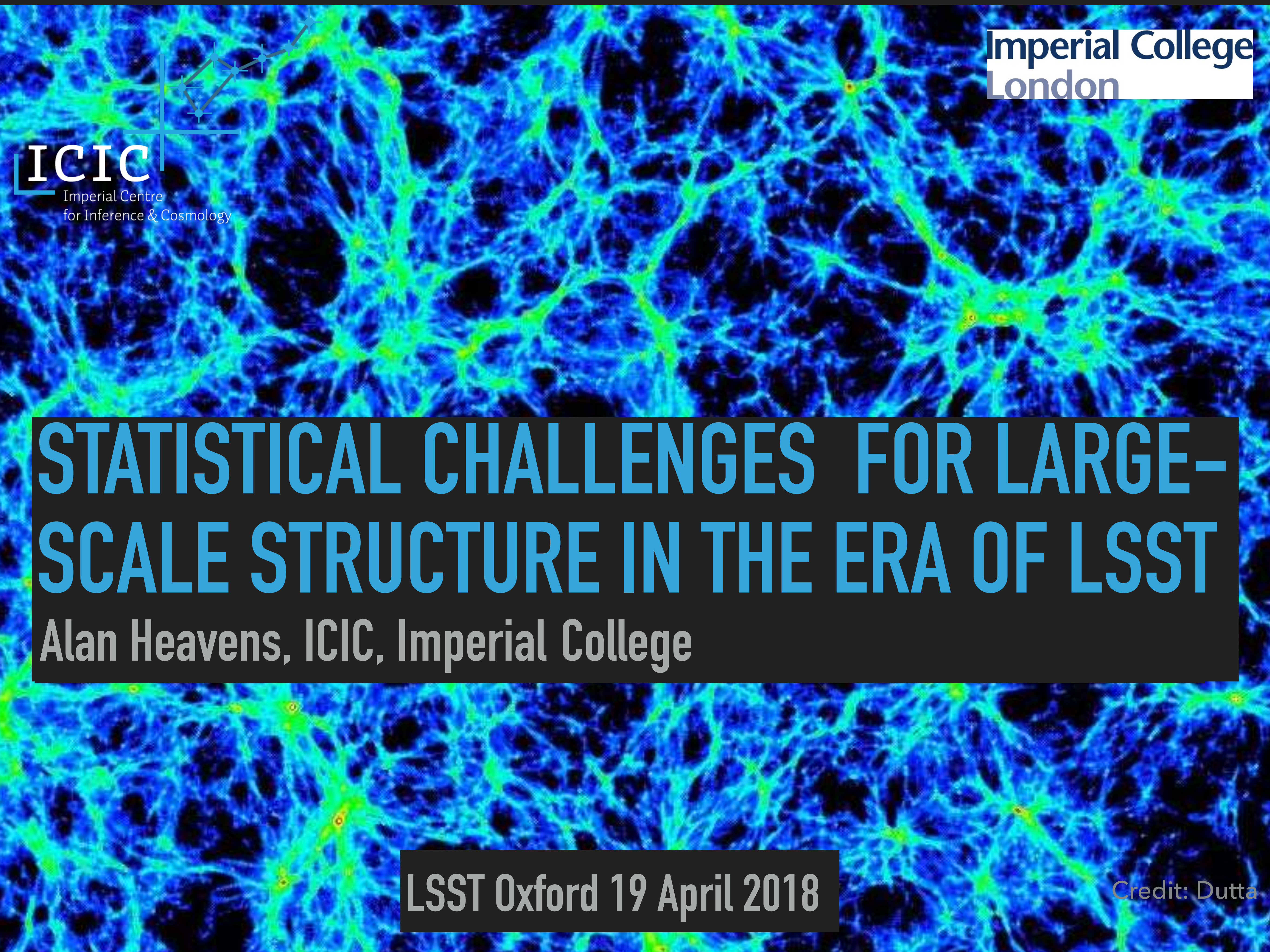


Imperial College London

ICIC Imperial Centre for Inference & Cosmology

STATISTICAL CHALLENGES FOR LARGE-SCALE STRUCTURE IN THE ERA OF LSST

Alan Heavens, ICIC, Imperial College

LSST Oxford 19 April 2018

Credit: Dutta

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

- Bayesian: we want the posterior: $p(\boldsymbol{\theta} \mid \mathbf{d})$

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

- Bayesian: we want the posterior: $p(\boldsymbol{\theta} \mid \mathbf{d})$
- $\boldsymbol{\theta}$ = cosmological parameters, $\mathbf{d}$ = data

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

- Bayesian: we want the posterior: $p(\boldsymbol{\theta} \mid \mathbf{d})$
- $\boldsymbol{\theta}$ = cosmological parameters, $\mathbf{d}$ = data
- Bayes: $p(\boldsymbol{\theta} \mid \mathbf{d}) \propto p(\mathbf{d} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

- Bayesian: we want the posterior: $p(\boldsymbol{\theta} \mid \mathbf{d})$
- $\boldsymbol{\theta}$ = cosmological parameters, $\mathbf{d}$ = data
- Bayes: $p(\boldsymbol{\theta} \mid \mathbf{d}) \propto p(\mathbf{d} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$
- Likelihood: $p(\mathbf{d} \mid \boldsymbol{\theta})$

WHAT ARE THE STATISTICAL CHALLENGES WITH LSST?

- Bayesian: we want the posterior: $p(\boldsymbol{\theta} \mid \mathbf{d})$
- $\boldsymbol{\theta}$ = cosmological parameters, $\mathbf{d}$ = data
- Bayes: $p(\boldsymbol{\theta} \mid \mathbf{d}) \propto p(\mathbf{d} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$
- Likelihood: $p(\mathbf{d} \mid \boldsymbol{\theta})$
- **d**? Typically summary statistics such as correlation function or power spectrum estimates. Already a massive data compression. Perhaps 10^2-10^4 summary statistics

GAUSSIAN LIKELIHOOD APPROXIMATION

GAUSSIAN LIKELIHOOD APPROXIMATION

- Often, we assume that the summary statistics are gaussian-distributed

GAUSSIAN LIKELIHOOD APPROXIMATION

- Often, we assume that the summary statistics are gaussian-distributed
- (Handwave, handwave, central limit theorem…)

GAUSSIAN LIKELIHOOD APPROXIMATION

- Often, we assume that the summary statistics are gaussian-distributed
- (Handwave, handwave, central limit theorem…)
- We rarely stop to question this, but we should. Let us run with it for now

REQUIREMENTS FOR GAUSSIAN LIKELIHOODS

- Data normally distributed: **d** ~ $N_d(\boldsymbol{\mu},\Sigma)$

$$p(\mathbf{d}|\theta) = |2\pi\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{d}-\mu)^{\mathbf{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{d}-\mu)\right]$$

- In general, both $\boldsymbol{\mu}$ and the covariance matrix Σ depend on cosmological parameters
- $\boldsymbol{\mu}$ would come from theory or simulation.
- Problem is Σ.

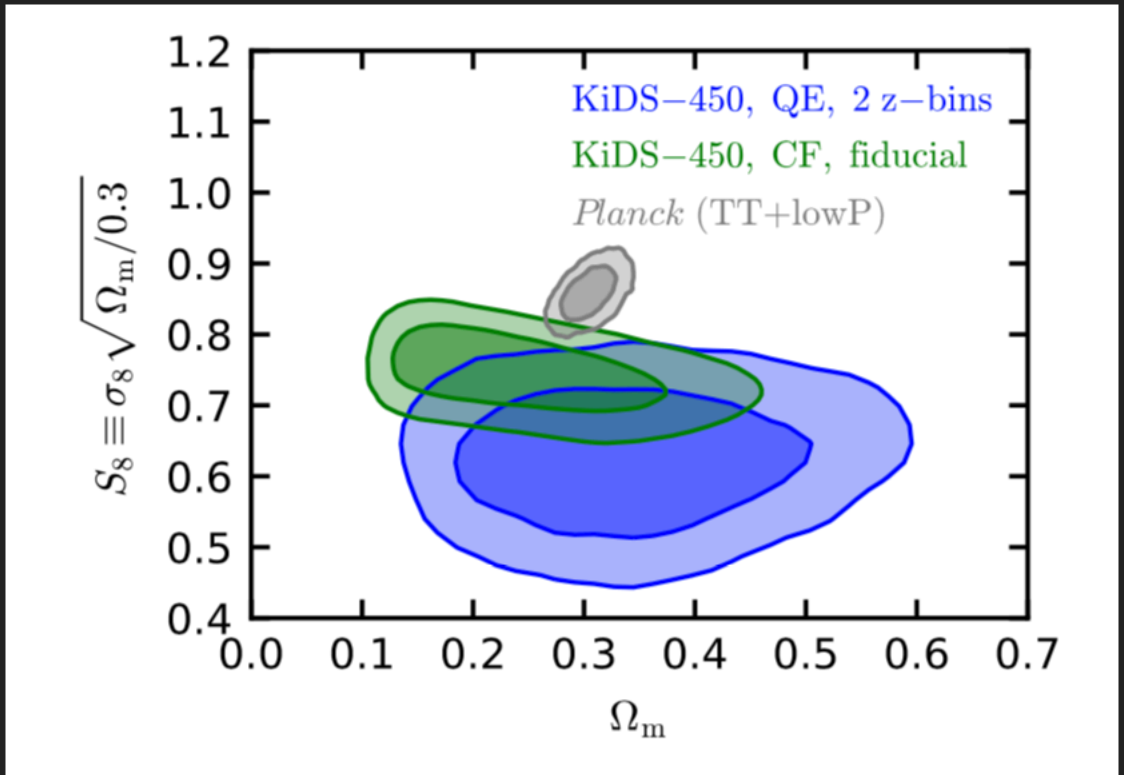
COVARIANCE MATRIX

- If summary statistics are 2-point functions, Σ is a 4-point function. Hard to compute for non-gaussian fields.
- Either use analytic covariance matrix, or simulate (or both)
- For simulated covariance matrices, $\hat{\Sigma}$ can be unbiased. Note that some effects are not included - e.g. super-sample covariance.
- However, $\hat{\Sigma}^{-1}$ is not unbiased. A fix is the Hartlap et al (2007) correction (N-1)/(N-p-2). p = number of data; N = no. of sims.
- Marginalise over Σ → likelihood of Sellentin & Heavens (2016)
- Further discussion: e.g. Friedrich & Eifler (2016), Joachimi (2017)

Elena Sellentin

CHANGING THE COVARIANCE MATRIX MATTERS

- e.g. KiDS weak lensing result (on S_8) shifts by 1σ when changing from an analytic to a simulated covariance matrix (Hildebrandt et al 2017)



Köhlinger et al 2017

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

CHALLENGE: NUMBER OF SIMULATIONS REQUIRED, N

CHALLENGE: NUMBER OF SIMULATIONS REQUIRED, N

- Need N>p+2, where p = number of summary statistics

CHALLENGE: NUMBER OF SIMULATIONS REQUIRED, N

- Need N>p+2, where p = number of summary statistics
- p could easily be 10^4 for LSST

CHALLENGE: NUMBER OF SIMULATIONS REQUIRED, N

- Need N>p+2, where p = number of summary statistics
- p could easily be 10^4 for LSST
- If Σ varies with cosmological parameters (as it will), then it's worse. Estimating Σ would be prohibitively expensive

CHALLENGE: NUMBER OF SIMULATIONS REQUIRED, N

- Need N>p+2, where p = number of summary statistics
- p could easily be 10^4 for LSST
- If Σ varies with cosmological parameters (as it will), then it's worse. Estimating Σ would be prohibitively expensive
- Solution: reduce p. Data compression

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

MOPED ALGORITHM

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)
- Variation by Zablocki & Dodelson (2016). Generalised by Alsing & Wandelt (2018). See also Charnock et al (2018)

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)
- Variation by Zablocki & Dodelson (2016). Generalised by Alsing & Wandelt (2018). See also Charnock et al (2018)
- $y_\alpha = \mathbf{b}_\alpha . \mathbf{x}$

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)
- Variation by Zablocki & Dodelson (2016). Generalised by Alsing & Wandelt (2018). See also Charnock et al (2018)
- $y_\alpha = \mathbf{b}_\alpha . \mathbf{x}$

$$\mathbf{b}_1 = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}^T \mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}}$$

and

$$\mathbf{b}_\alpha = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T \mathbf{b}_\beta)\mathbf{b}_\beta}{\sqrt{\boldsymbol{\mu}_{,\alpha}^T \mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T \mathbf{b}_\beta)^2}} \qquad 1 < \alpha \le m,$$

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)
- Variation by Zablocki & Dodelson (2016). Generalised by Alsing & Wandelt (2018). See also Charnock et al (2018)
- $y_\alpha = \mathbf{b}_\alpha . \mathbf{x}$
- Size of dataset reduced to the ***number of parameters***. *Same* Fisher Matrix.

$$\mathbf{b}_1 = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}^T\mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}}$$

and

$$\mathbf{b}_\alpha = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T\mathbf{b}_\beta)\mathbf{b}_\beta}{\sqrt{\boldsymbol{\mu}_{,\alpha}^T\mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T\mathbf{b}_\beta)^2}} \qquad 1 < \alpha \leq m,$$

MOPED ALGORITHM

- Massively Optimised Parameter Estimation and Data compression (Heavens, Jimenez, Lahav 2000)

- Variation by Zablocki & Dodelson (2016). Generalised by Alsing & Wandelt (2018). See also Charnock et al (2018)

- $y_\alpha = \mathbf{b}_\alpha . \mathbf{x}$
- Size of dataset reduced to the ***number of parameters***. *Same* Fisher Matrix.

- MOPED proposed to solve the simulations problem by Heavens et al (2017) and Gualdi et al (2018).

$$\mathbf{b}_1 = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}^T \mathsf{C}^{-1}\boldsymbol{\mu}_{,1}}}$$

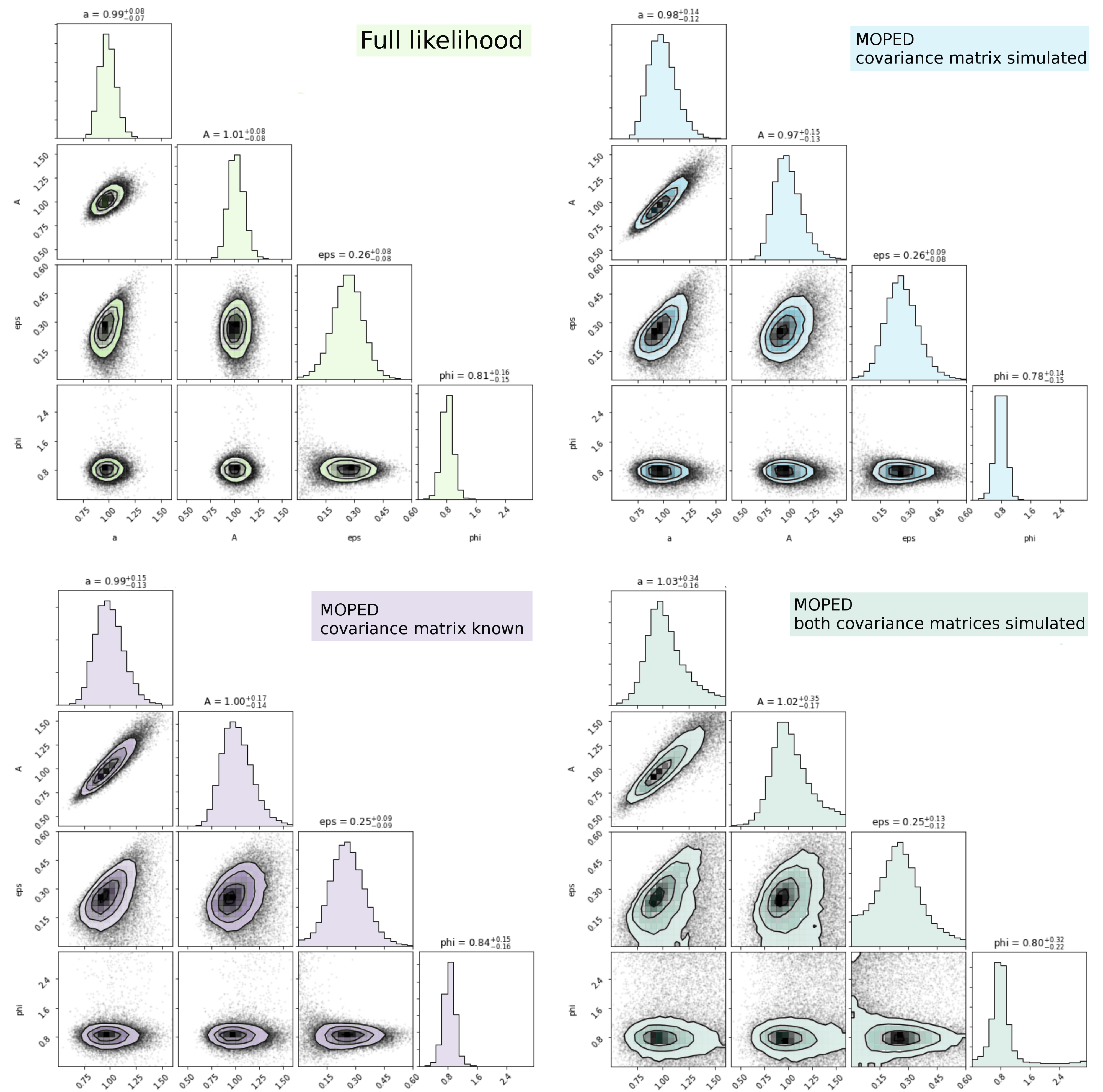
and

$$\mathbf{b}_\alpha = \frac{\mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T \mathbf{b}_\beta)\mathbf{b}_\beta}{\sqrt{\boldsymbol{\mu}_{,\alpha}^T \mathsf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{\beta=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T \mathbf{b}_\beta)^2}} \qquad 1 < \alpha \leq m,$$

MOPED

INFERENCE

- Need full Σ for compression
- Do it once
- Degradation: optimal only for correct parameters
- Estimate compressed Σ as well



Heavens et al 2017

ICIC Imperial Centre for Inference & Cosmology

NUMBER OF SIMULATIONS NEEDED

- Reduces number by up to six orders of magnitude

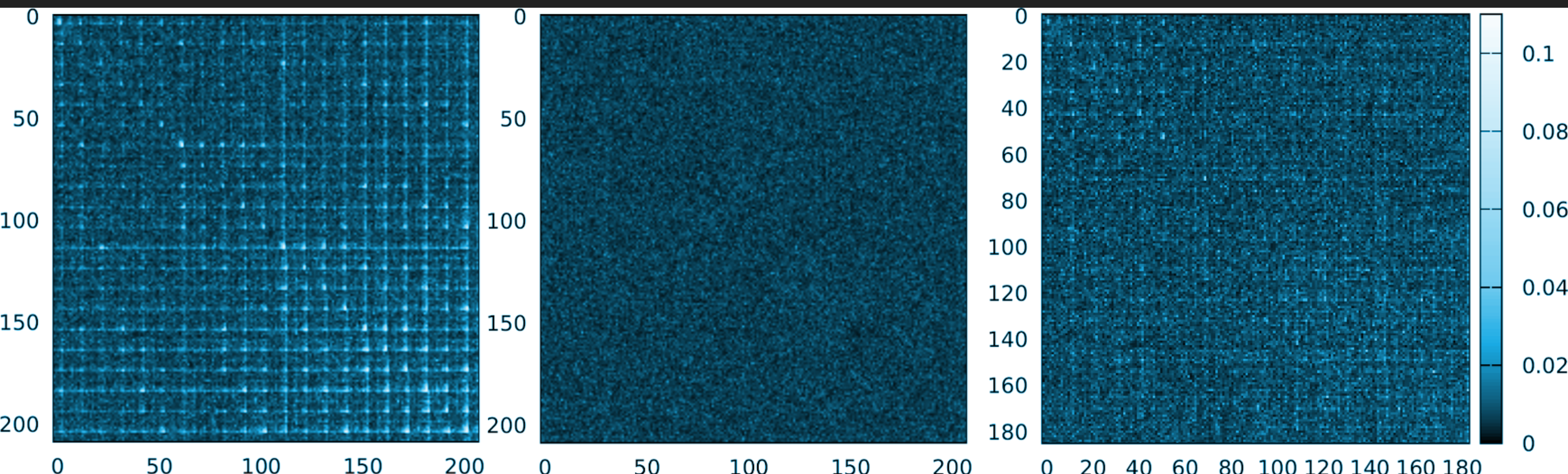
Estimating C^y at:	emulator locations;	each MCMC point.
No compression	10^6	10^9
MOPED compression, using simulated C^x	10^4	10^6
MOPED compression, using analytic/theoretical C^x	10^3	10^6

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

ARE WE FOCUSSING ON THE WRONG PROBLEM?

- The data are not Gaussian-distributed, even when the CLT handwave suggests otherwise…



CFHTLenS data | Gaussian data | CFHTLenS without the most NG terms

Sellentin & Heavens 2017

NON-GAUSSIAN LIKELIHOODS I

- From a Bayesian perspective, we want the full likelihood (sampling distribution) e.g. p(all Fourier Coefficients | parameters)

NON-GAUSSIAN LIKELIHOODS I

- From a Bayesian perspective, we want the full likelihood (sampling distribution) e.g. p(all Fourier Coefficients | parameters)
- Challenging: few tools. e.g. Edgeworth expansion
- Schematically
- B = bispectrum, T = trispectrum

NON-GAUSSIAN LIKELIHOODS I

- From a Bayesian perspective, we want the full likelihood (sampling distribution) e.g. p(all Fourier Coefficients | parameters)
- Challenging: few tools. e.g. Edgeworth expansion
- Schematically

$$p(a_{\mathbf{k}}|\theta) = |\mathrm{diag}[2\pi P(k)]|^{-1/2} \exp\left[-\frac{|a_{\mathbf{k}}|^2}{2P(k)}\right] \left\{1 + B(\theta) + T(\theta) + B^2(\theta) + \ldots\right\}$$

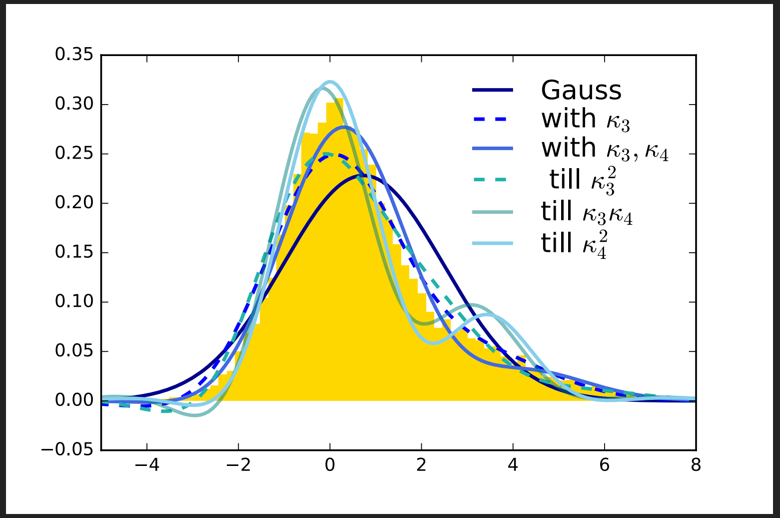
- B = bispectrum, T = trispectrum

NON-GAUSSIAN LIKELIHOODS I

- From a Bayesian perspective, we want the full likelihood (sampling distribution) e.g. p(all Fourier Coefficients | parameters)
- Challenging: few tools. e.g. Edgeworth expansion
- Schematically

$$p(a_{\mathbf{k}}|\theta) = |\mathrm{diag}[2\pi P(k)]|^{-1/2} \exp\left[-\frac{|a_{\mathbf{k}}|^2}{2P(k)}\right]\{1 + B(\theta) + T(\theta) + B^2(\theta) + \ldots\}$$

- B = bispectrum, T = trispectrum
- Gaussianising transforms? Alex (Hall and Mead)
- Large-deviation theory? (Sandrine Codis' talk)



Sellentin, Jaffe & Heavens 2018

Imperial Centre for Inference & Cosmology

Imperial College London

NG LIKELIHOODS II: FIT THE LIKELIHOOD FUNCTION NUMERICALLY

- Run many simulations; fit the sampling distribution of mocks
- e.g. Hahn et al (2018)
- Feasible in relatively small numbers of dimensions
- Probably impossible in very high dimensions
- Data compression needed again

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

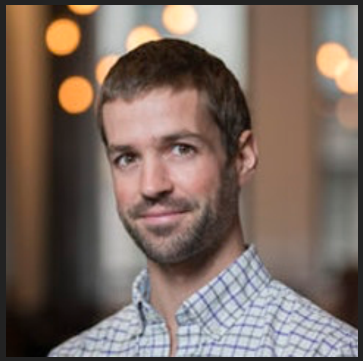
- Try to get $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint $p(\boldsymbol{\theta}, \mathbf{d})$ (Alsing et al 2018)

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get p(**θ** | **d**) directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint p(**θ** , **d**) (Alsing et al 2018)



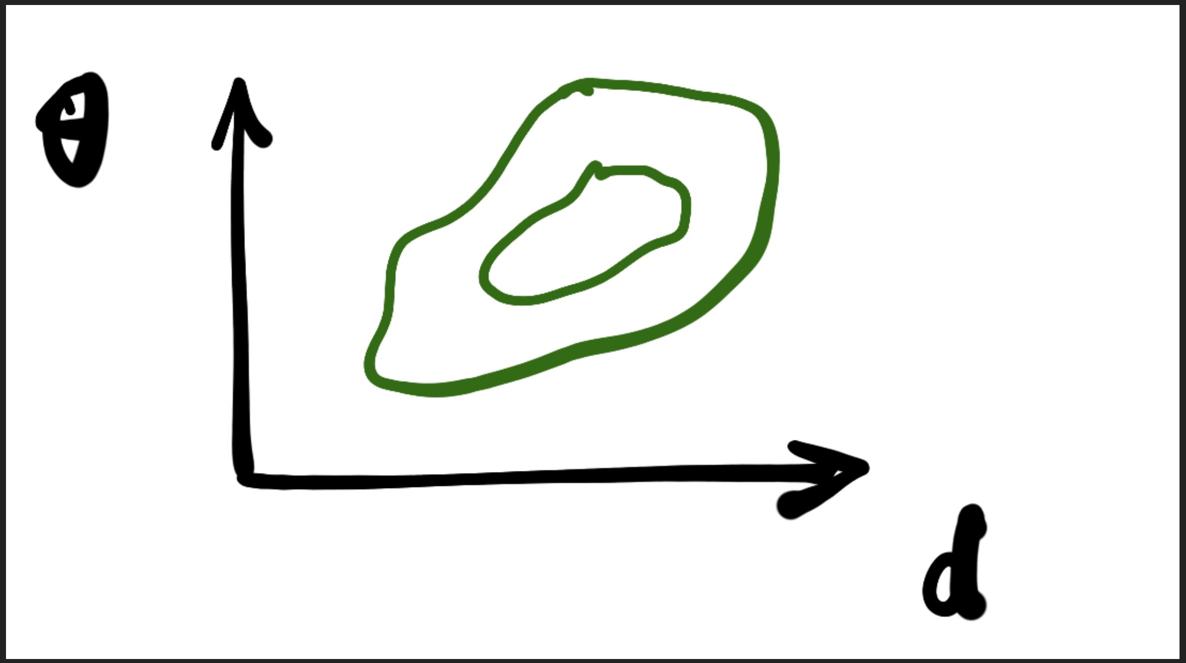
Justin Alsing

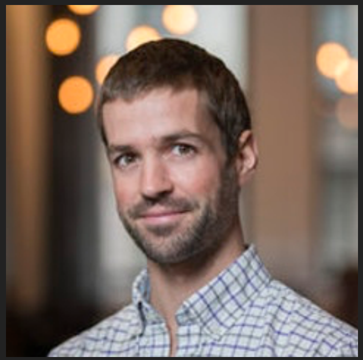
ICIC Imperial Centre for Inference & Cosmology

Imperial College London

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get $p(\boldsymbol{\theta} \mid \mathbf{d})$ directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint $p(\boldsymbol{\theta}, \mathbf{d})$ (Alsing et al 2018)





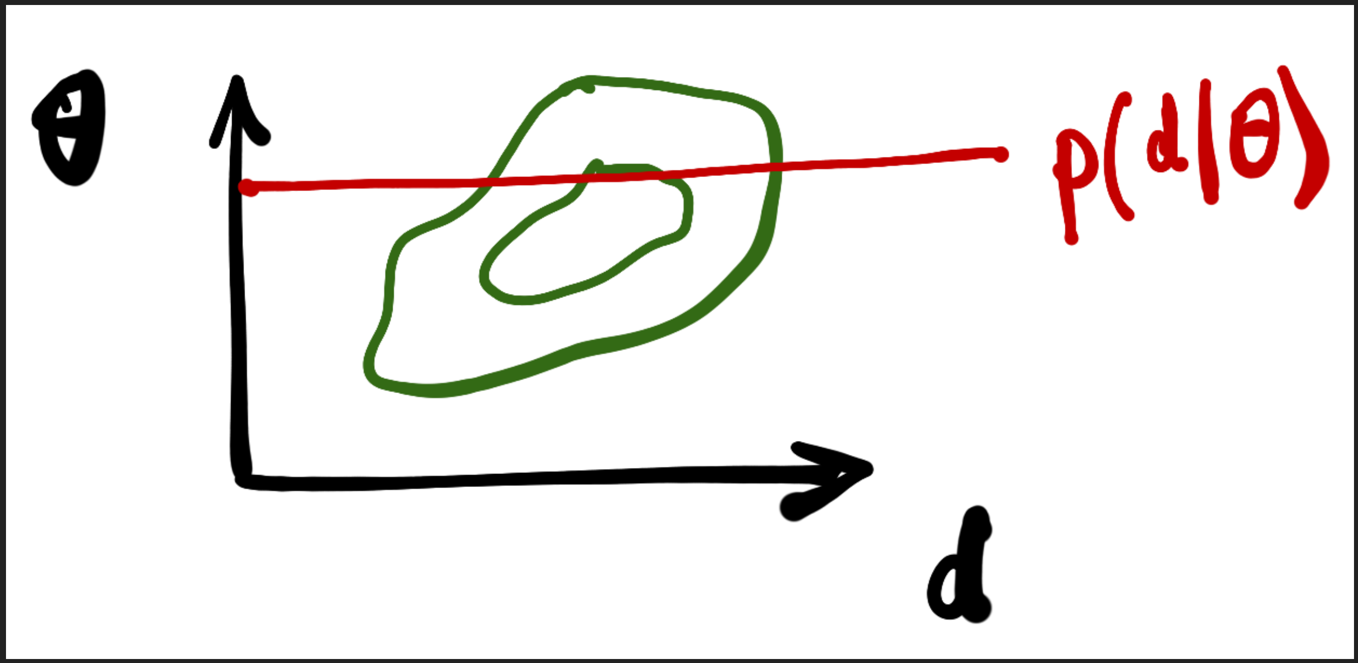
Justin Alsing

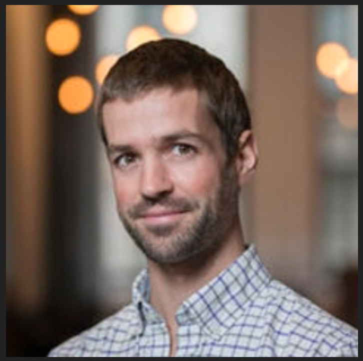
ICIC Imperial Centre for Inference & Cosmology

Imperial College London

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get p(**θ** | **d**) directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint p(**θ** , **d**) (Alsing et al 2018)

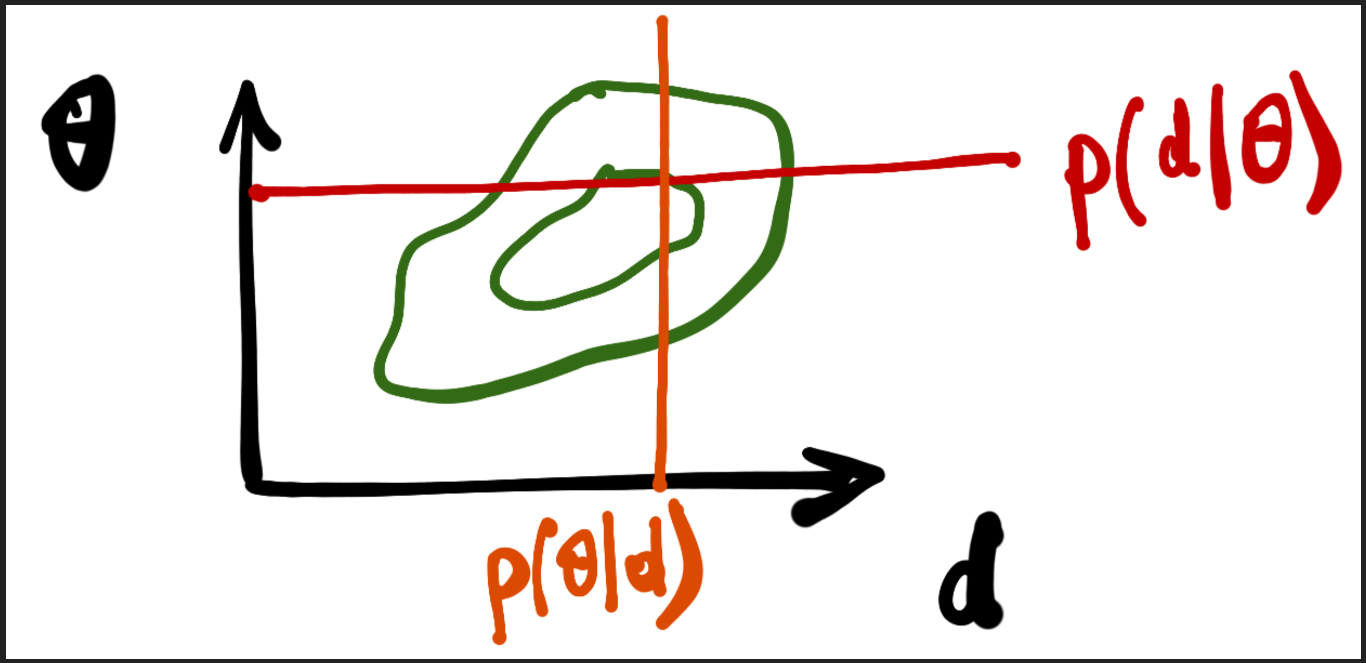


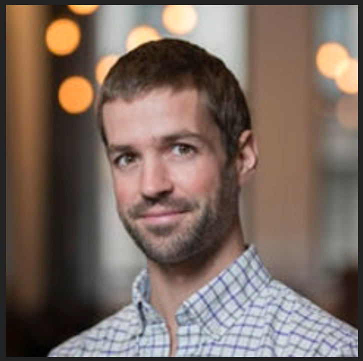


Justin Alsing

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get **p(θ | d)** directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint **p(θ , d)** (Alsing et al 2018)

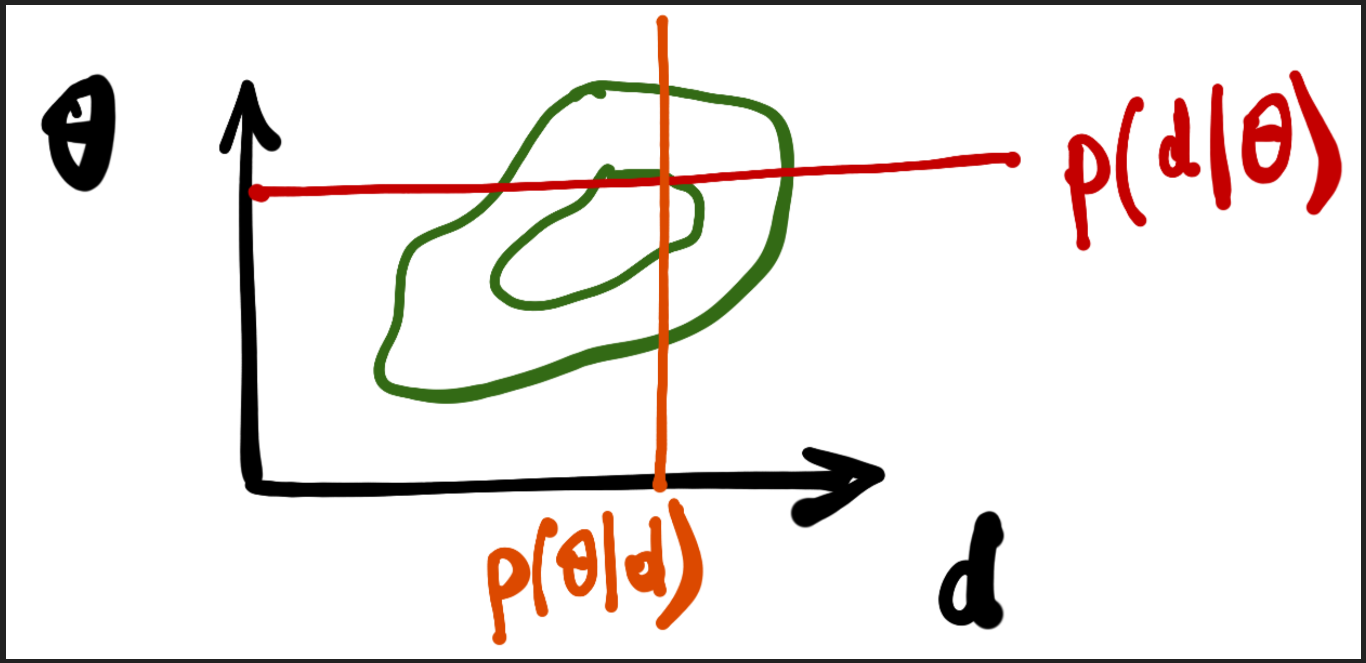


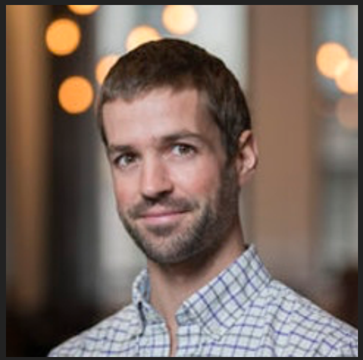


Justin Alsing

NG LIKELIHOODS III: LIKELIHOOD-FREE INFERENCE

- Try to get p(**θ** | **d**) directly. E.g. simulate with many different cosmological parameter sets, and keep those that match the data
- Naively, very inefficient
- 'Match': not everything, but match some summary statistics
- Try to approximate joint p(**θ** , **d**) (Alsing et al 2018)
  - Which statistics? Use MOPED again.





Justin Alsing

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

BAYESIAN HIERARCHICAL MODELS

BAYESIAN HIERARCHICAL MODELS

- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)

BAYESIAN HIERARCHICAL MODELS

- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)
- Mask is easy to include (infinite variance pixels)

BAYESIAN HIERARCHICAL MODELS

- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)
- Mask is easy to include (infinite variance pixels)
- Introduce latent variables e.g. the true map **s**:

BAYESIAN HIERARCHICAL MODELS

- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)
- Mask is easy to include (infinite variance pixels)
- Introduce latent variables e.g. the true map **s**:

$$p(\mathbf{d}|\theta) = \int p(\mathbf{d}, \mathbf{s}|\theta) d\mathbf{s}$$

BAYESIAN HIERARCHICAL MODELS

- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)
- Mask is easy to include (infinite variance pixels)
- Introduce latent variables e.g. the true map **s**:

$$p(\mathbf{d}|\theta) = \int p(\mathbf{d},\mathbf{s}|\theta)d\mathbf{s} = \int p(\mathbf{d}|\mathbf{s},\theta)\,p(\mathbf{s}|\theta)d\mathbf{s}$$

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

BAYESIAN HIERARCHICAL MODELS

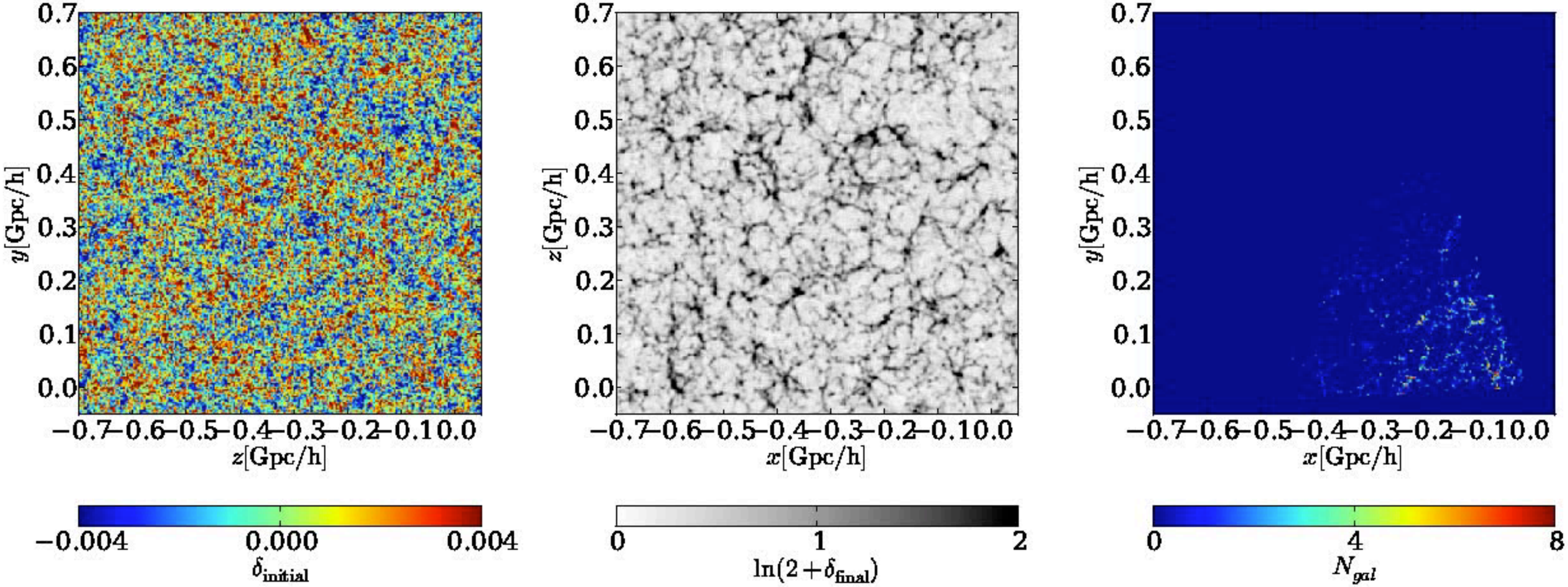
- This is the way to do it, if we can. Can add in many systematic effects, e.g. redshift distributions (Leistedt, Mortlock, Peiris 2016. See Boris' talk; also Alex Malz)
- Mask is easy to include (infinite variance pixels)
- Introduce latent variables e.g. the true map **s**:

$$p(\mathbf{d}|\theta) = \int p(\mathbf{d},\mathbf{s}|\theta)d\mathbf{s} = \int p(\mathbf{d}|\mathbf{s},\theta)\, p(\mathbf{s}|\theta)d\mathbf{s}$$

Jointly sample θ and **s**. It is a <u>very</u> high dimensional space ~10^6. Use HMC or Gibbs

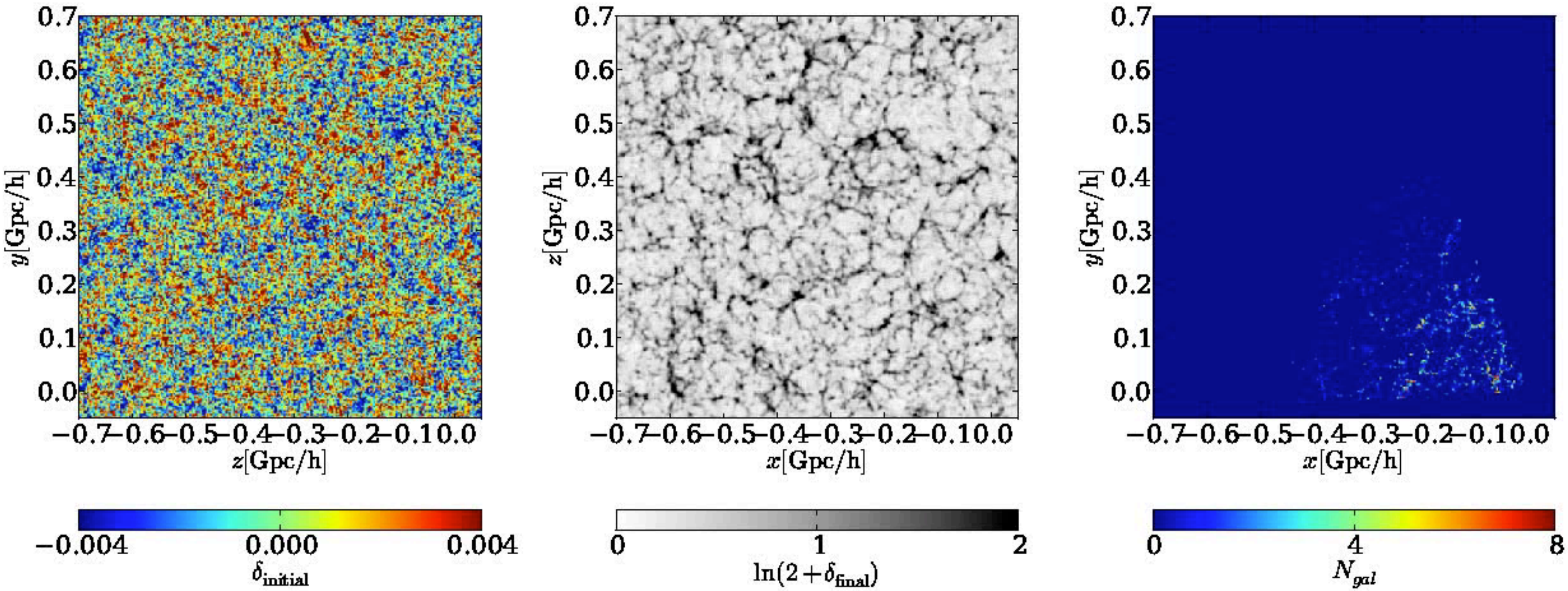
BORG, ALTAIR AND VARIANTS

- See Jens Jasche, Guilhem Lavaux and Doogesh Kodi Ramanagh's talks (also papers by Wandelt, Leclercq, Elsner, Anderes).

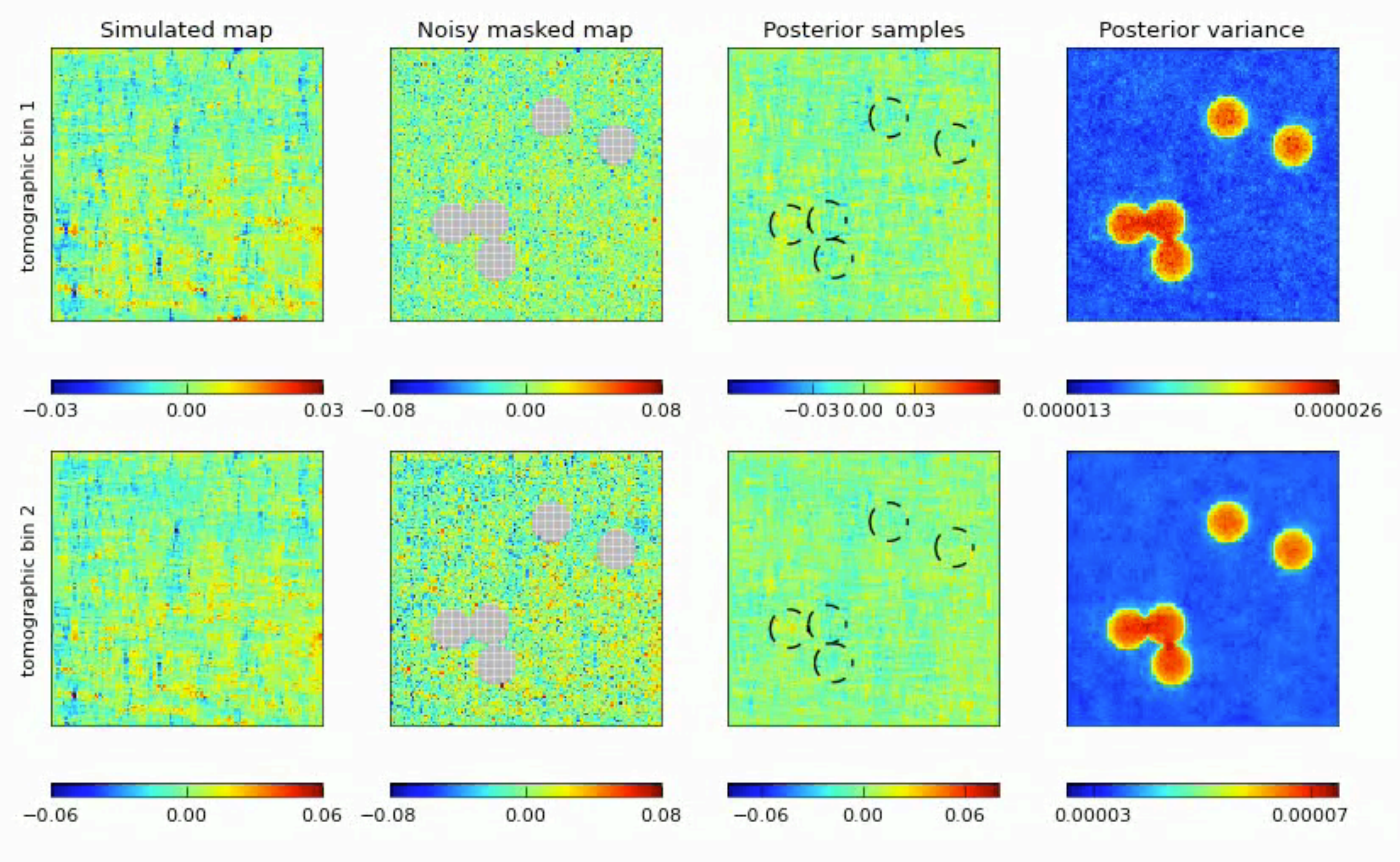


BORG, ALTAIR AND VARIANTS

- See Jens Jasche, Guilhem Lavaux and Doogesh Kodi Ramanagh's talks (also papers by Wandelt, Leclercq, Elsner, Anderes).

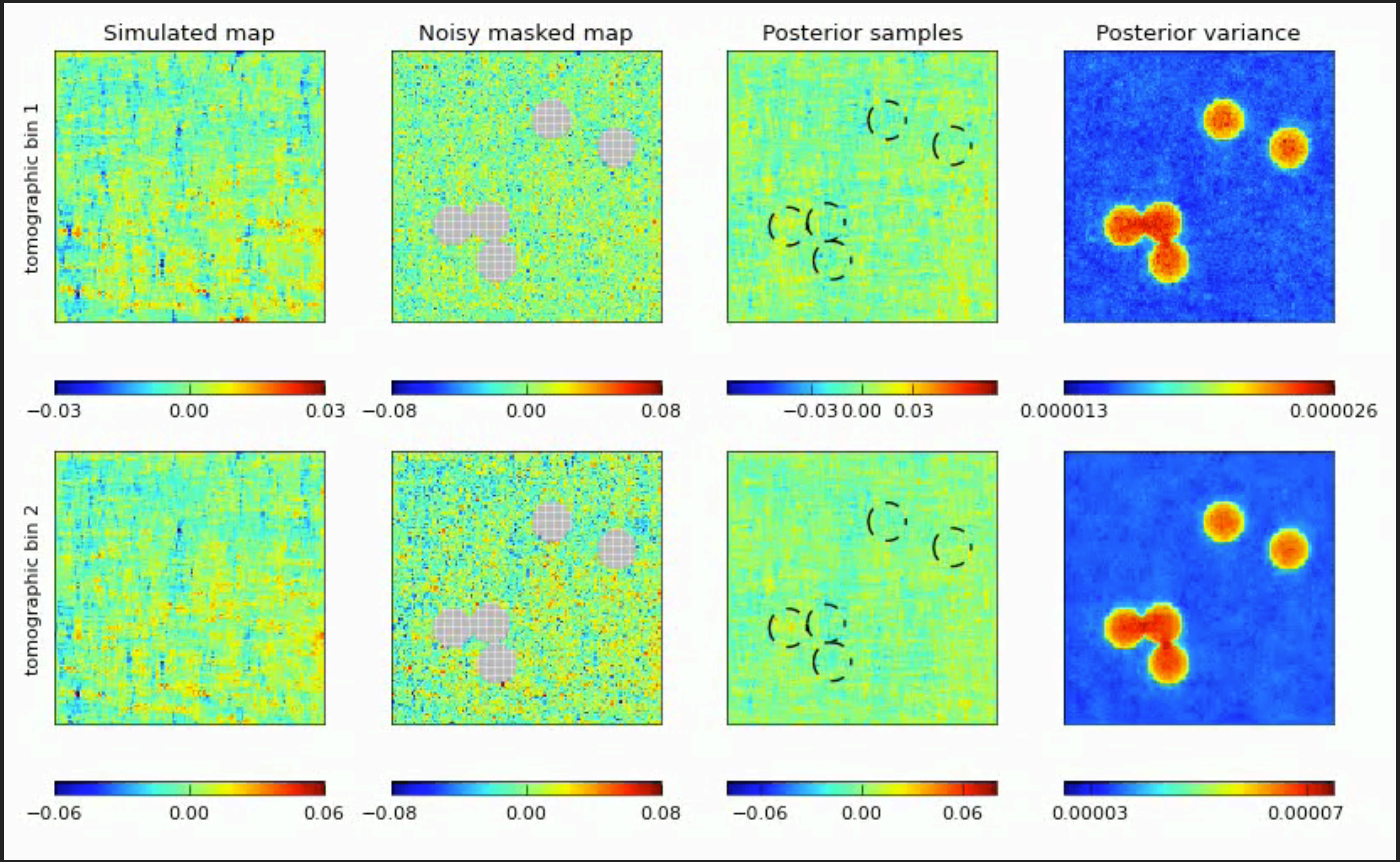


WEAK LENSING: SIMULATION



Alsing et al 2016

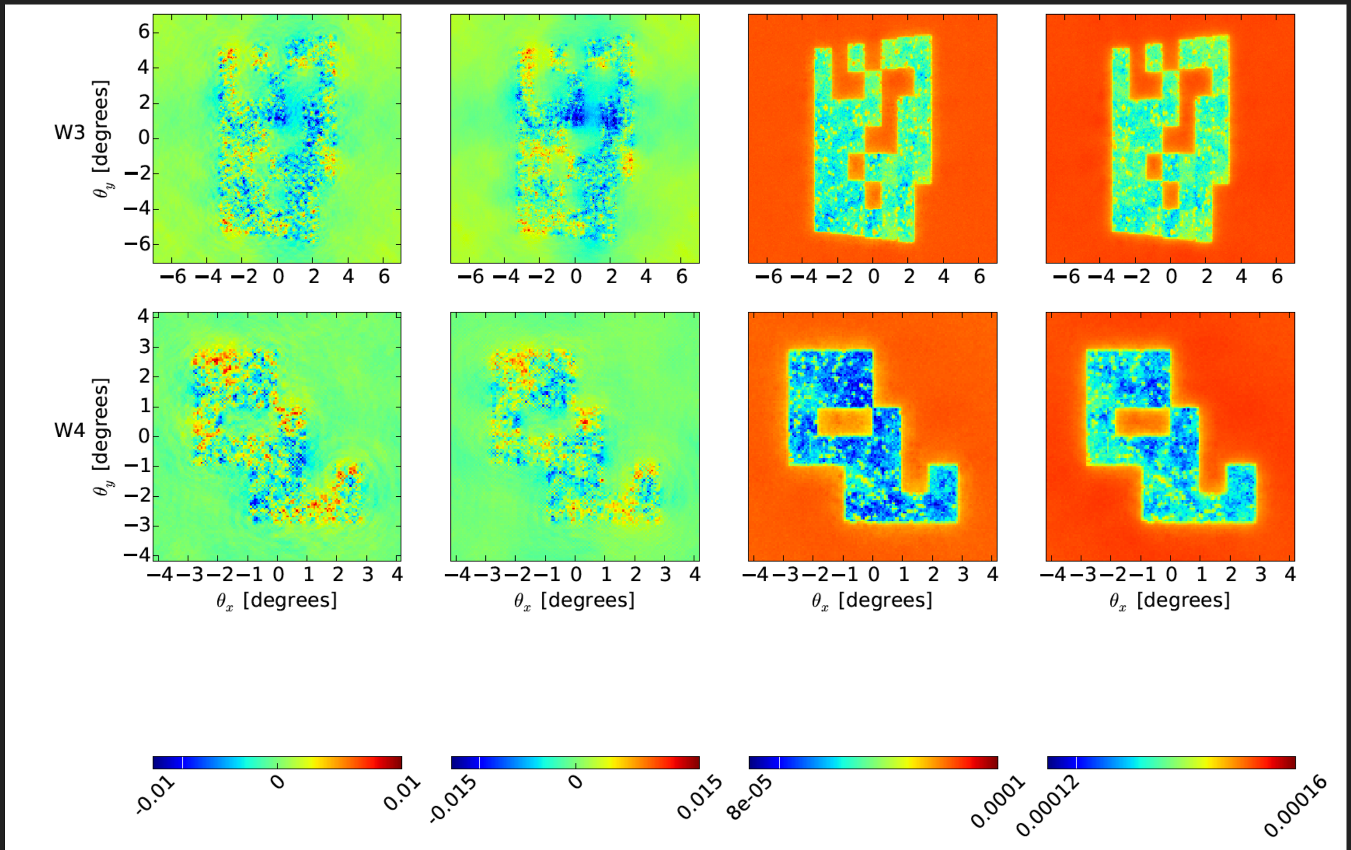
WEAK LENSING: SIMULATION

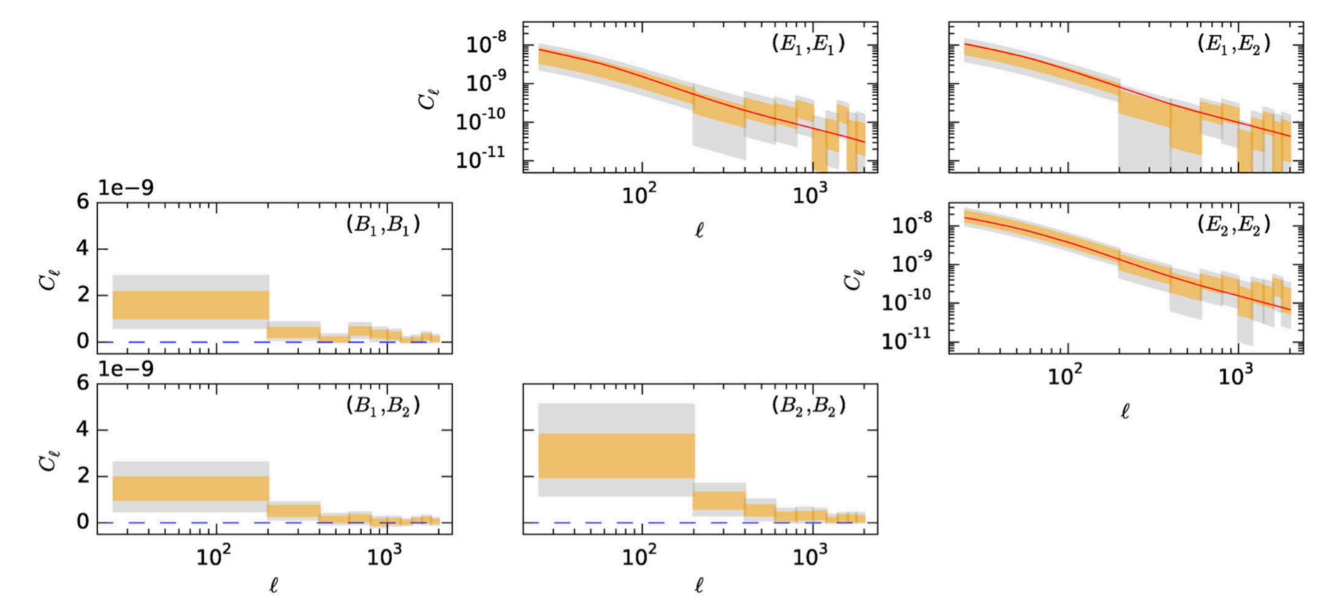


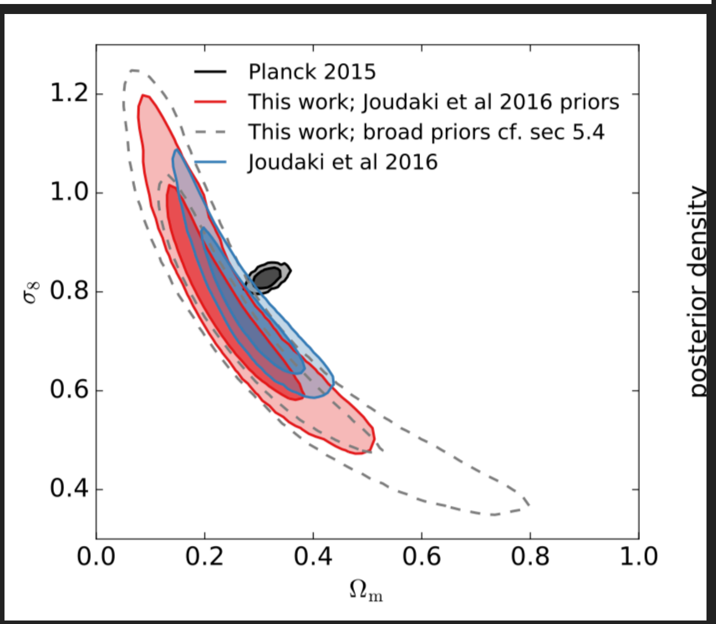
Alsing et al 2016

MASS MAPS AND POWER SPECTRA OR COSMOLOGICAL PARAMETERS

- Sample map and cosmology
- Marginalise over the maps to get cosmology
- Marginalise over cosmology to get maps



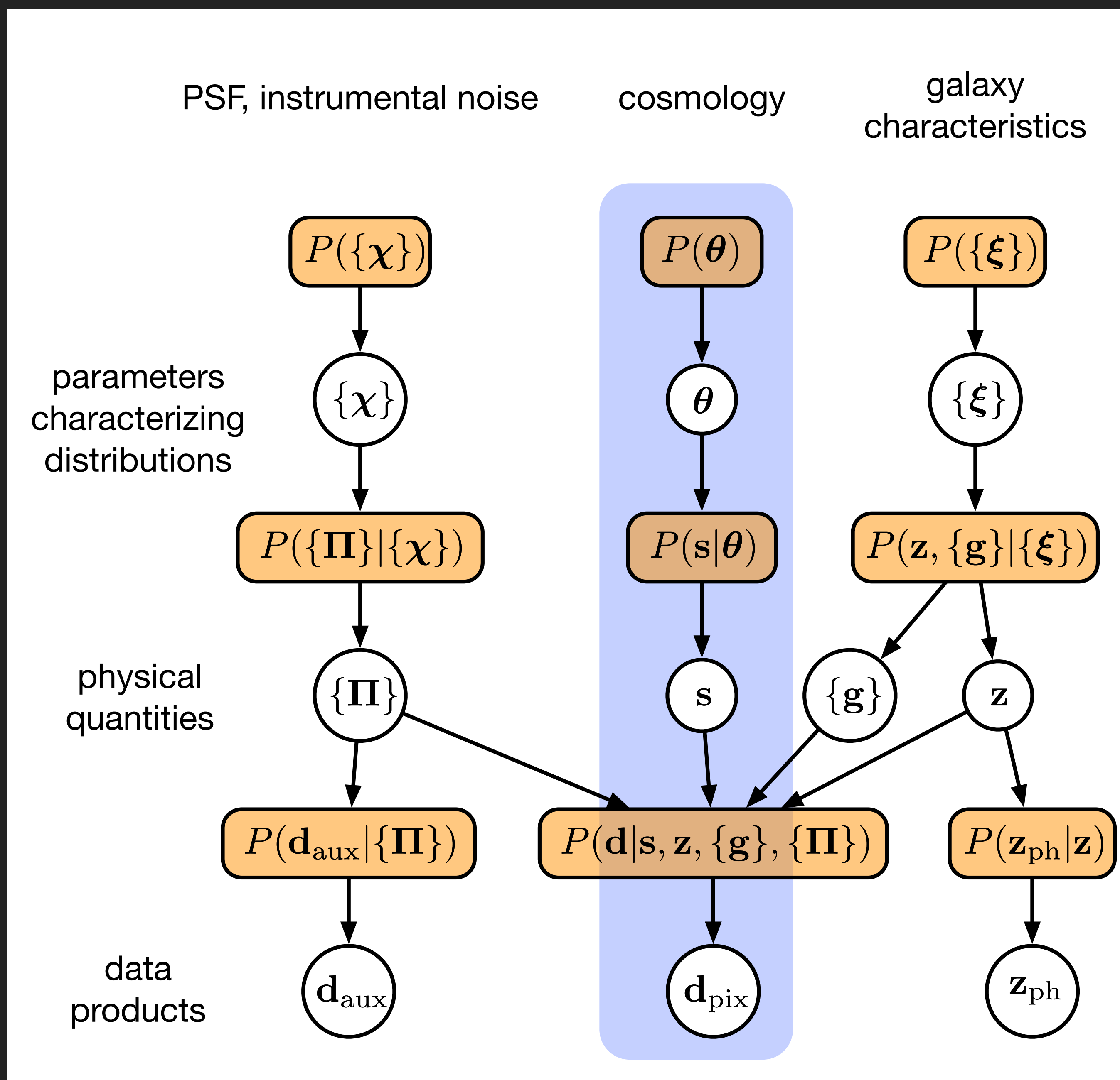




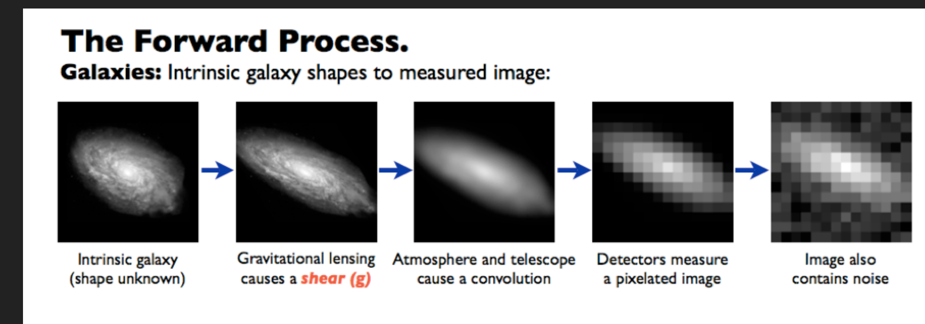
Alsing et al 2017

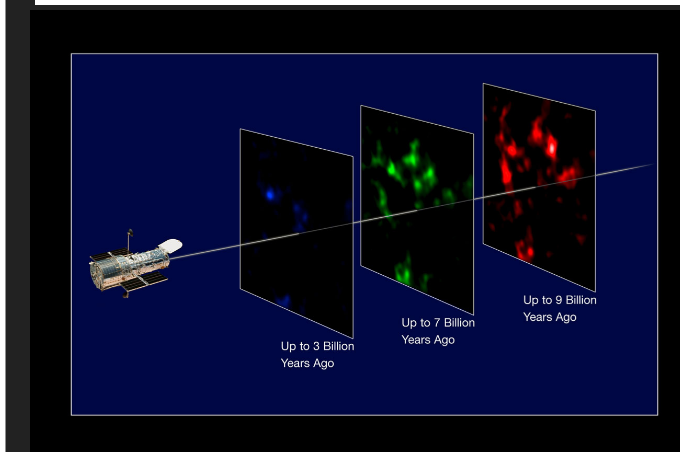
ICIC Imperial Centre for Inference & Cosmology

FULL BHM FOR WEAK LENSING

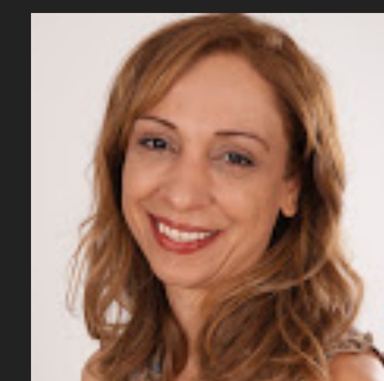


Alsing et al 2016

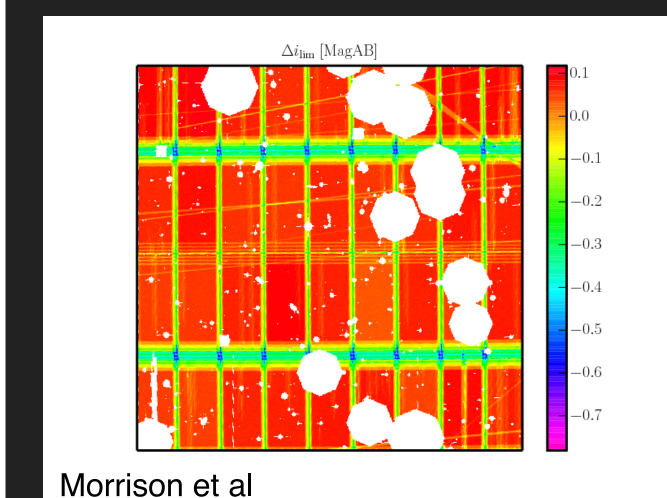




Credit:NASA,ESA;J.Hartlap,P.Simon,T.Schrabback



Malak Olamaie



Florent Leclercq

FEASIBILITY OF COSMIC SHEAR BHM

- Relatively simple BHMs with existing data can be analysed in ~1 day
- Scaling N log N (FFT), $N^{3/2}$ (Spherical harmonics); N = number of pixels
- Possible for LSST analysis
- Ideally sample from initial density field and evolve with 2LPT or ICE-COLA, for example
- Timescales then similar to galaxy clustering Bayesian analysis

CONCLUSIONS

- For gaussian distributed data, estimation of the covariance matrix will require data compression to avoid unfeasibly many simulations
- Assuming that data are gaussian-distributed will almost certainly not be good enough
- For likelihood-free parameter inference, or for approximating sampling distributions, massive data compression will also be necessary
- MOPED offers a way to do this without loss of information
- Bayesian Hierarchical Modelling is the principled solution to the analysis challenge

ICIC Imperial Centre for Inference & Cosmology

Imperial College London