & & & G_2G_2,G_2G_2 & G_2G_2,TG_2 & G_2G_2,G_1G_2 \\ & \text{symmetric} & & & TG_2,TG_2 & TG_2,G_1G_2 \\ & & & & & G_1G_2,G_1G_2 \end{pmatrix}$$

0.2 GB at nside 16 (our masked case)

43 GB at nside 64 (our masked case)

where each entry of the matrix stands for

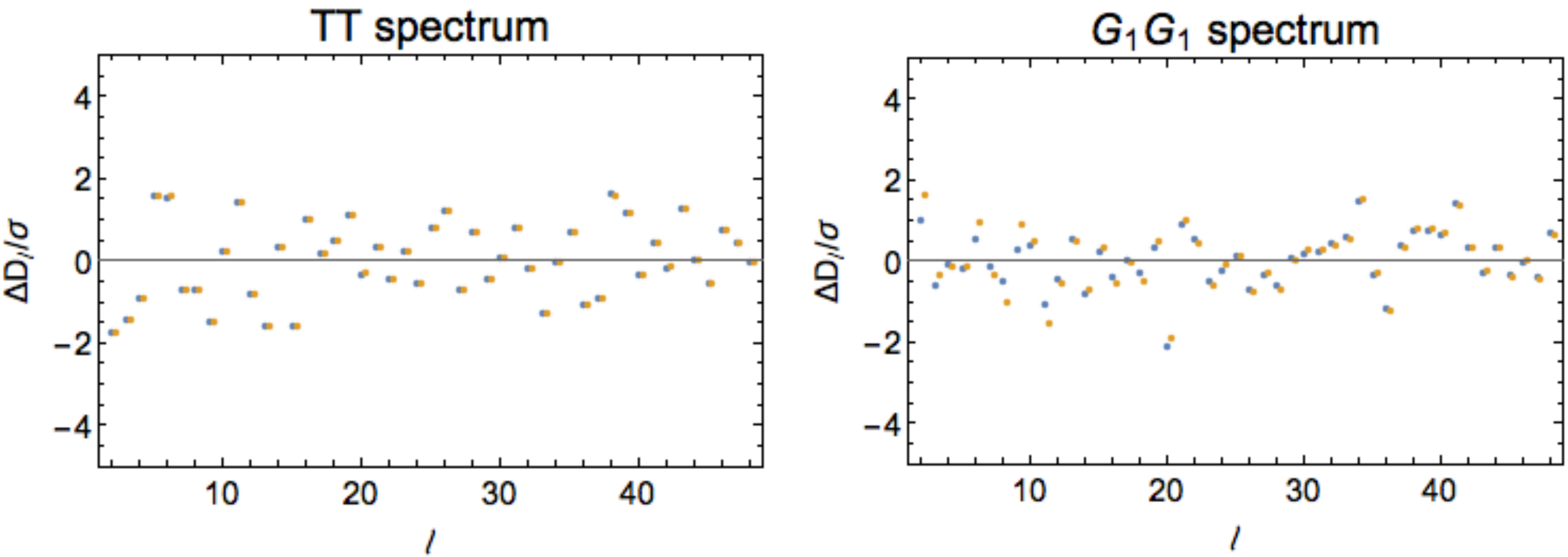
$$F^{XY}_{\ell\ell'} = \frac{1}{2} Tr \left[C^{-1} P^X_\ell C^{-1} P^Y_{\ell'} \right]$$

$$X, Y = TT, TG_1, G_1G_1, G_2G_2, TG_2, G_1G_2$$

Validation (nside=16)

Blue estimates = estimator built with the same fiducial used for the MC

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

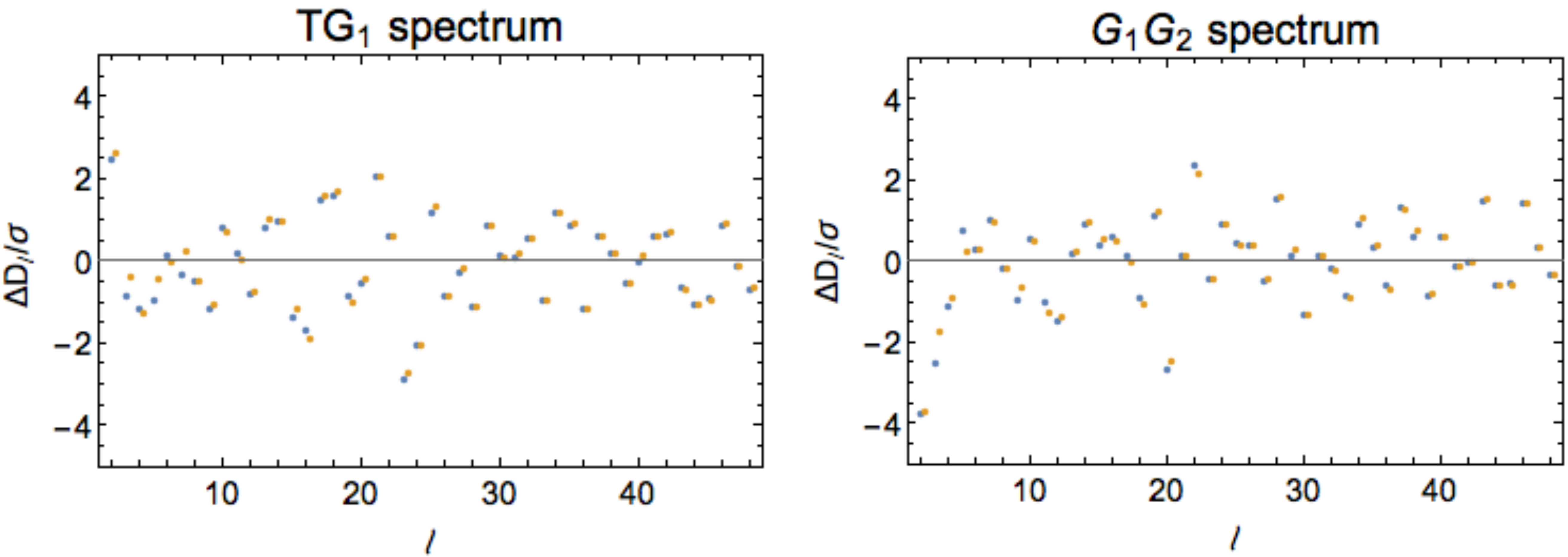


Distance of MC average (1000 sims) from the fiducial spectra in units of standard deviation of the mean

Validation (nside=16)

Blue estimates = estimator built with the same fiducial used for the MC

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

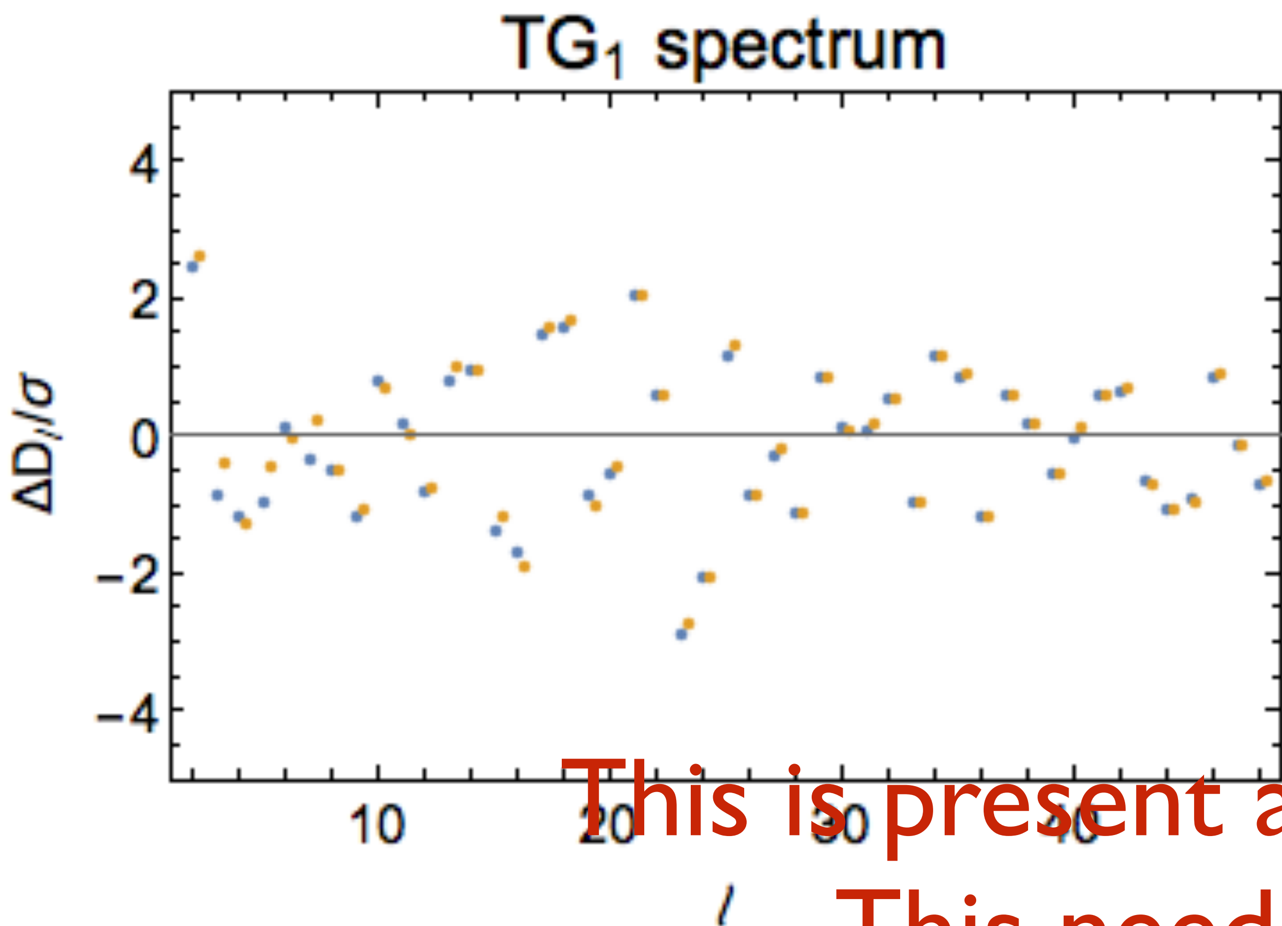


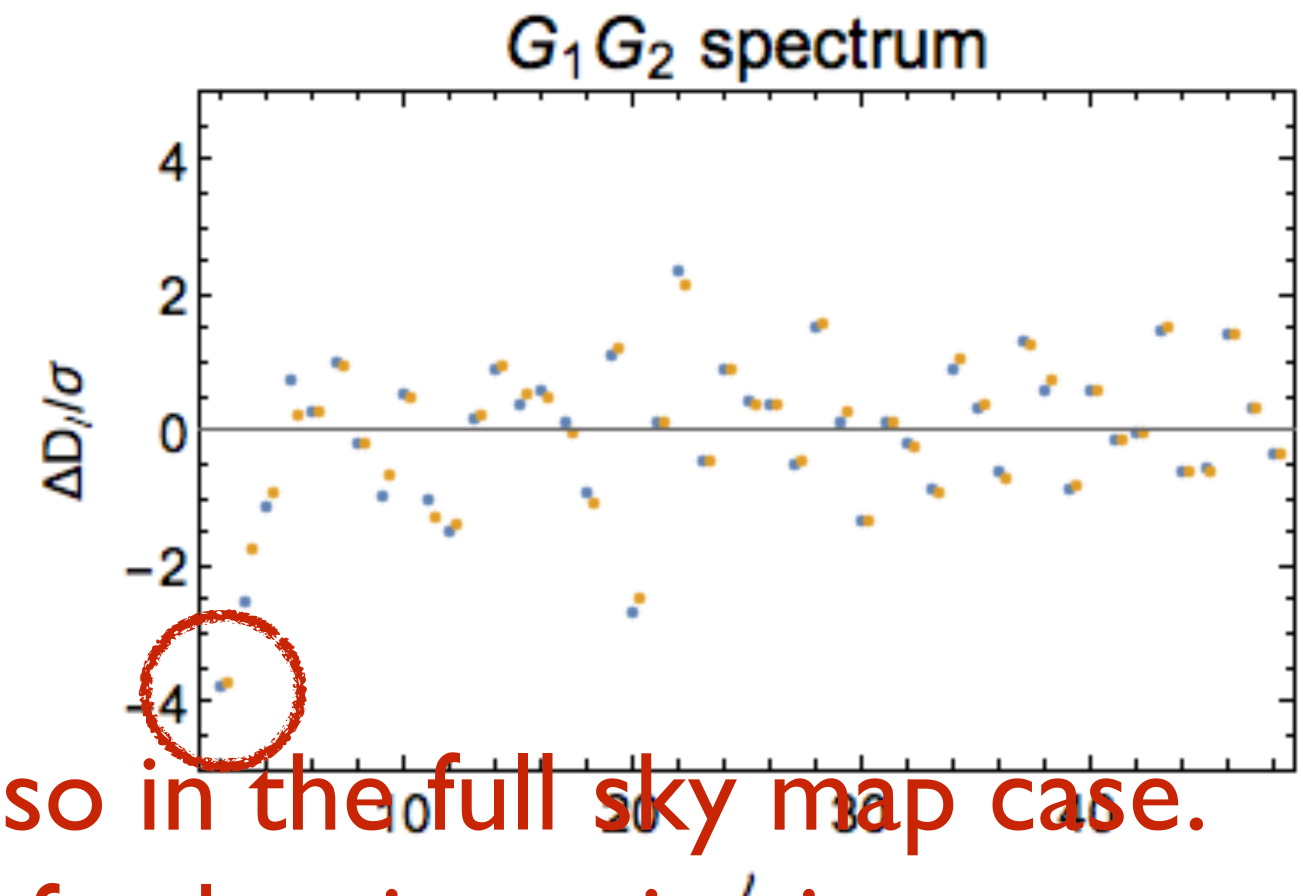
Distance of MC average (1000 sims) from the fiducial spectra in units of standard deviation of the mean

Validation (nside=16)

Blue estimates = estimator built with the same fiducial used for the MC

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)





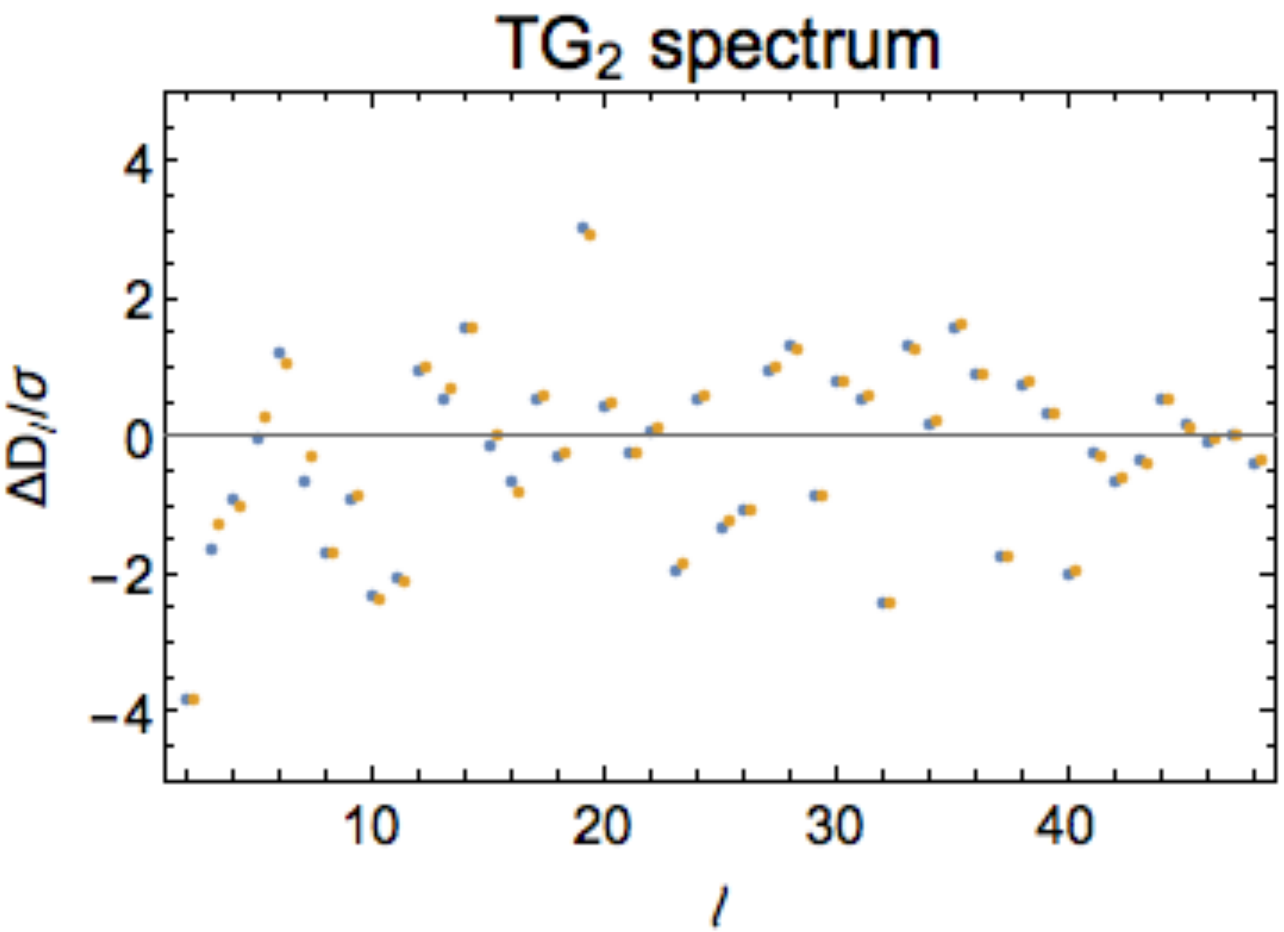
This is present also in the full sky map case.
This needs further investigation

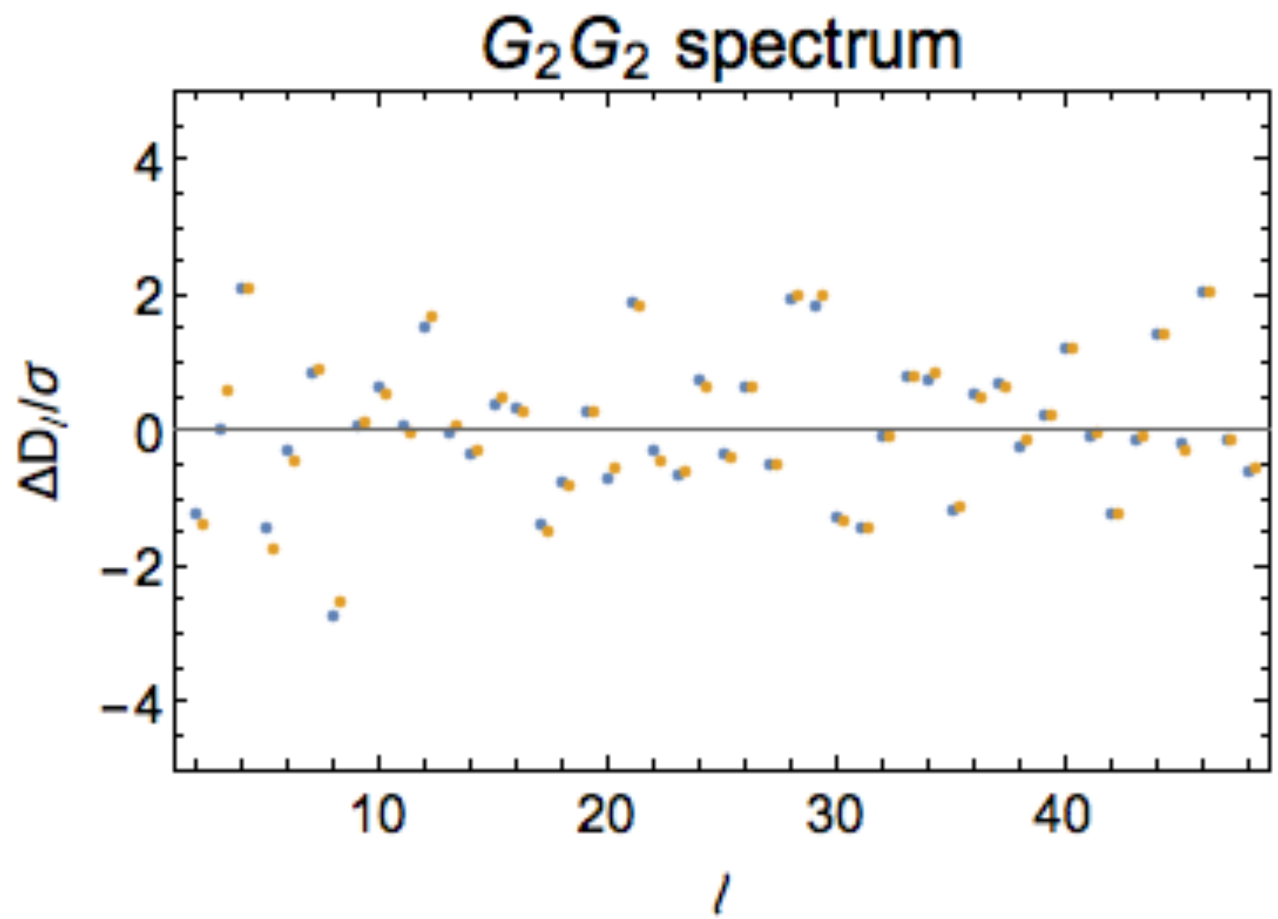
Distance of MC average (1000 sims) from the fiducial spectra in units of standard deviation of the mean

Validation (nside=16)

Blue estimates = estimator built with the same fiducial used for the MC

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)



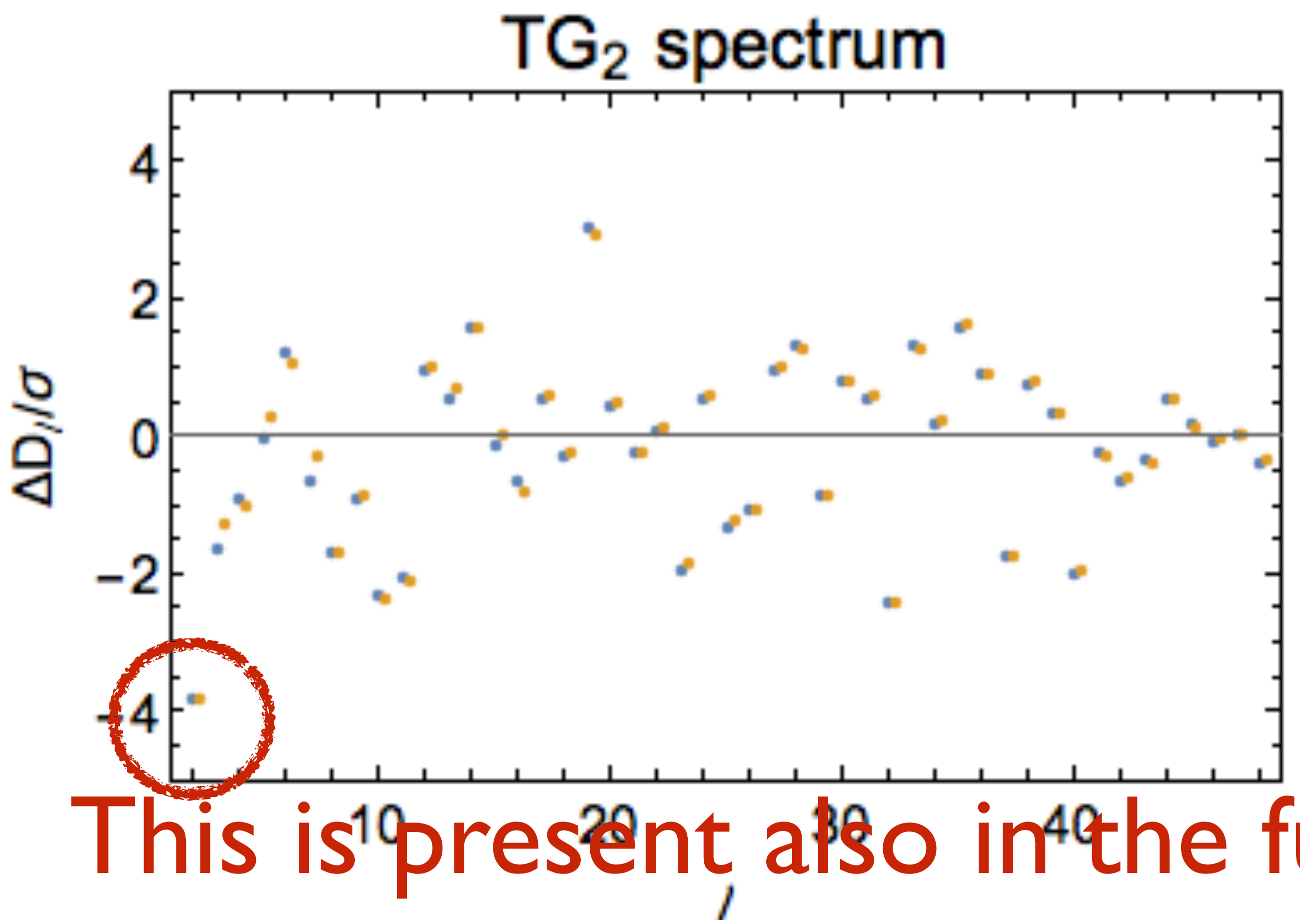


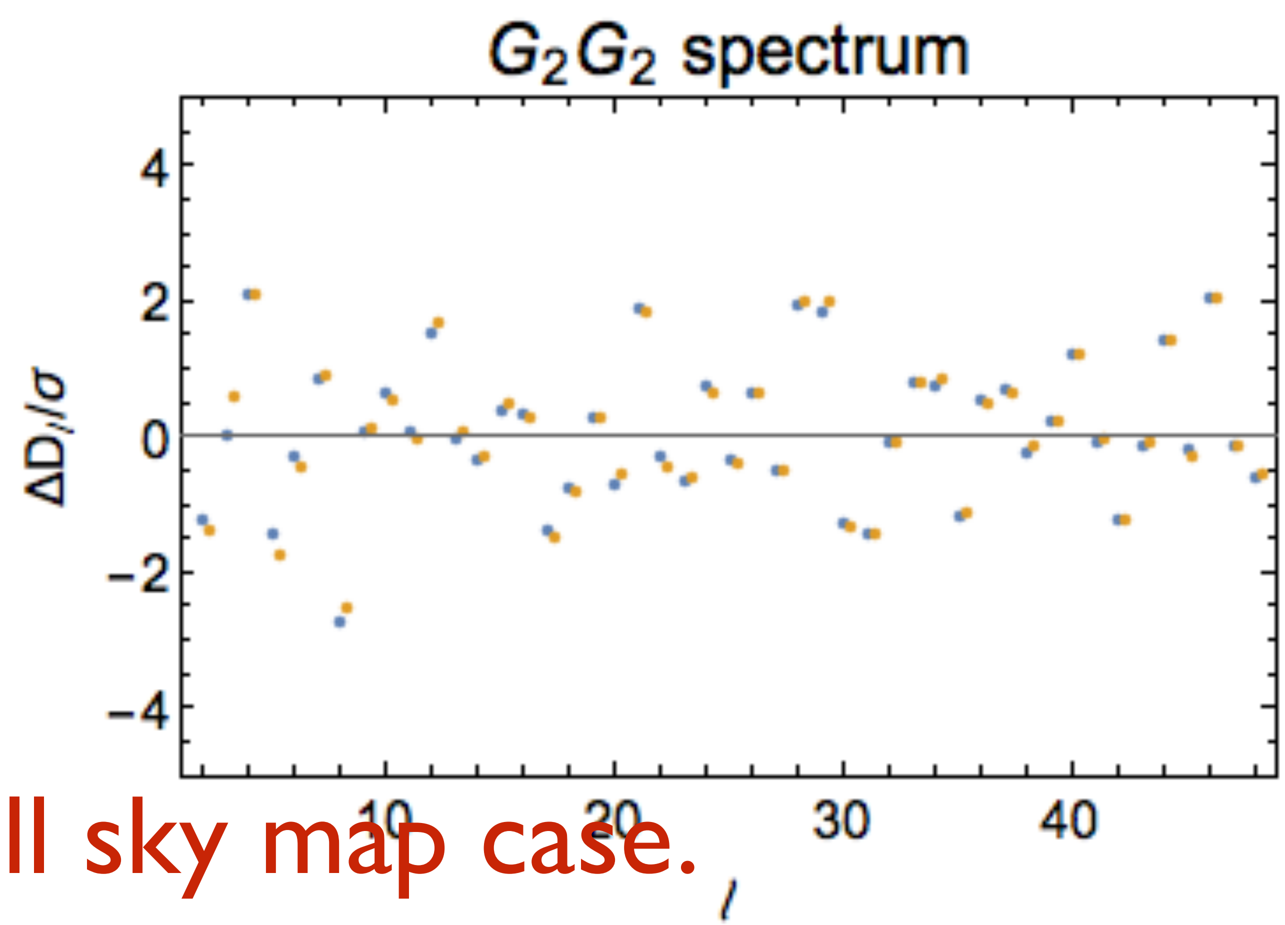
Distance of MC average (1000 sims) from the fiducial spectra in units of standard deviation of the mean

Validation (nside=16)

Blue estimates = estimator built with the same fiducial used for the MC

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

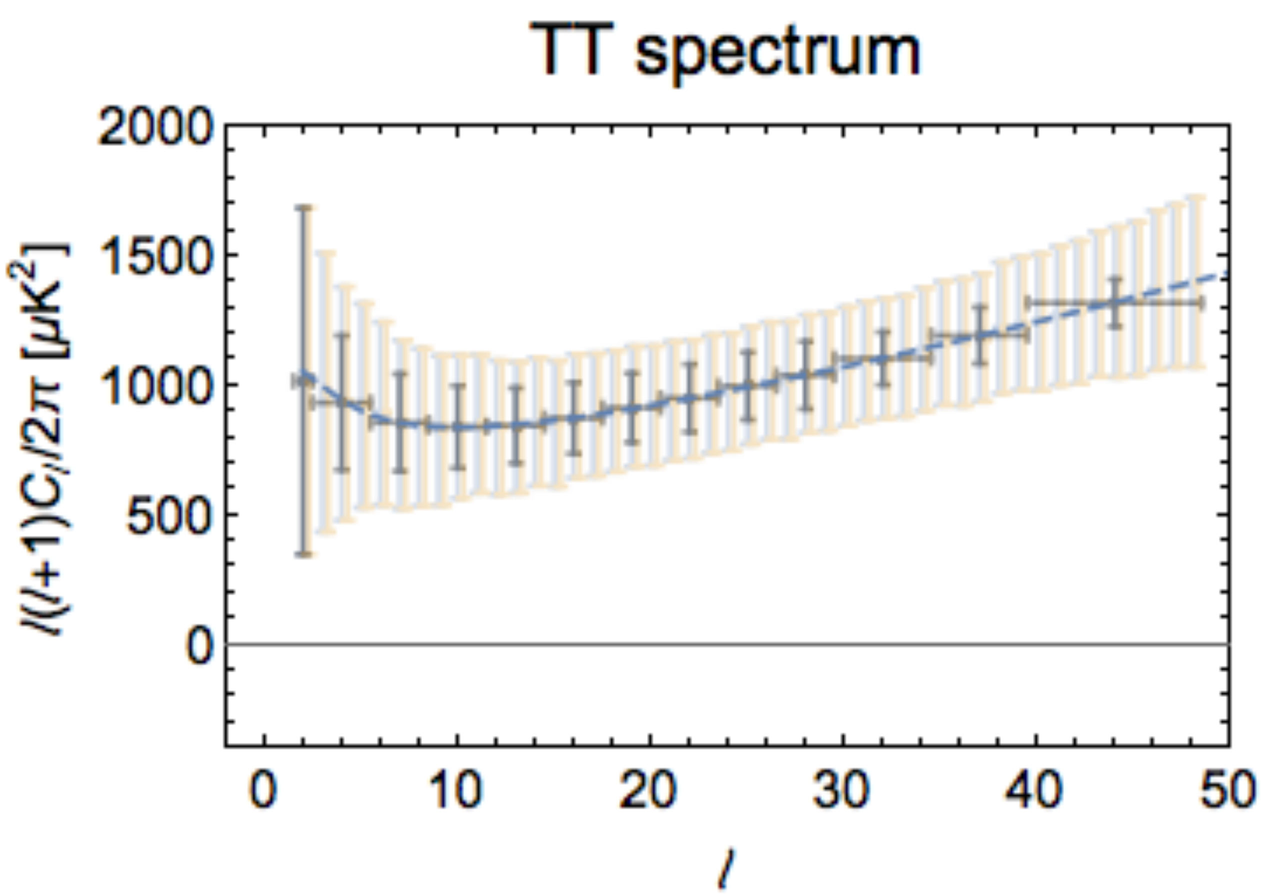


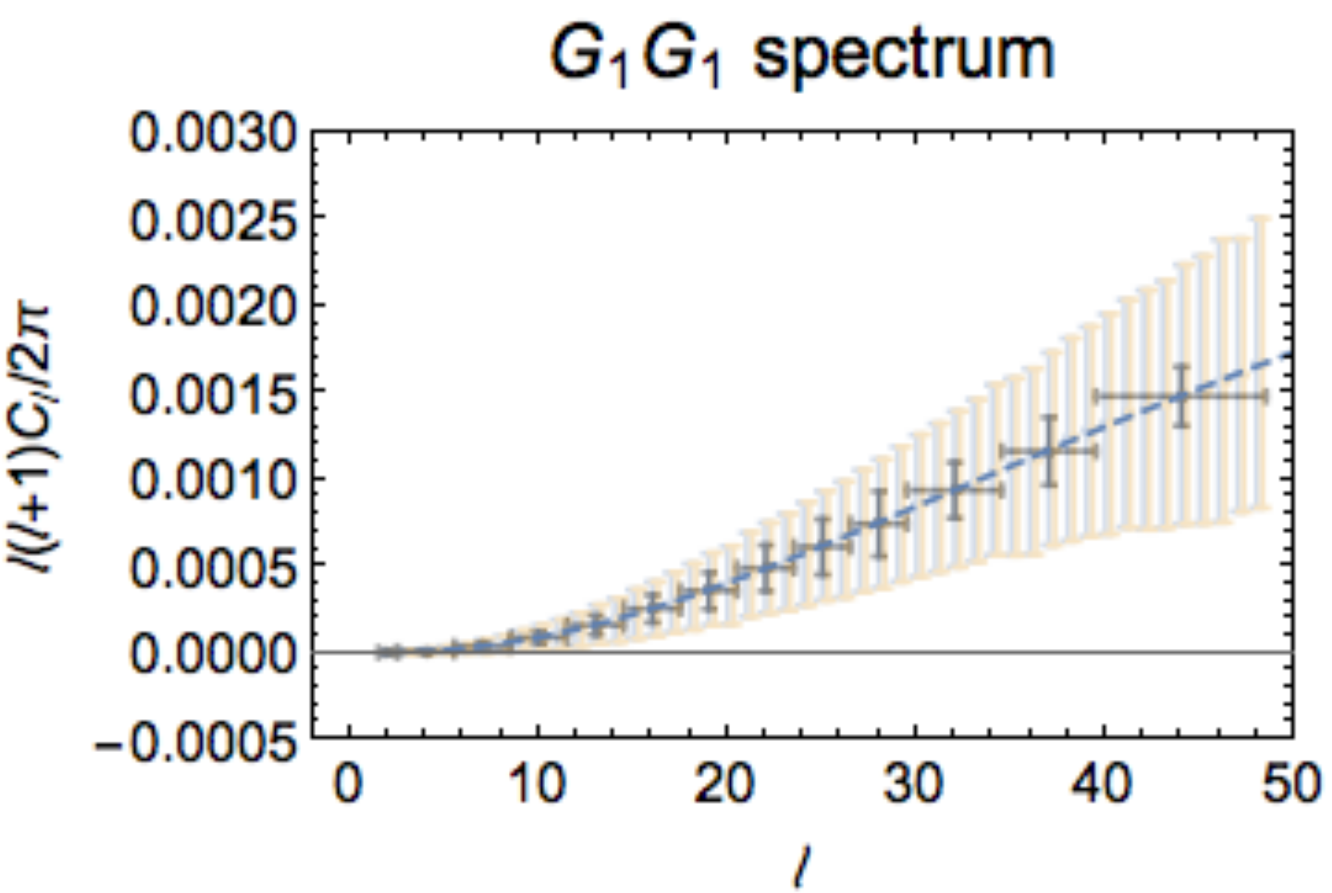


This is present also in the full sky map case.
This needs further investigation

Distance of MC average (1000 sims) from the fiducial spectra in units of standard deviation of the mean

Spectra (nside=16)



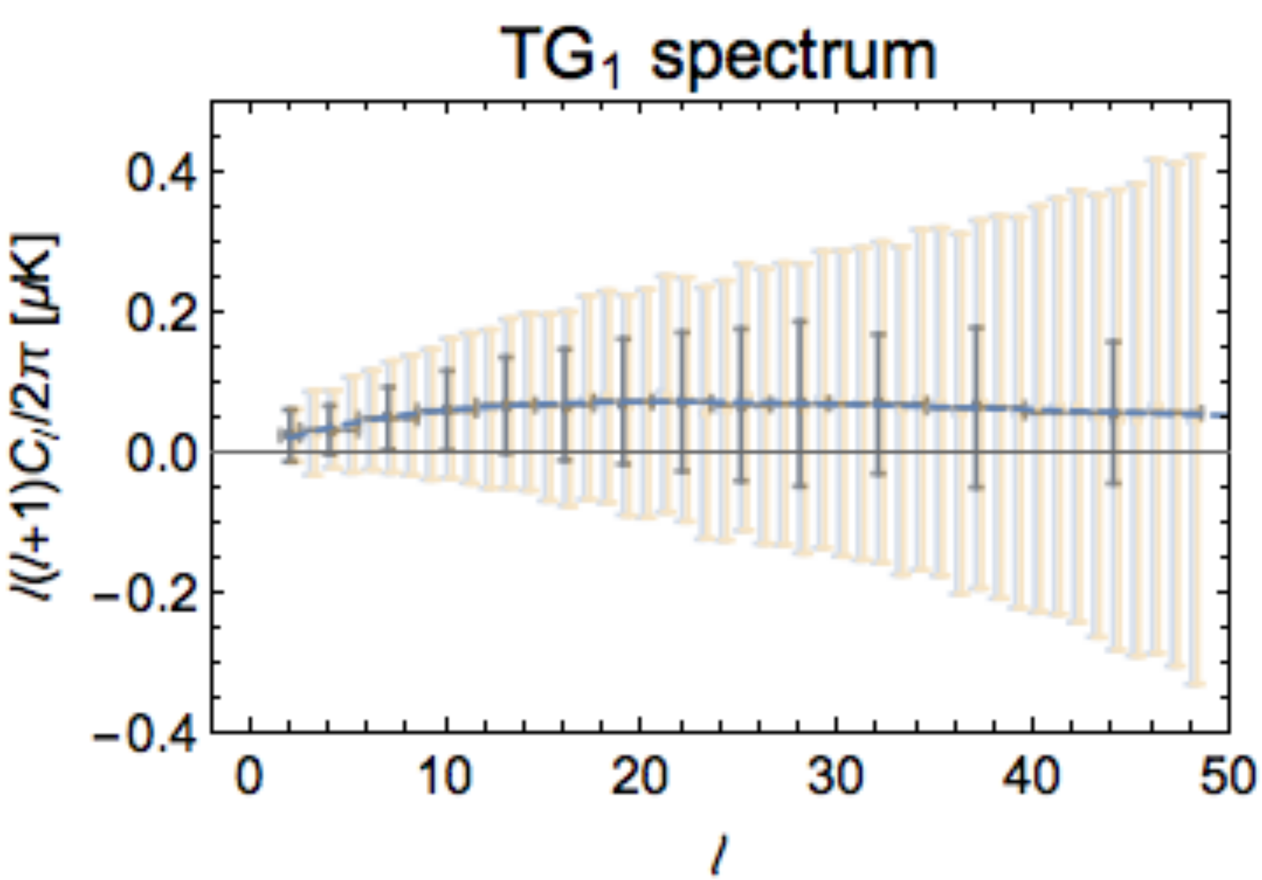


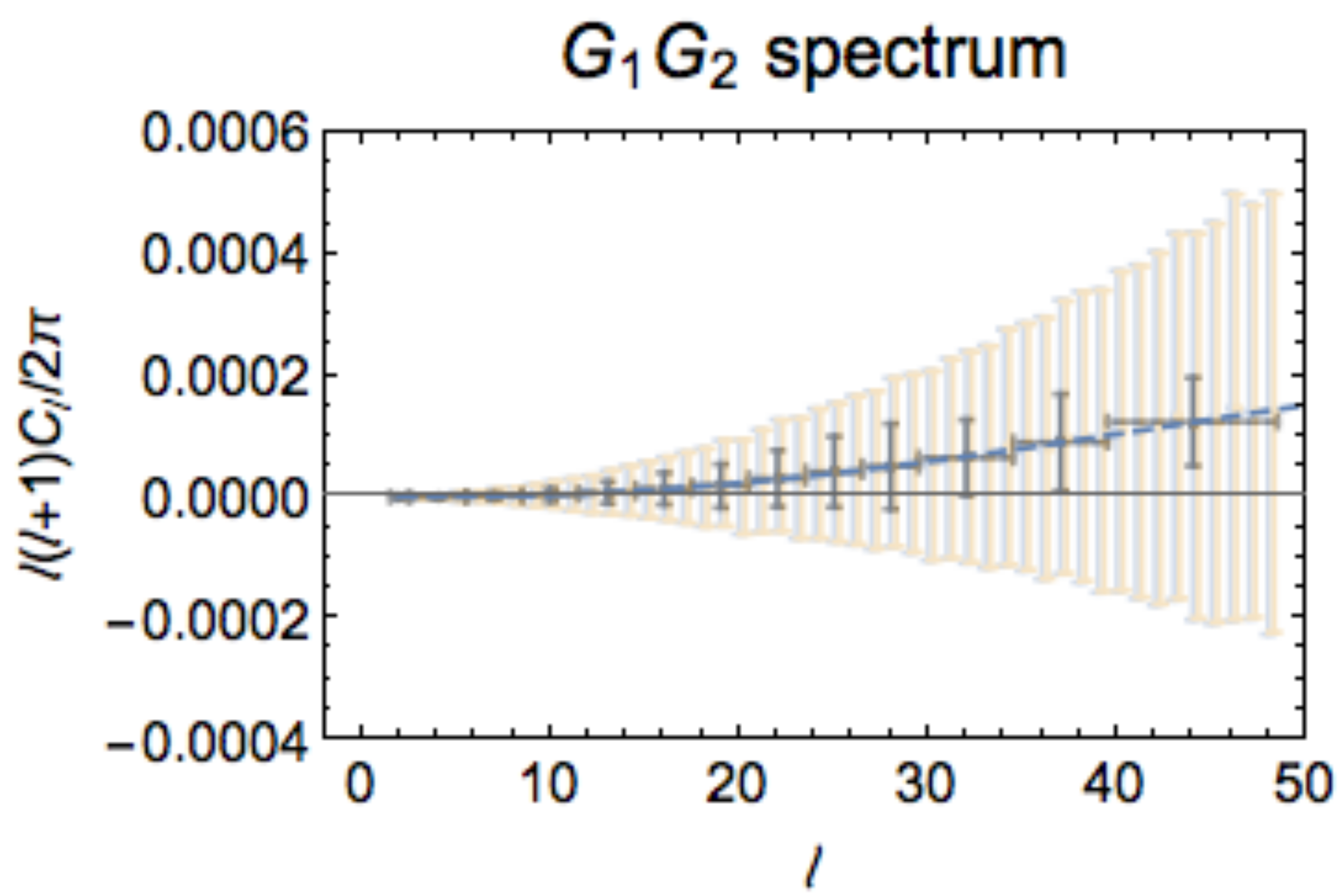
Blue estimates = estimator built with the same fiducial used for the MC

Error = Fisher matrix (single realisation)

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

Spectra (nside=16)



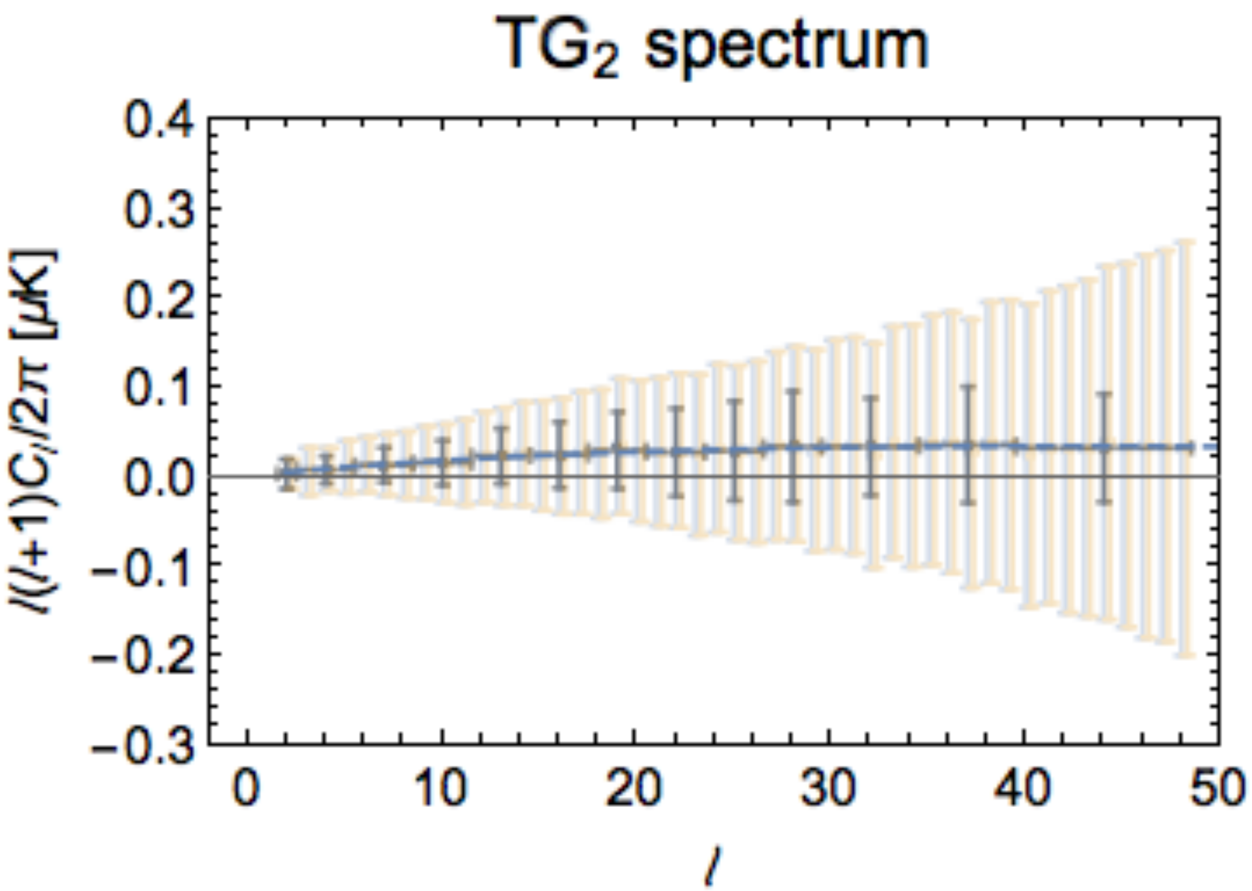


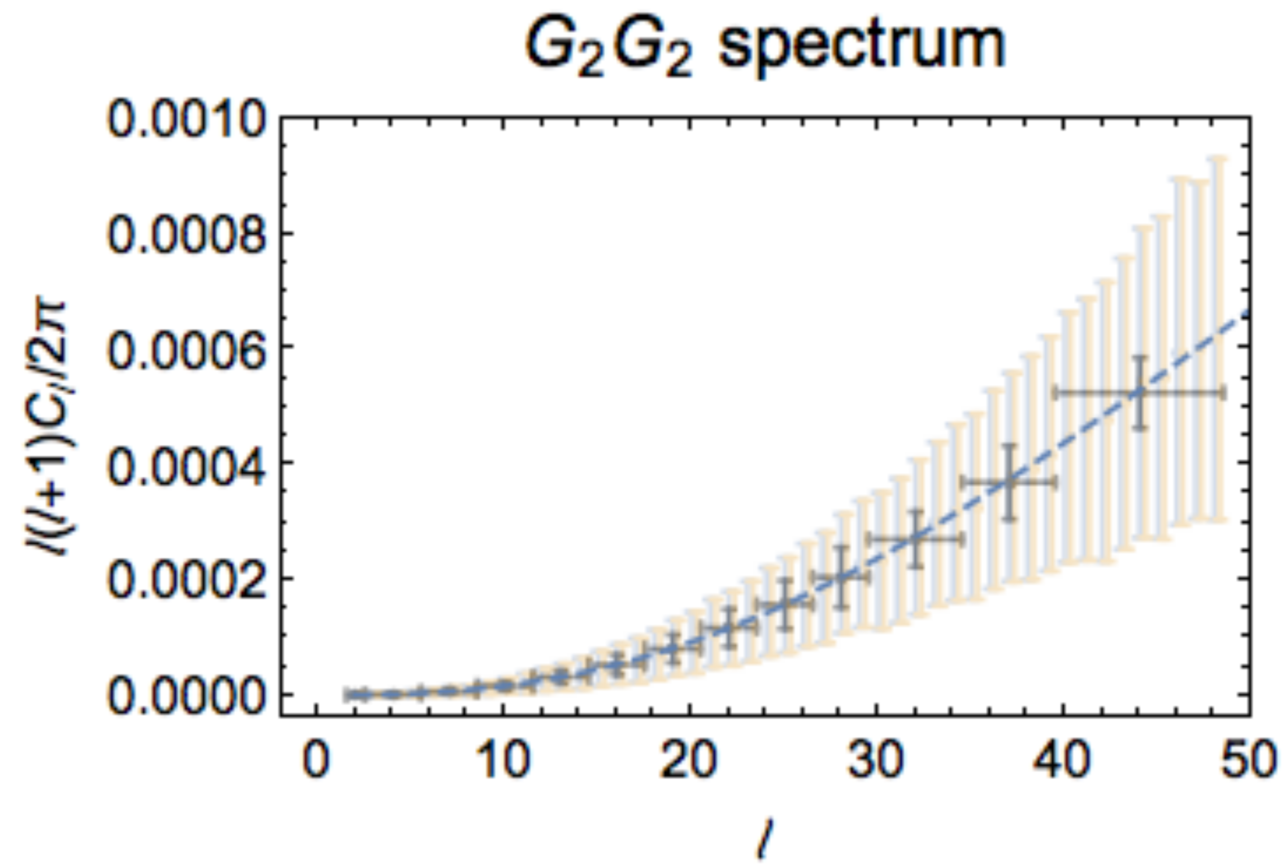
Blue estimates = estimator built with the same fiducial used for the MC

Error = Fisher matrix (single realisation)

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

Spectra (nside=16)





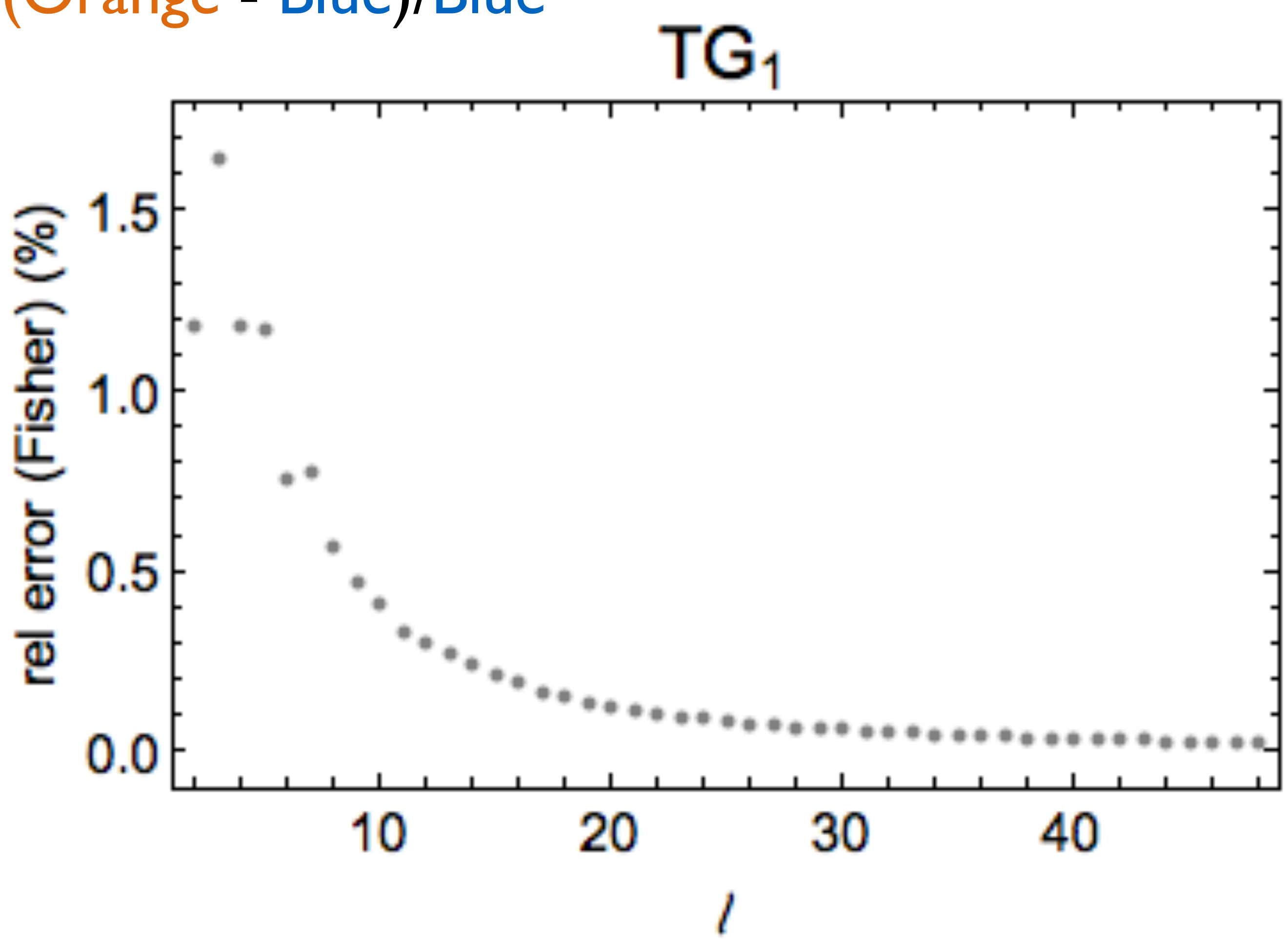
Blue estimates = estimator built with the same fiducial used for the MC

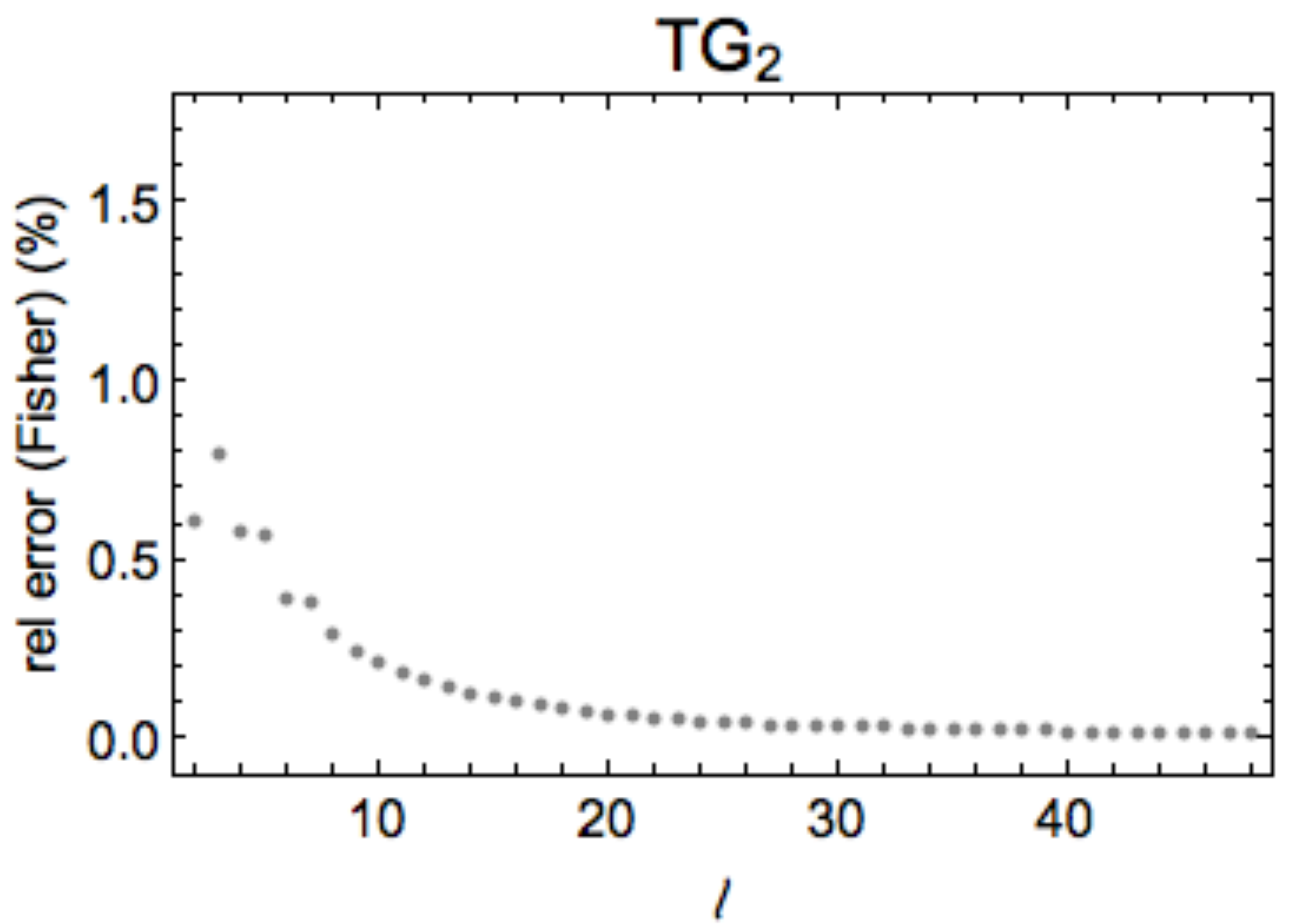
Error = Fisher matrix (single realisation)

Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

Impact on Fisher uncertainties

“(Orange - Blue)/Blue “



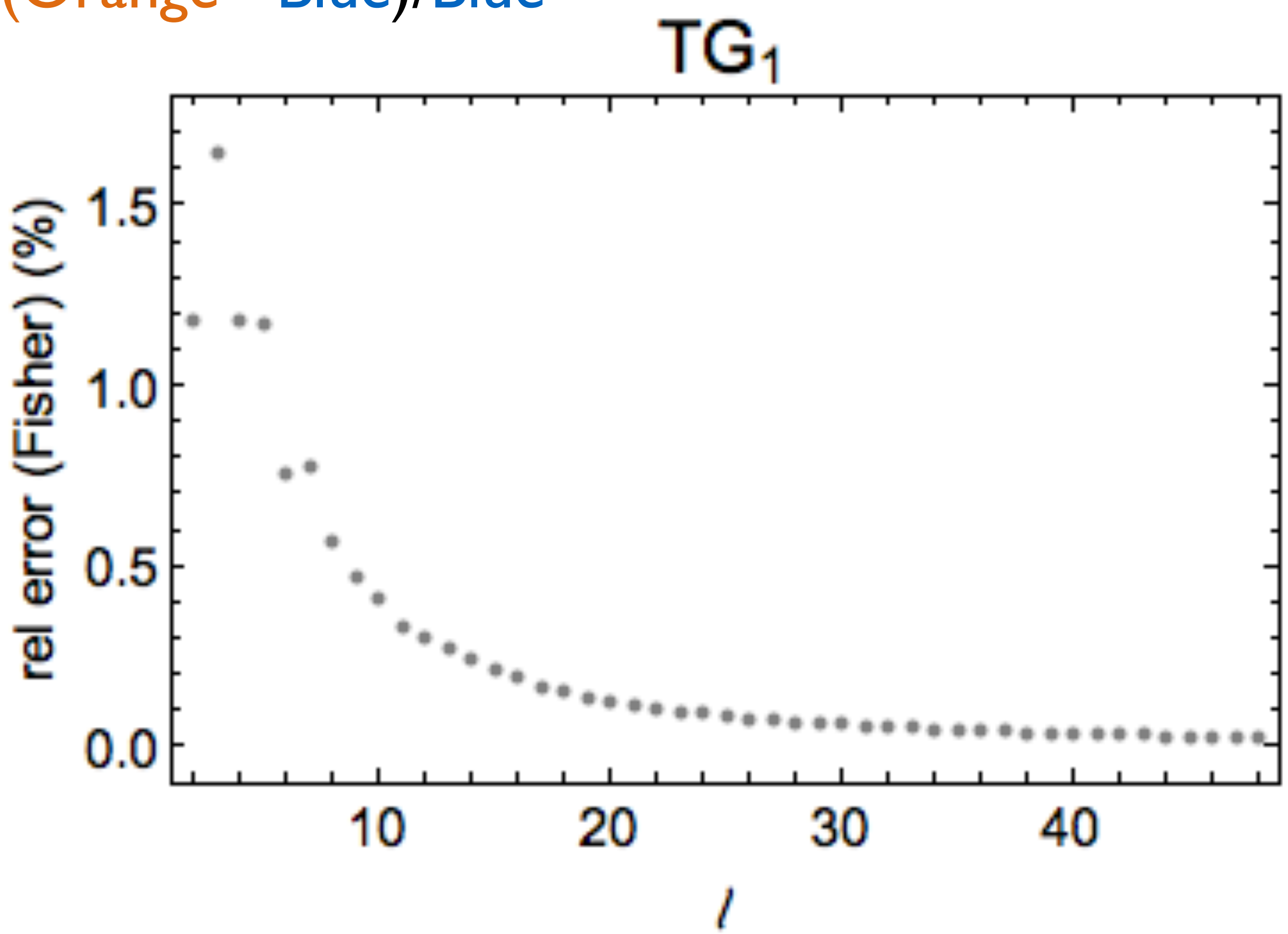


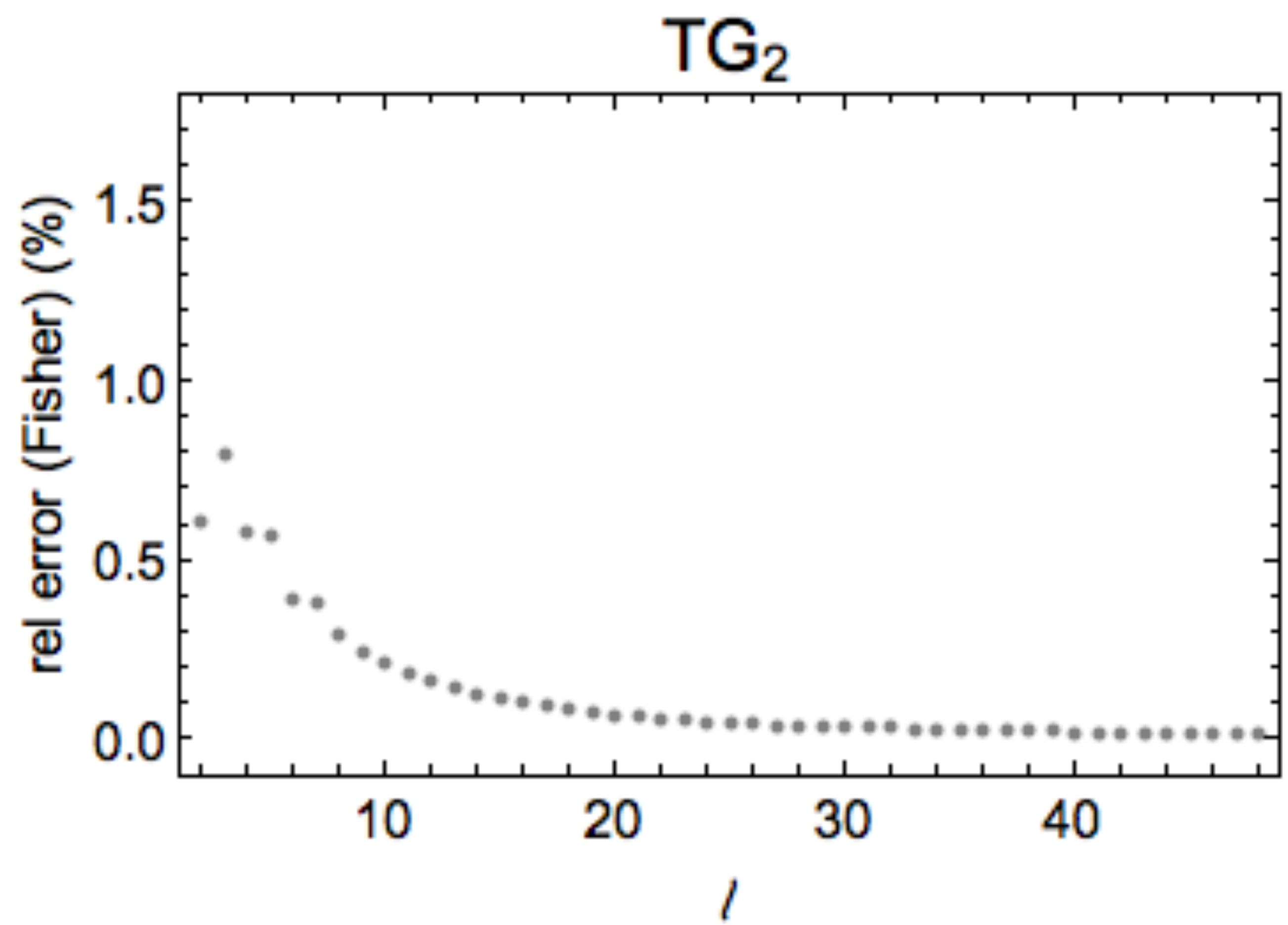
When the fiducial is not exactly the one used generate the MC, the QML is not exactly optimal anymore. Here we compare how much error bars are larger because of this non-optimality.

This quantifies the increase of the error in percentage for each multipole when we use this simplification

Impact on Fisher uncertainties

“(Orange - Blue)/Blue “





Note that if the peak of the cross correlation is not below ~10 the impact might be mild. However this has to be quantified at the level of cosmological parameters.

How to make the computation lighter

Example: case of a CMB map and one Galaxy survey

QML algebra

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG} \\ \hat{C}_\ell^{GG} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG} \\ y_{\ell'}^{GG} \end{pmatrix}$$

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$C = \langle x x^t \rangle = \begin{pmatrix} C_{CMB} & C_{TG} \\ C_{TG}^t & C_G \end{pmatrix}$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} \text{TT TT} & \text{TT TG} & \text{TT GG} \\ \text{TG TT} & \text{TG TG} & \text{TG GG} \\ \text{GG TT} & \text{GG TG} & \text{GG GG} \end{pmatrix}$$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} Tr \left[C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y \right]$$

$$X, Y = TT, TG, GG$$

How to make the computation lighter

Example: case of a CMB map and one Galaxy survey

QML algebra

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG} \\ \hat{C}_\ell^{GG} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG} \\ y_{\ell'}^{GG} \end{pmatrix}$$

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$C = \langle x x^t \rangle = \begin{pmatrix} C_{CMB} & \cancel{C_{TG}} \\ \cancel{C_{TG}} & C_G \end{pmatrix}$$

$$C_\ell^{fid,TG} = 0$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} \mathrm{TT\,TT} & \cancel{\mathrm{TT\,TG}} & \cancel{\mathrm{TT\,GG}} \\ \cancel{\mathrm{TG\,TT}} & \mathrm{TG\,TG} & \cancel{\mathrm{TG\,GG}} \\ \cancel{\mathrm{GG\,TT}} & \cancel{\mathrm{GG\,TG}} & \mathrm{GG\,GG} \end{pmatrix}$$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} Tr \left[C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y \right]$$

$$X, Y = TT, TG, GG$$

How to make the computation lighter

Example: case of a CMB map and one Galaxy survey

THREE INDEPENDENT QMLS

$$\hat{C}_\ell^{TT} = \left(F_{\ell\ell'}^{TT,TT}\right)^{-1} y_{\ell'}^{TT}$$

$$\hat{C}_\ell^{TG} = \left(F_{\ell\ell'}^{TG,TG}\right)^{-1} y_{\ell'}^{TG}$$

$$\hat{C}_\ell^{GG} = \left(F_{\ell\ell'}^{GG,GG}\right)^{-1} y_{\ell'}^{GG}$$

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$C = \langle x x^t \rangle = \begin{pmatrix} C_{CMB} & C_{TG} \\ C_{TG} & C_G \end{pmatrix}$$

$$C_\ell^{fid,TG} = 0$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} \text{TT TT} & \text{TT TG} & \text{TT GG} \\ \text{TG TT} & \text{TG TG} & \text{TG GG} \\ \text{GG TT} & \text{GG TG} & \text{GG GG} \end{pmatrix}$$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} Tr\left[C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y\right]$$

$$X, Y = TT, TG, GG$$

This means going from 43 GB to 5.4GB at nside 64 (our case, with mask)

e computation lighter

CMB map and one Galaxy survey

THREE INDEPENDENT QMLS

$$x = (x_{CMB}, x_G)$$

$$\hat{C}_\ell^{TT} = \left(F_{\ell\ell'}^{TT,TT}\right)^{-1} y_{\ell'}^{TT}$$

$$\hat{C}_\ell^{TG} = \left(F_{\ell\ell'}^{TG,TG}\right)^{-1} y_{\ell'}^{TG}$$

$$\hat{C}_\ell^{GG} = \left(F_{\ell\ell'}^{GG,GG}\right)^{-1} y_{\ell'}^{GG}$$

and the covariance matrix is

$$C = \langle x x^t \rangle = \begin{pmatrix} C_{CMB} & C_{TG} \\ C_{TG} & C_G \end{pmatrix}$$

$$C_\ell^{fid,TG} = 0$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} \text{TT TT} & \text{TT TG} & \text{TT GG} \\ \text{TG TT} & \text{TG TG} & \text{TG GG} \\ \text{GG TT} & \text{GG TG} & \text{GG GG} \end{pmatrix}$$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} Tr\left[C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y\right]$$

$$X, Y = TT, TG, GG$$

Example at nside=64

The job is submitted and queuing at NERSC

Summary and next steps

- A QML code has been validated and put to work (it deals with one CMB map and two galaxy counts maps)
- Validation successfully completed. Only issue, outliers at specific multipoles. Turns out to be a simulation (not a QML) issue, which needs to be investigated.
- Binned option included. Important to go to nside=64.
- We can start the comparison with pseudoCell
- We can quantify in terms of uncertainties of the cosmological parameters the impact of setting to zero the cross-terms in the covariance matrix (Likelihood code is needed). This might lead us to build a lighter QML only for TG.