Simulator Expansion for Likelihood-Free Inference

Prospects for Euclid

Florent Leclercq

www.florent-leclercq.eu

Imperial Centre for Inference and Cosmology
Imperial College London

Wolfgang Enzi, Alan Heavens,
Jens Jasche, Guilhem Lavaux,

and the Aquila Consortium
www.aquila-consortium.org

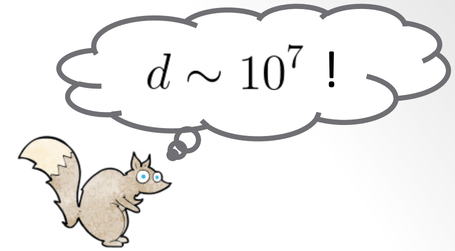
December 17th, 2019

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

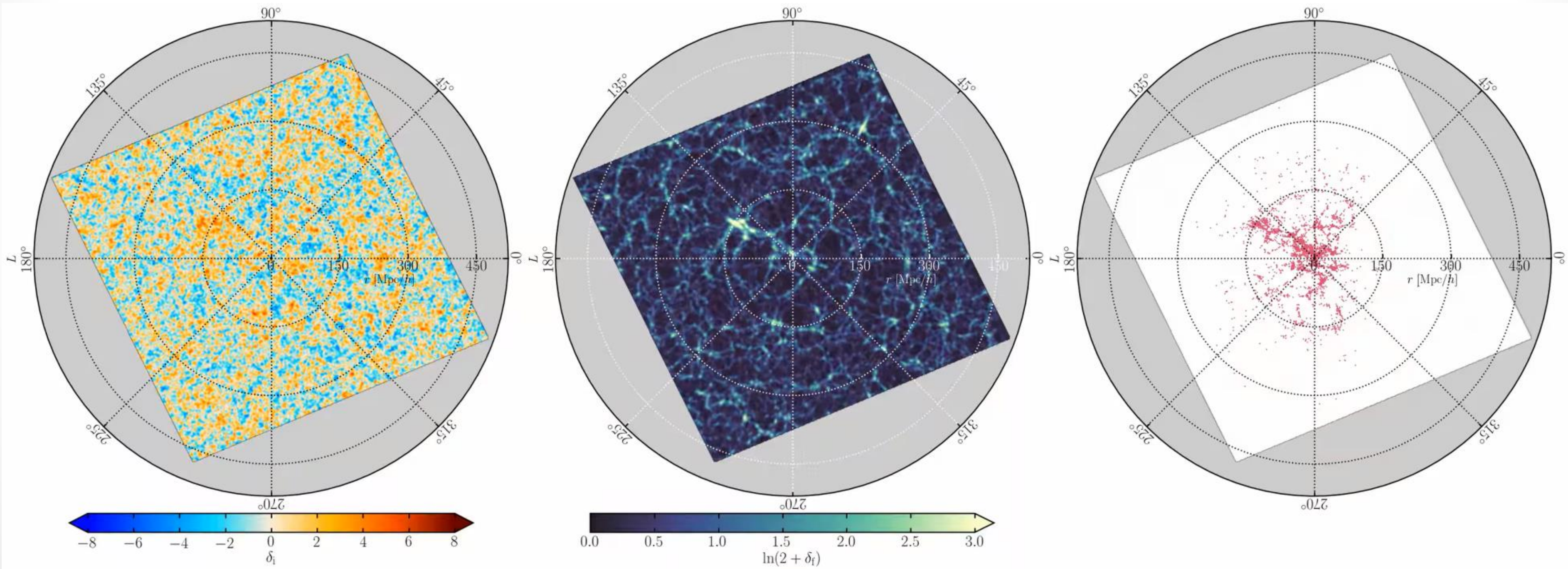
Vocabulary consideration:

What is the likelihood?



In cosmology, the (true?) likelihood should live at the level of the map of the CMB or LSS.
e.g. Wiener filtering for the CMB, BORG for the LSS (a 256^3-dimensional Poisson likelihood):

Initial conditions **Final conditions** **Observations**

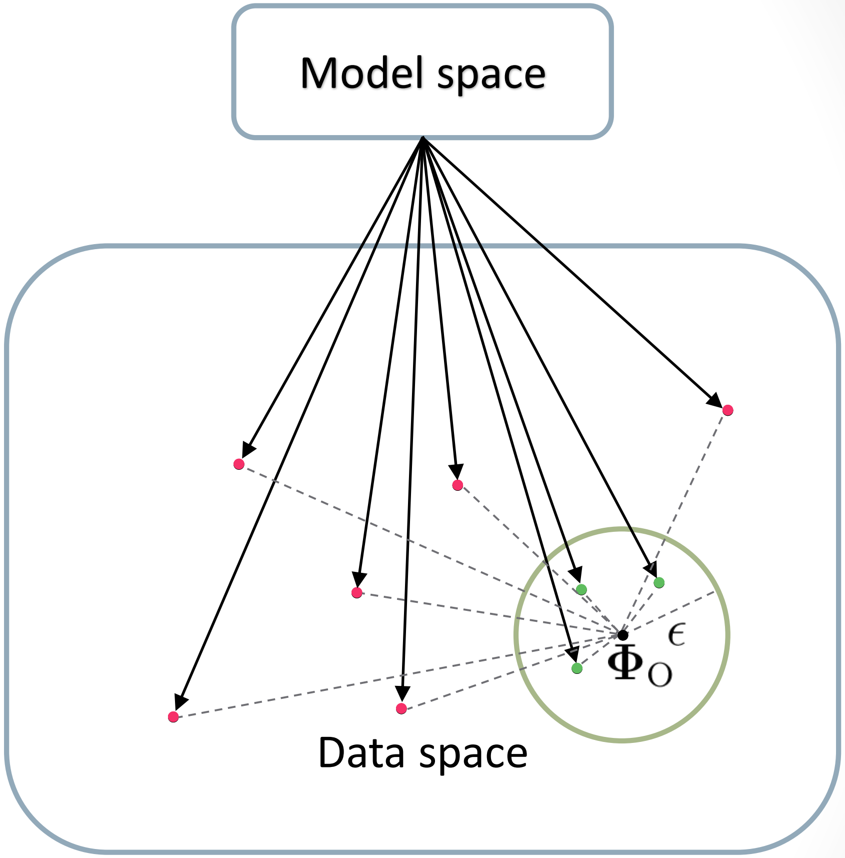


Jasche & Lavaux 2019, 1806.11117 – FL, Lavaux & Jasche, in prep.

Expert knowledge of the likelihood is needed to beat the curse of dimensionality:
conditionals/gradients of the likelihood are required by the samplers (Gibbs/Hamiltonian).

Likelihood-free rejection sampling (LFRS)

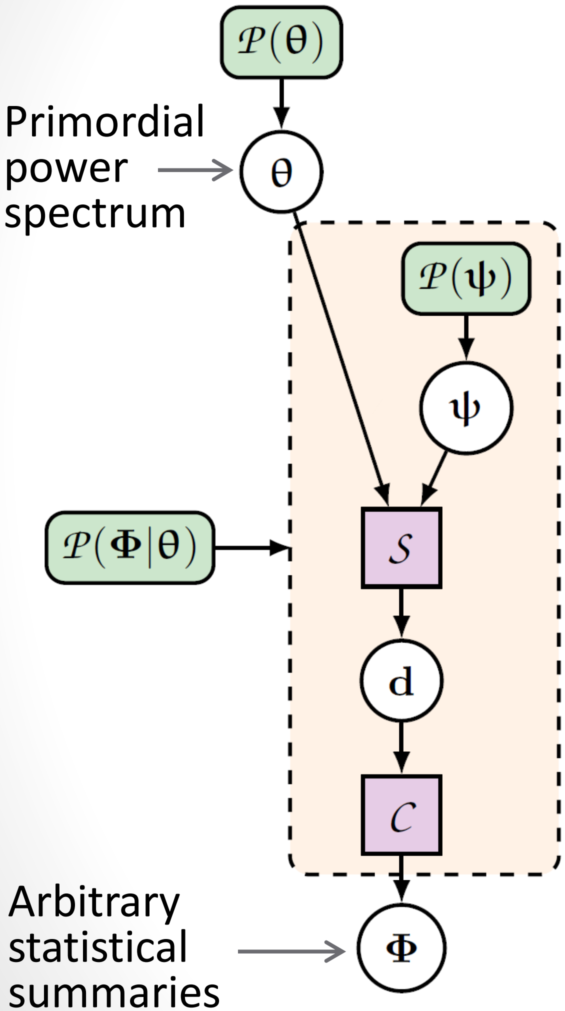
- Iterate many times:
  - Sample θ from a proposal distribution $q(\theta)$
  - Simulate Φ_θ using the black-box
  - Compute the distance $\Delta(\Phi_\theta, \Phi_\mathrm{O})$ between simulated and observed data
  - Retain θ if $\Delta(\Phi_\theta, \Phi_\mathrm{O}) \leq \epsilon$, otherwise reject



ϵ can be adaptively reduced (Population Monte Carlo)

Beyond LFRS: the SELFI approach

Simulator Expansion for Likelihood-Free Inference

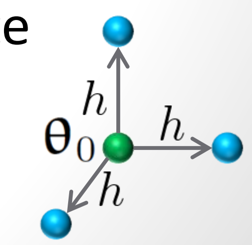


- We aim at inferring the primordial power spectrum, which contains (almost?) all of the information
- This requires doing LFI in $d = \mathcal{O}(100) - \mathcal{O}(1,000)$
- If we trust the results of earlier experiments, we can Taylor-expand the black-box around an expansion point θ_0:

$$\hat{\Phi}_\theta \approx \mathbf{f}_0 + \nabla \mathbf{f}_0 \cdot (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)^\intercal \cdot \mathbf{H} \cdot (\theta - \theta_0) + \dots$$

SELFI-2 (second-order): coming soon!

- Gradients, Hessian matrix, etc. of the black-box can be evaluated via finite differences in parameter space



SELFI-1: linearization of the black-box

- Linearization of the black-box:

  $$\hat{\boldsymbol{\Phi}}_{\boldsymbol{\theta}} \approx \mathbf{f}_0 + \nabla\mathbf{f}_0 \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

- Gaussian prior + Gaussian effective likelihood

➡ The posterior is Gaussian and analogous to a Wiener filter:

expansion point — observed summaries

$$\boldsymbol{\gamma} \equiv \boldsymbol{\theta}_0 + \boldsymbol{\Gamma}\,(\nabla\mathbf{f}_0)^\intercal\,\mathbf{C}_0^{-1}(\boldsymbol{\Phi}_\mathrm{O} - \mathbf{f}_0)$$

$$\boldsymbol{\Gamma} \equiv \left[(\nabla\mathbf{f}_0)^\intercal\,\mathbf{C}_0^{-1}\nabla\mathbf{f}_0 + \mathbf{S}^{-1}\right]^{-1}$$

prior covariance

covariance of summaries — gradient of the black-box

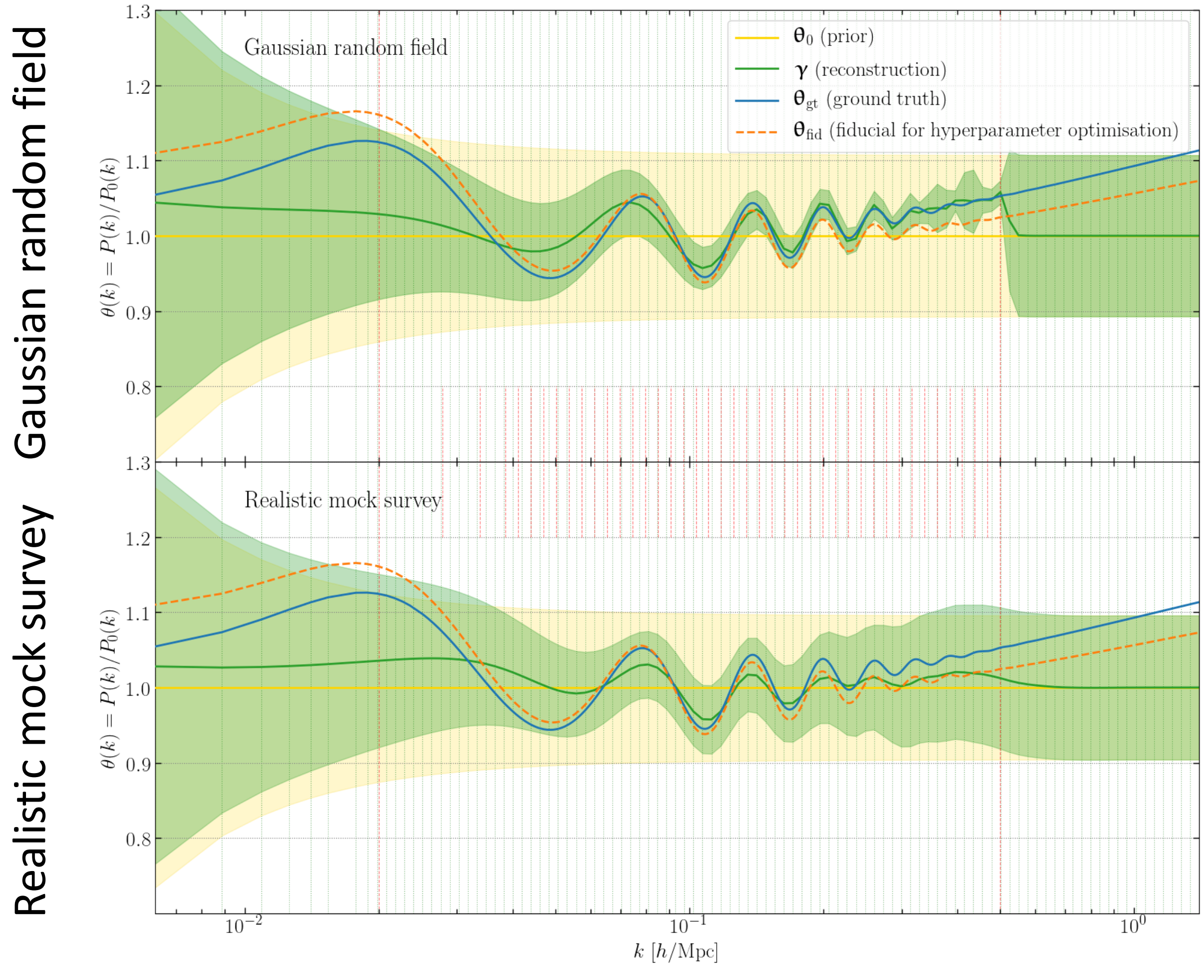
$\mathbf{f}_0$, $\mathbf{C}_0$ and $\nabla\mathbf{f}_0$ can be evaluated through simulations only.

The number of required simulations is fixed *a priori* (contrary to MCMC).

The workload is perfectly parallel.

FL, Enzi, Jasche & Heavens 2019, 1902.10149

SELFI + numerical model: Proof-of-concept



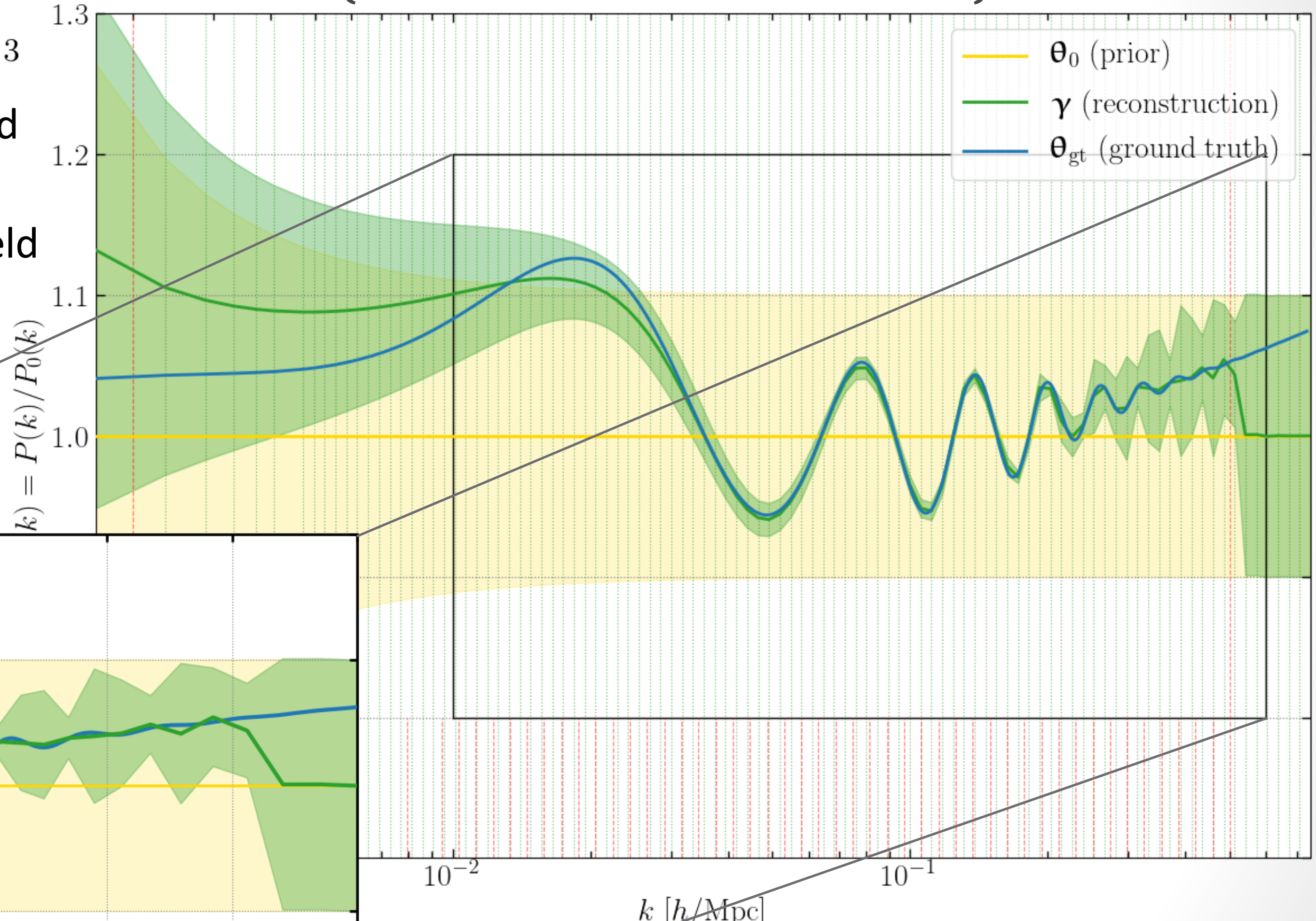
100 parameters are simultaneously inferred from a black-box data model

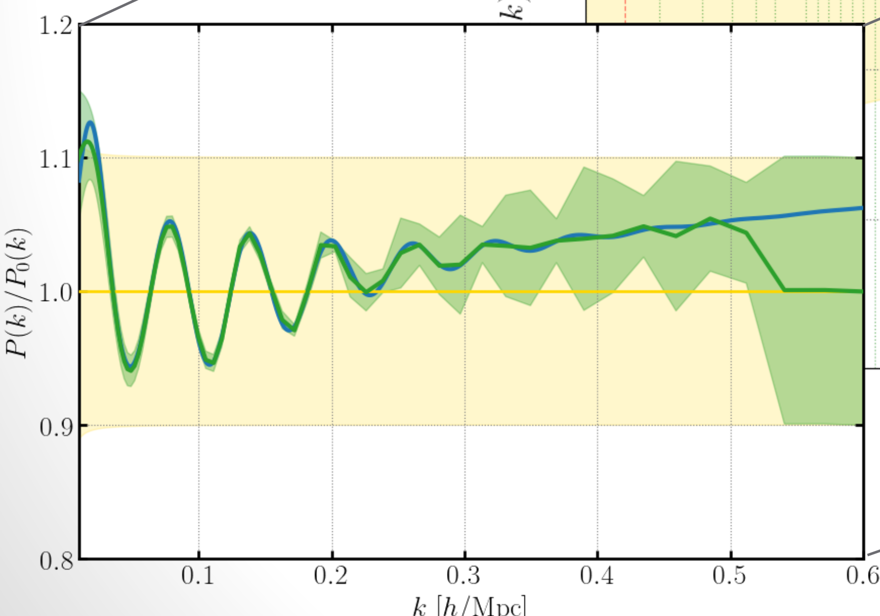
1 (Gpc/h)^3 only! Much more potential for upcoming data…

FL, Enzi, Jasche & Heavens 2019, 1902.10149

SELFI-1 Euclid forecast (cosmic variance limit)

- $V = (3780\ \mathrm{Mpc}/h)^3$ (volume of the Euclid flagship simulation)
- Gaussian random field data model
- 6,060 simulations



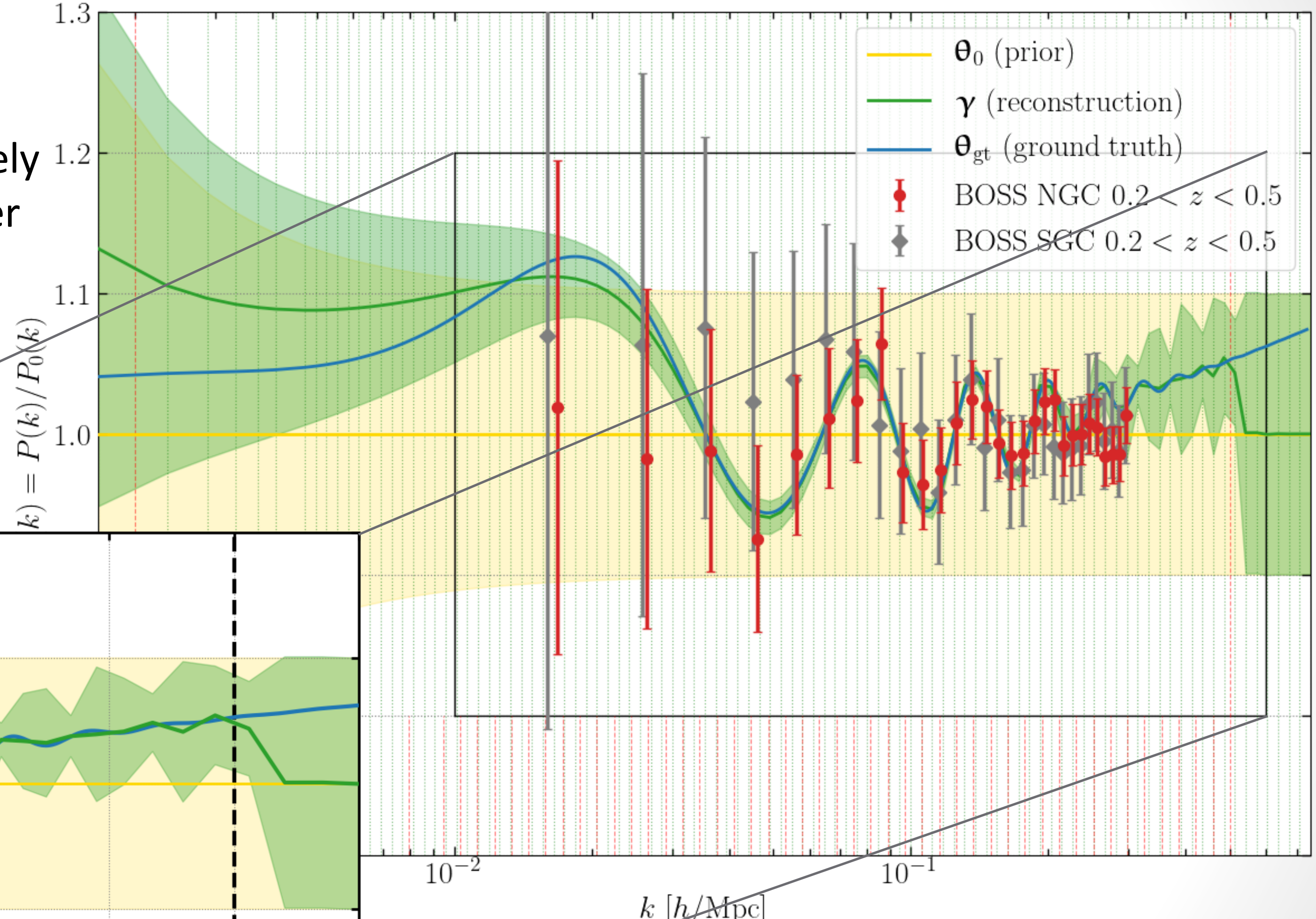


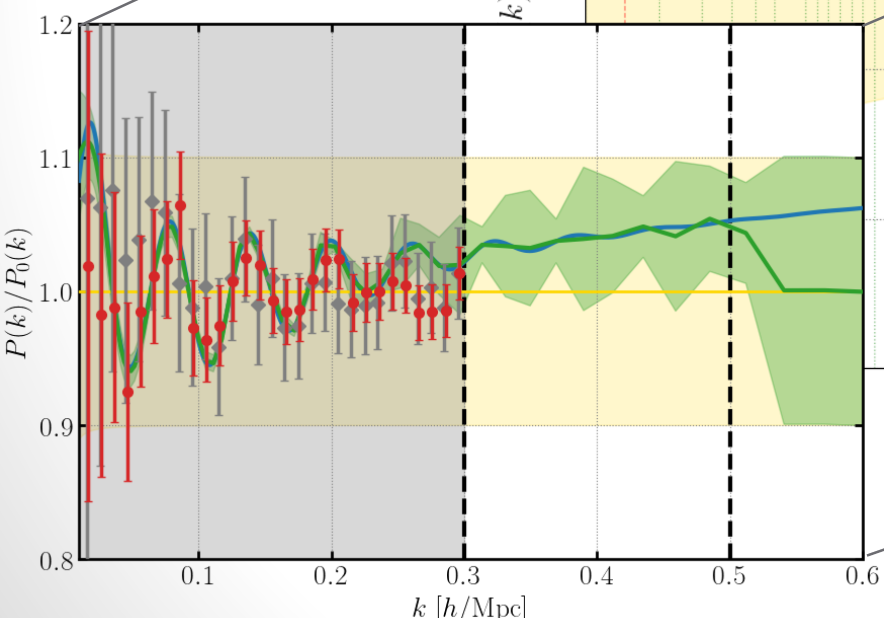
$$\boldsymbol{\gamma} \equiv \boldsymbol{\theta}_0 + \boldsymbol{\Gamma} \left(\nabla \mathbf{f}_0\right)^\intercal \mathbf{C}_0^{-1} (\boldsymbol{\Phi}_\mathrm{O} - \mathbf{f}_0)$$

SELFI-1 Euclid versus BOSS

Data points from
Beutler *et al.* 2016, 1607.03149

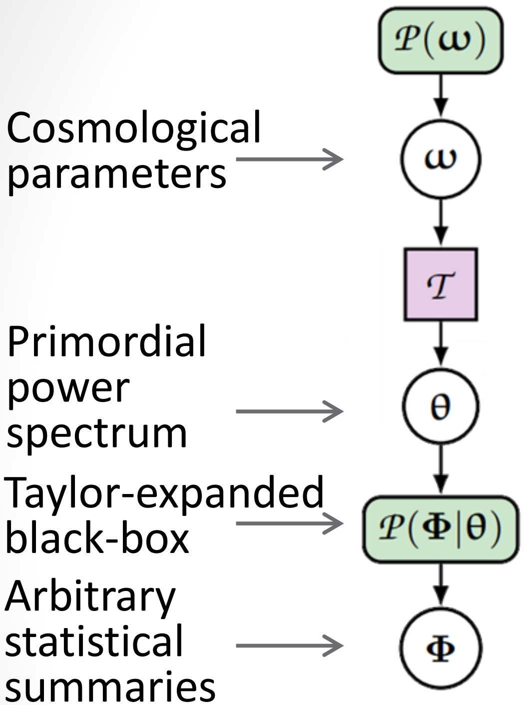
- Numerical data models allow to safely use the galaxy power spectrum up to at least
  $k \gtrsim 0.5\ h/\mathrm{Mpc}$



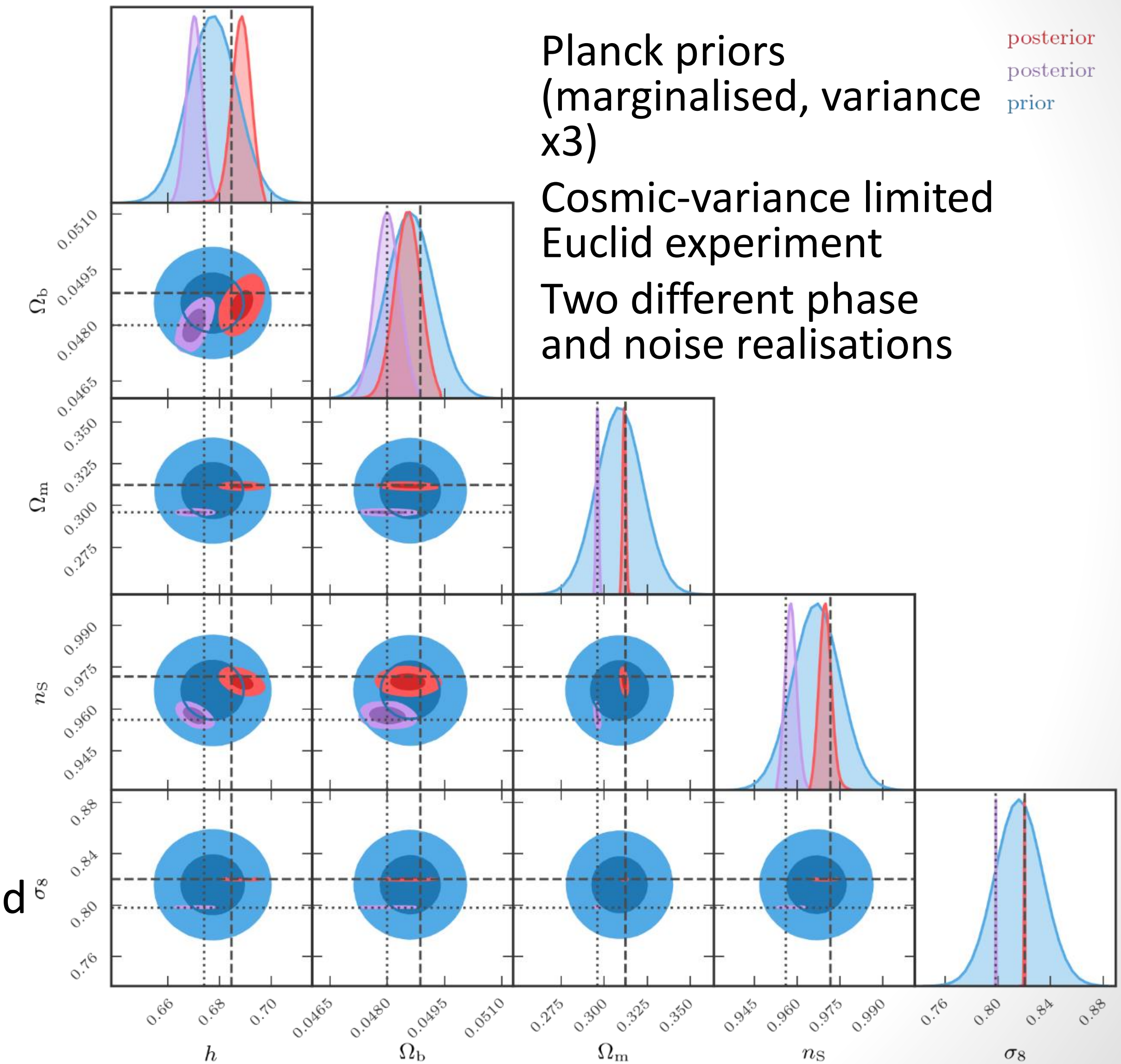


$N_{\mathrm{modes}} \propto k^3$: 5 times more modes are used in the analysis

From primordial power spectrum to cosmology



- Robust inference of cosmological parameters can be easily performed *a posteriori* once the linearized data model is learnt



Planck priors (marginalised, variance x3)

Cosmic-variance limited Euclid experiment

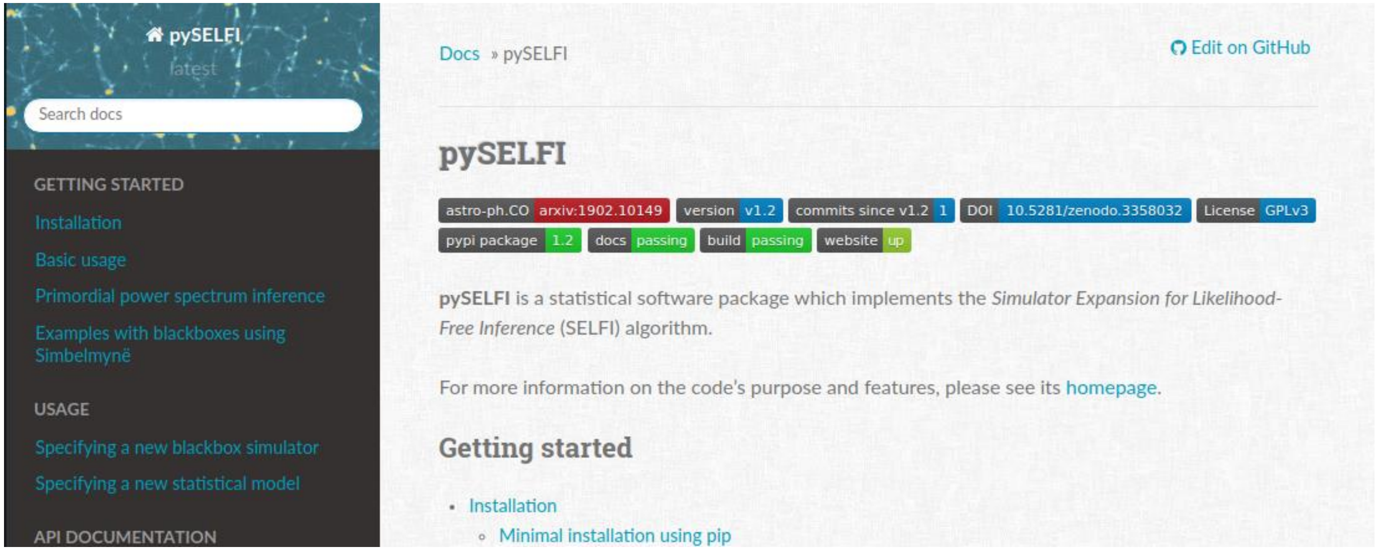
Two different phase and noise realisations

FL, Enzi, Jasche & Heavens 2019, 1902.10149

pySELFI is publicly available

- Code homepage: http://pyselfi.florent-leclercq.eu/
- Source on GitHub: https://github.com/florent-leclercq/pyselfi/
- Documentation on ReadtheDocs: https://pyselfi.readthedocs.io/en/latest/

(with templates to use your own black-box)




```
pip install pyselfi
```

Concluding thoughts

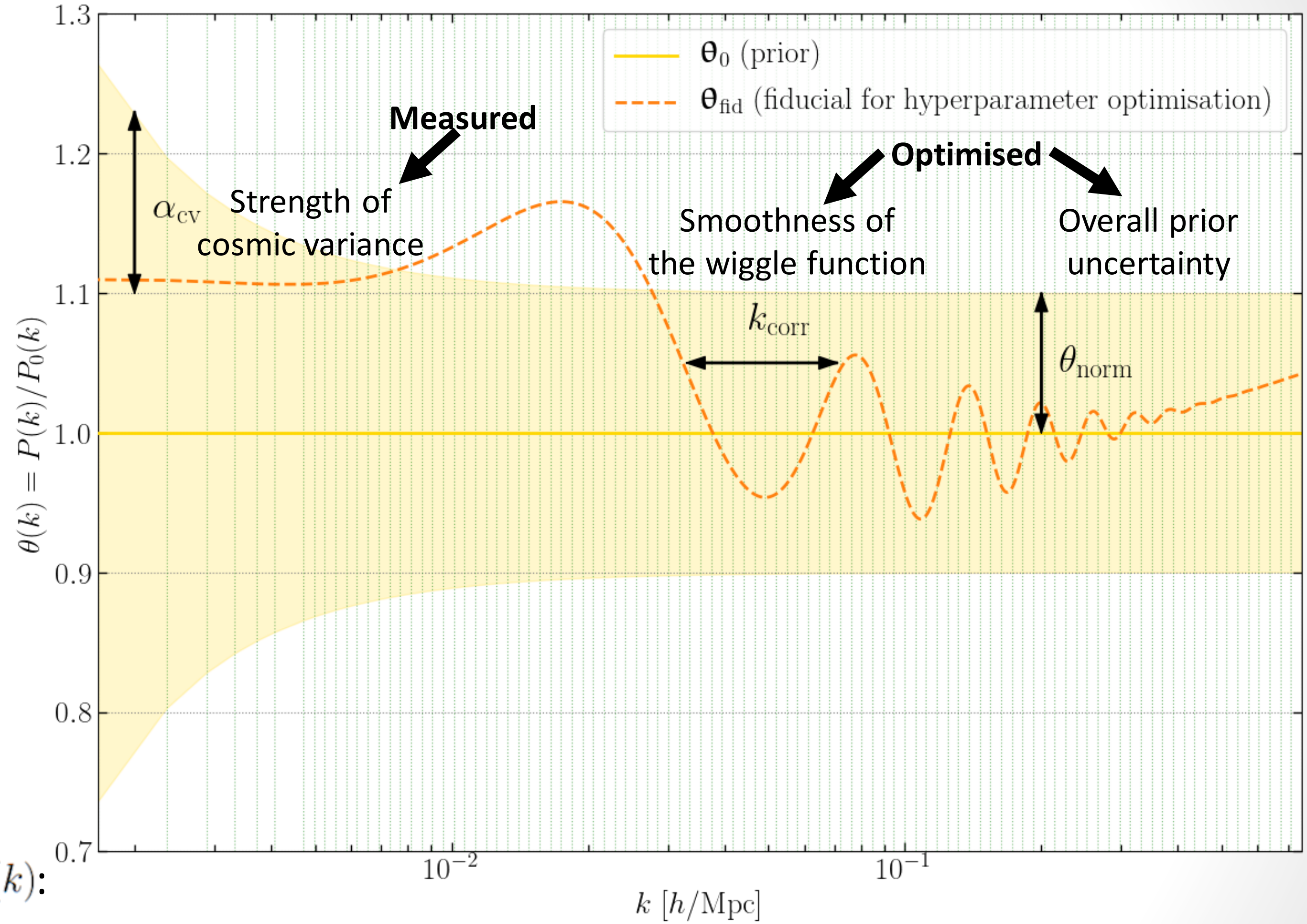
- Goal: developing an algorithm for targeted questions, allowing the use of simulators including all relevant physical and observational effects.
- Bayesian analyses of galaxy surveys with fully non-linear numerical black-box models is not an impossible task!
- SELFI allows inference of the primordial power spectrum and cosmological parameters.

Additional Slides

A prior for the primordial power spectrum

Assumptions:

1. the power spectrum is Gaussian-distributed
2. it is strongly constrained to live close to P_0,
3. it is a smooth function of wavenumber,
4. and the power spectrum P_0 is subject to cosmic variance



Prior for $\theta(k) = P(k)/P_0(k)$:

Mean: $\boldsymbol{\theta}_0 = \mathbf{1}_{\mathbb{R}^S}$ (without baryon acoustic oscillations wiggles)

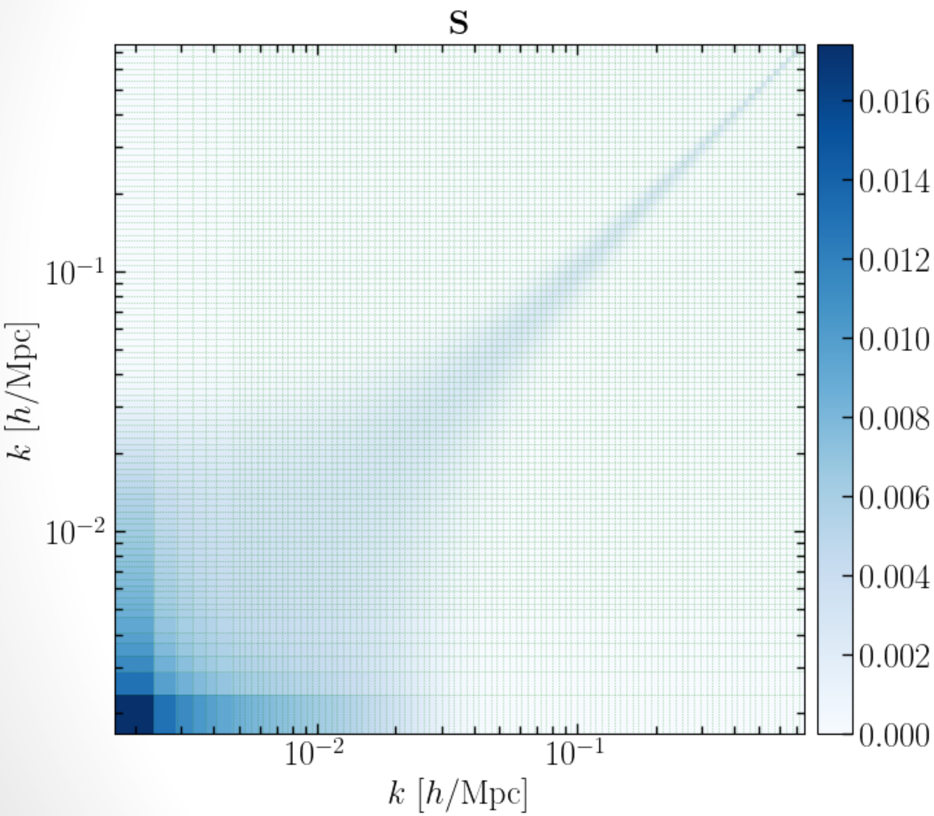
Covariance: $\mathbf{S} \equiv \theta^2_{\mathrm{norm}} \mathbf{u}\mathbf{u}^\intercal \circ \mathbf{K}$ $\quad (\mathbf{K})_{ss'} \equiv \exp\left[-\frac{1}{2}\left(\frac{k_s - k_{s'}}{k_{\mathrm{corr}}}\right)^2\right]$ $\quad (\mathbf{u})_s \equiv 1 + \sigma_s = 1 + \frac{\alpha_{\mathrm{cv}}}{k_s^{3/2}}$

FL, Enzi, Jasche & Heavens 2019, 1902.10149

Uncertainty quantification

$$\mathbf{\Gamma} \equiv \left[(\nabla \mathbf{f}_0)^\intercal \, \mathbf{C}_0^{-1} \nabla \mathbf{f}_0 + \mathbf{S}^{-1} \right]^{-1}$$

Prior covariance matrix



Posterior covariance matrix