Bayesian inference with black-box cosmological models

Florent Leclercq

www.florent-leclercq.eu

Imperial Centre for Inference and Cosmology
Imperial College London

Wolfgang Enzi, Jens Jasche, Alan Heavens,
and the Aquila Consortium
www.aquila-consortium.org

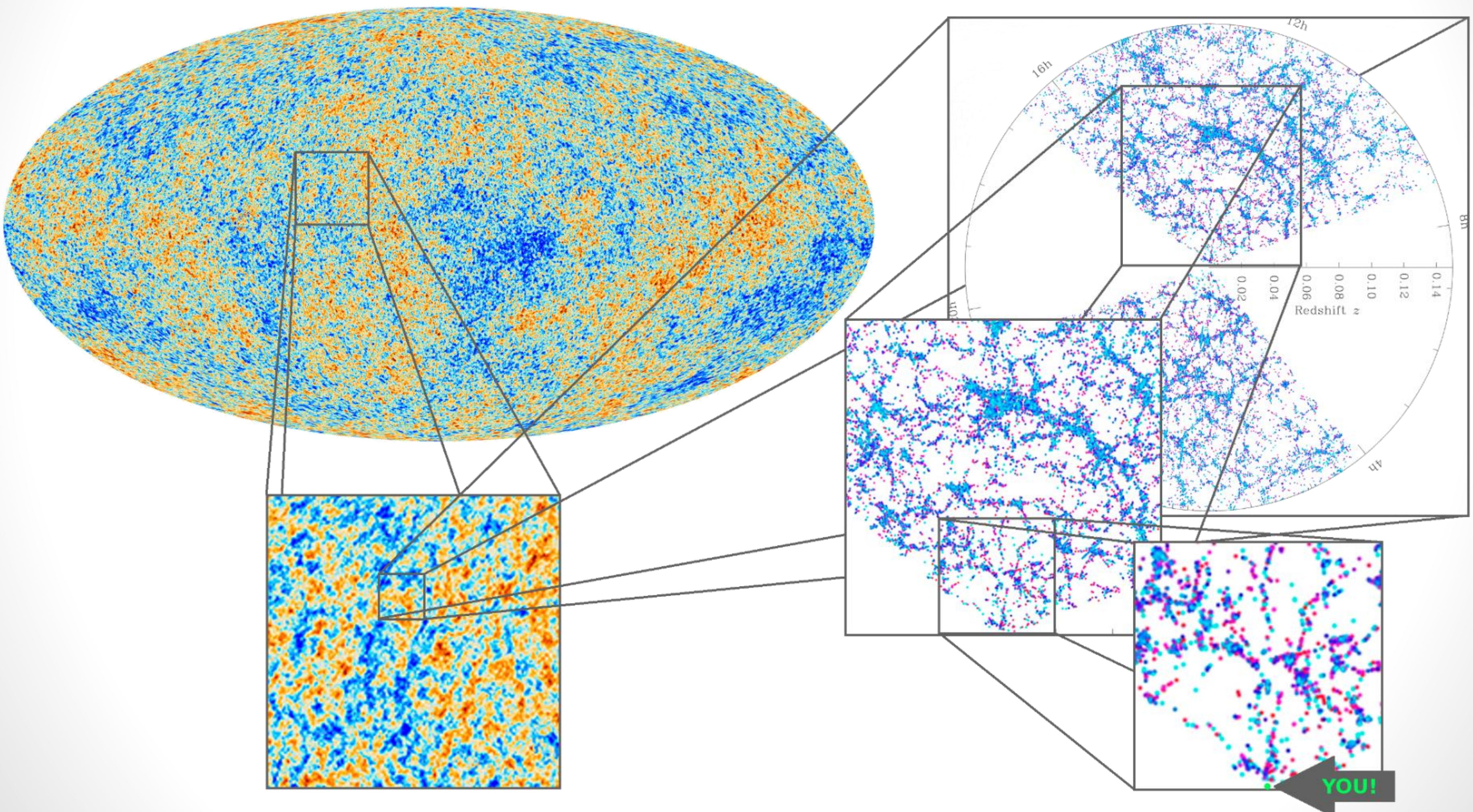
May 30th, 2019

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

The big picture: the Universe is highly structured

You are here. Make the best of it…



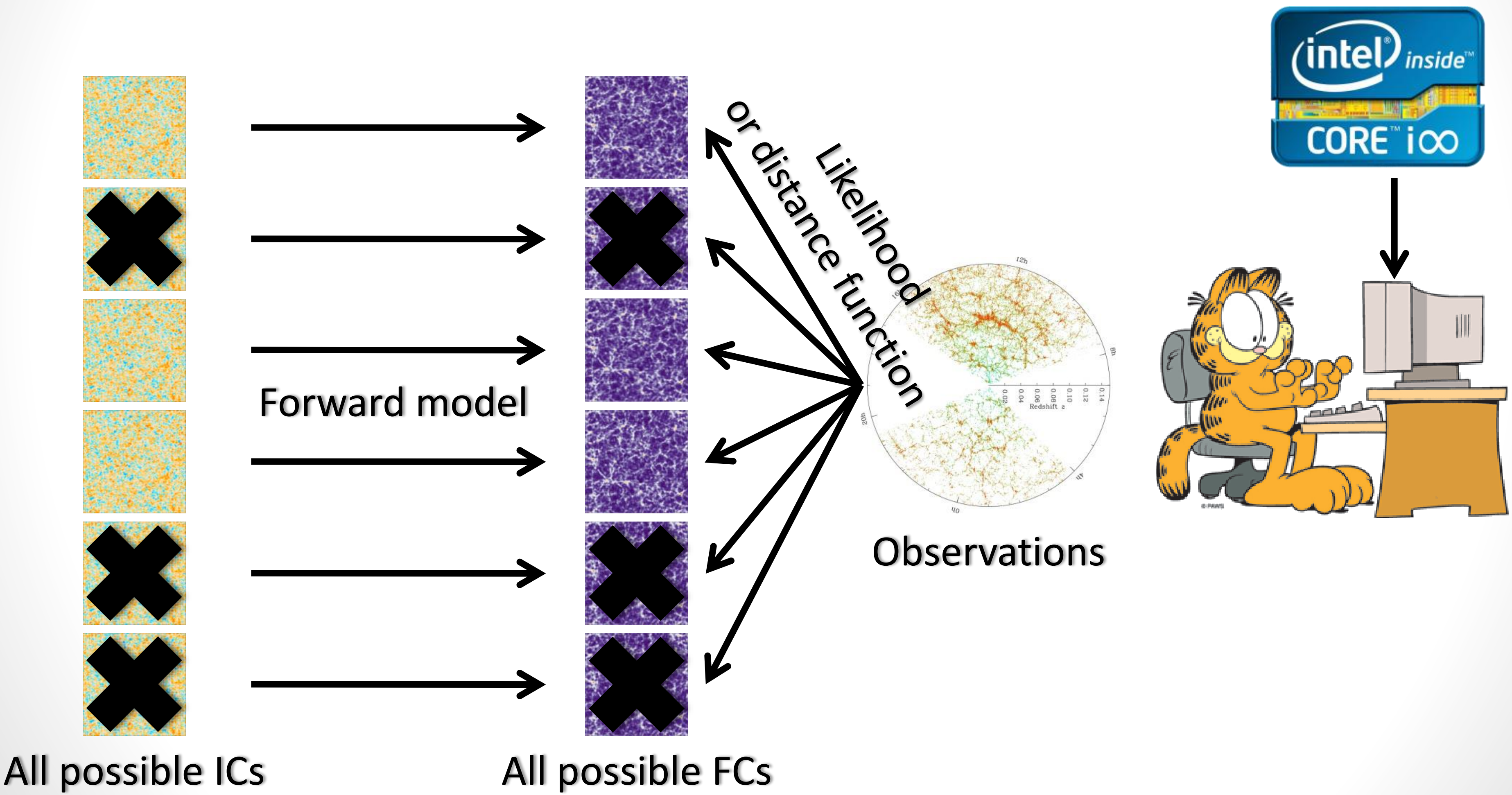
Planck collaboration (2013-2015)   M. Blanton and the Sloan Digital Sky Survey (2010-2013)

What we want to know from the large-scale structure

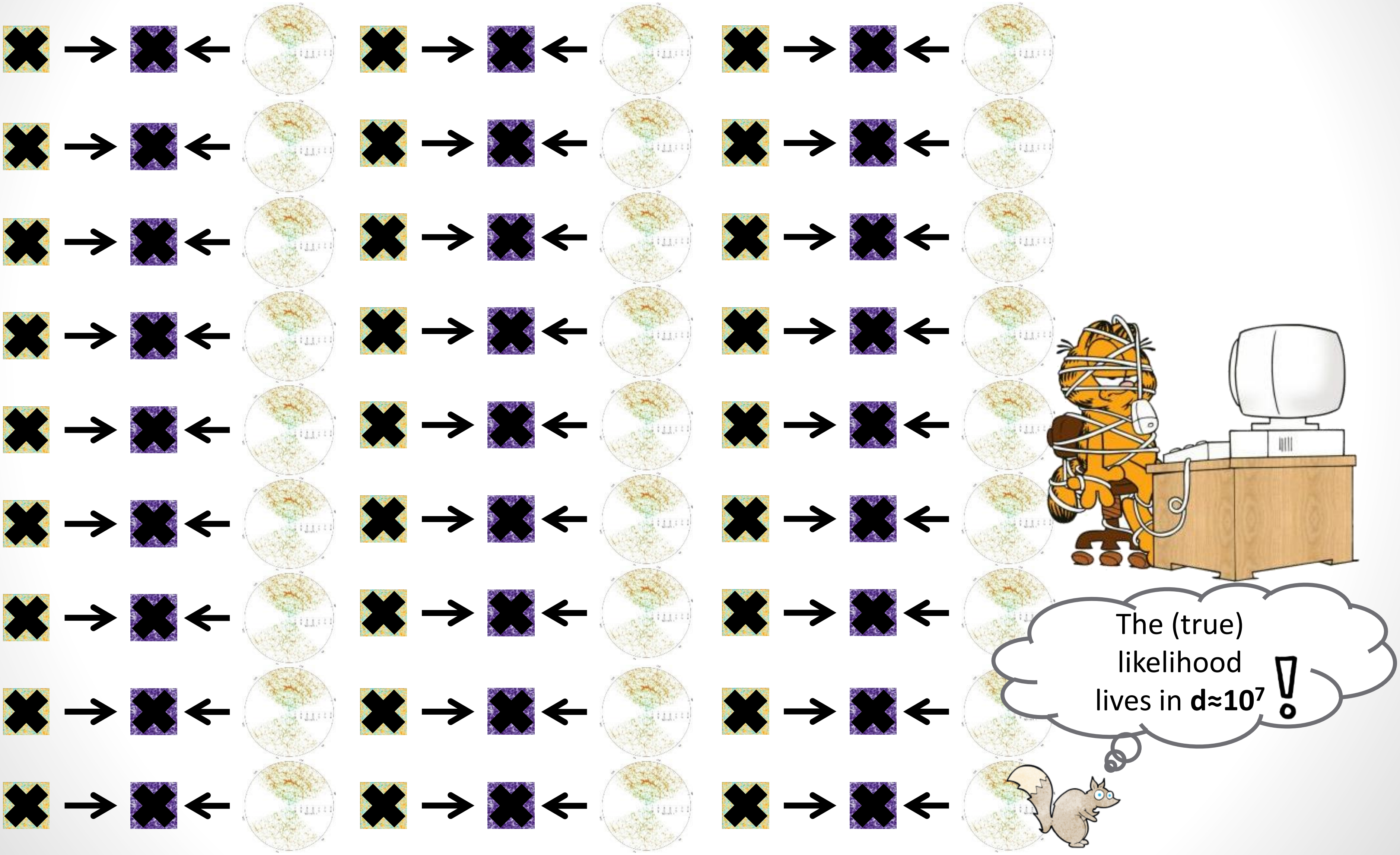
The LSS is a vast source of knowledge:

- Cosmology:
  - ΛCDM : cosmological parameters and tests against alternatives,
  - Physical nature of the dark components,
  - Neutrinos : number and masses,
  - Geometry of the Universe,
  - Tests of General Relativity,
  - Initial conditions and link to high energy physics
- Astrophysics: galaxy formation and evolution as a function of their environment
  - Galaxy properties (colours, chemical composition, shapes),
  - Intrinsic alignments, intrinsic size-magnitude correlations

Bayesian forward modelling: the ideal scenario



Bayesian forward modelling: the challenge



Likelihood-based solution: BORG

Bayesian Origin Reconstruction from Galaxies

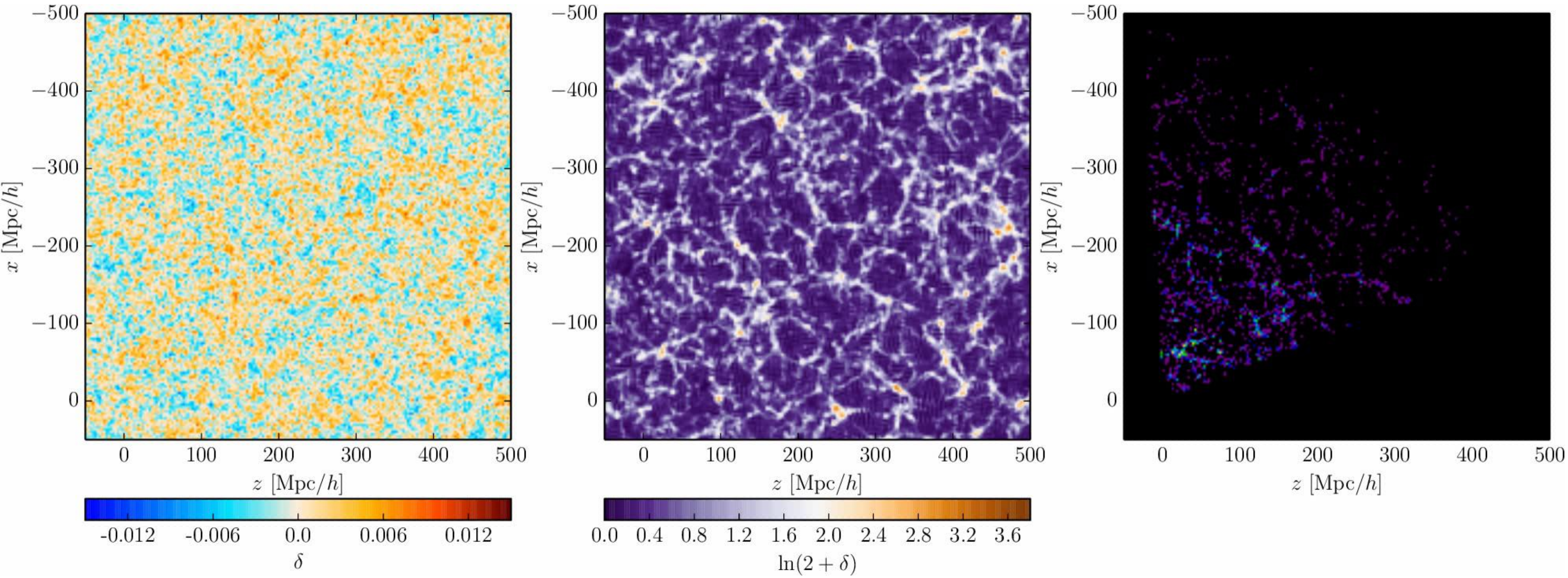
Likelihood-based solution:
Exact statistical analysis
Approximate data model

Data assimilation

?

Likelihood-based solution: BORG at work

www.aquila-consortium.org/



Initial conditions | Final conditions | Observations

334,074 galaxies, ≈ 17 million parameters, 3 TB of primary data products,
12,000 samples, ≈ 250,000 data model evaluations, 10 months on 32 cores

All data products are publicly available:

Jasche, FL & Wandelt 2015, arXiv:1409.6308

https://github.com/florent-leclercq/borg_sdss_data_release, doi: 10.5281/zenodo.1455729

Likelihood-free solution: BOLFI & SELFI

Bayesian Optimisation for Likelihood-Free Inference
Simulator Expansion for Likelihood-Free Inference

Likelihood-based solution:
Exact statistical analysis
Approximate data model

Data assimilation

?

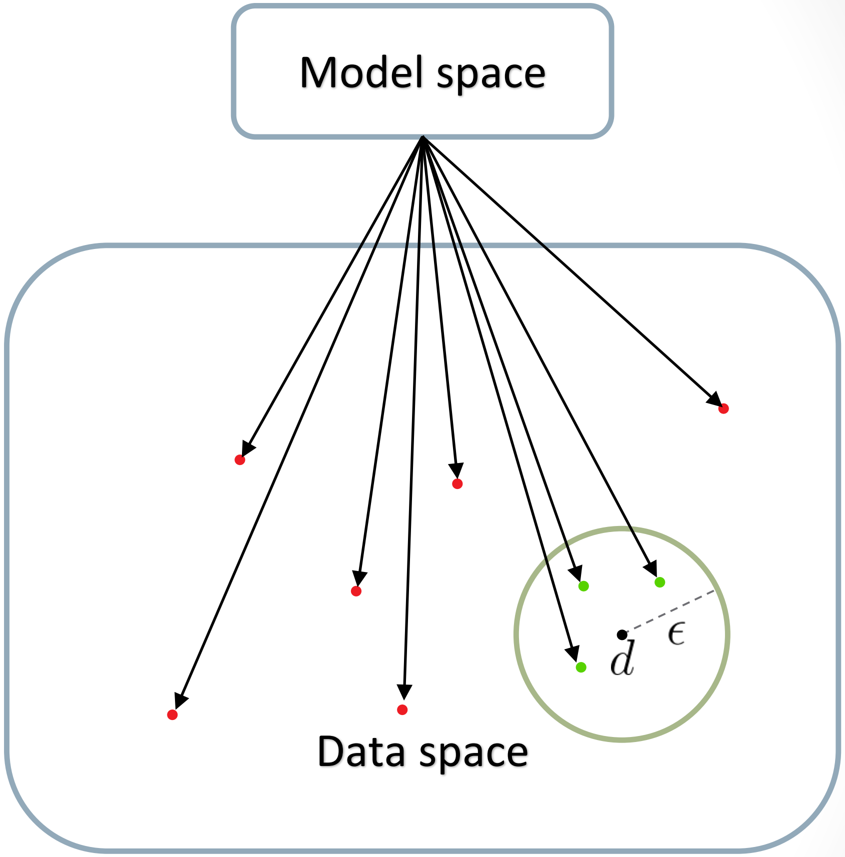
Likelihood-free solution:
Approximate statistical analysis
Arbitrary data model

Generative inference

Likelihood-free rejection sampling (LFRS)

- Iterate many times:
  - Sample θ from a proposal distribution $q(\theta)$
  - Simulate $\tilde{d}(\theta)$ according to the data model
  - Compute distance $\mathrm{d}(\tilde{d}(\theta), d)$ between simulated and observed data
  - Retain θ if $\mathrm{d}(\tilde{d}(\theta), d) \leq \epsilon$, otherwise reject
- Effective likelihood approximation:

$$L(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(\mathrm{d}(\tilde{d}(\theta), d) \leq \epsilon\right)$$



ϵ can be adaptively reduced (Population Monte Carlo)

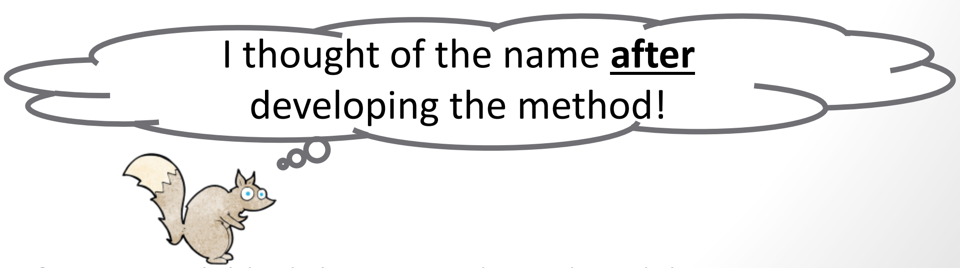
Beyond LFRS: two scenarios

The "number of simulations" route:

- Specific cosmological models ($d \lesssim 10$), general exploration of parameter space
- Density Estimation for Likelihood-Free Inference (DELFI)
  Papamakarios & Murray 2016, arXiv:1605.06376
  Alsing, Feeney & Wandelt 2018, arXiv:1801.01497
- Bayesian Optimisation for Likelihood-Free Inference (BOLFI)
  Gutmann & Corander 2016, arXiv:1501.03291
  FL 2018, arXiv:1805.07152

The "number of parameters" route:

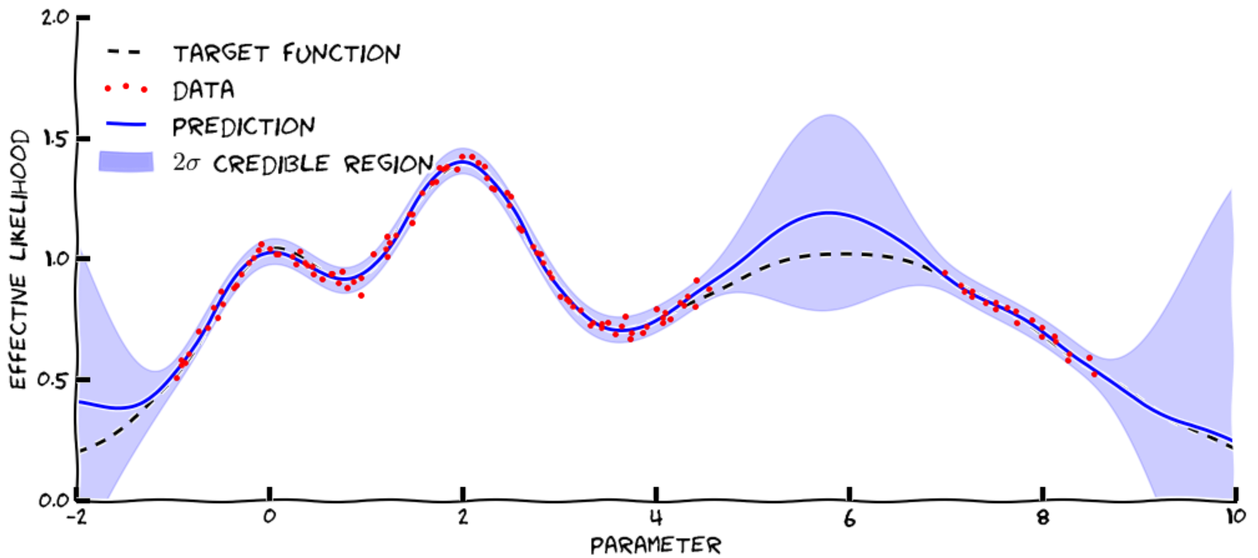
- Model-independent theoretical parametrisation ($d \gtrsim 100$), strong existing constraints in parameter space
- Simulator Expansion for Likelihood-Free Inference (SELFI)
  FL, Enzi, Jasche & Heavens 2019, arXiv:1902.10149



The “number of simulations” route: BOLFI

Bayesian Optimisation for Likelihood-Free Inference

BOLFI: Regression of the effective likelihood



1. "LFRS rejects most samples when ϵ is small"

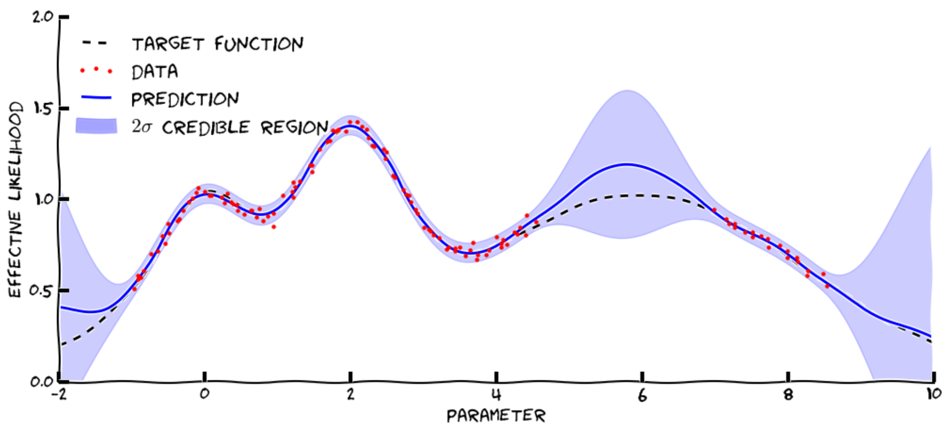
- Keep all values (θ_i, d_i) $\qquad \mathrm{d}_i = \mathrm{d}(\tilde{d}(\theta_i), d)$

2. "LFRS does not make assumptions about the shape of $L(\theta)$"

- Model the conditional distribution of distances given this training set

Gutmann & Corander JMLR 2016, arXiv:1501.03291

Gaussian process regression (a.k.a. kriging)



- Why?
  - It is a **general purpose regressor**: it will be able to deal with a large variety of complex/non-linear features of likelihood functions.
  - It provides not only a prediction, but also the **uncertainty of the regression**.
  - It allows to **extrapolate** in regions where we have no data points.

$$p(\mathbf{f}|\mathbf{X}) \propto \exp\left[-\frac{1}{2}\sum_{mn}(f(\mathbf{x}_m)-\mu(\mathbf{x}_m))^\intercal K(\mathbf{x}_m,\mathbf{x}_n)(f(\mathbf{x}_n)-\mu(\mathbf{x}_n))\right]$$

$$K(\mathbf{x}_m,\mathbf{x}_n) = \underbrace{C_1}_{K_\mathrm{C}(C_1)} \times \underbrace{\exp\left[-\frac{1}{2}\left(\frac{\mathbf{x}_m-\mathbf{x}_n}{C_2}\right)^2\right]}_{K_\mathrm{RBF}(C_2)} + \underbrace{C_3\delta_\mathrm{K}^{mn}}_{K_\mathrm{GN}(C_3)}$$

The prediction and uncertainty for a new point is:

$$p(f_\star|\mathbf{x}_\star,\mathbf{X},\mathbf{f}) \propto \exp\left[-\frac{1}{2}\left(\frac{f_\star-\alpha(\mathbf{x}_\star)}{\sigma(\mathbf{x}_\star)}\right)^2\right]$$

$$\alpha(\mathbf{x}_\star) = \mu(\mathbf{x}_\star) + K(\mathbf{x}_\star,\mathbf{x}_m)^\intercal K^{-1}(\mathbf{x}_m,\mathbf{x}_n)(\mathbf{f}-\mu(\mathbf{X}))_n$$

$$\sigma(\mathbf{x}_\star)^2 = K(\mathbf{x}_\star,\mathbf{x}_\star) - K(\mathbf{x}_\star,\mathbf{x}_m)^\intercal K^{-1}(\mathbf{x}_m,\mathbf{x}_n)K(\mathbf{x}_\star,\mathbf{x}_n)$$

Hyperparameters C_1, C_2, C_3 are automatically adjusted during the regression.

e.g. Rasmussen & Williams 2006

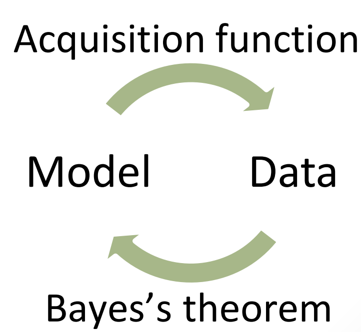
BOLFI: Data acquisition

3. "LFRS uses only a fixed proposal distribution, not all information available"

- Samples are obtained from sampling an **adaptively-constructed proposal distribution**, using the regressed effective likelihood

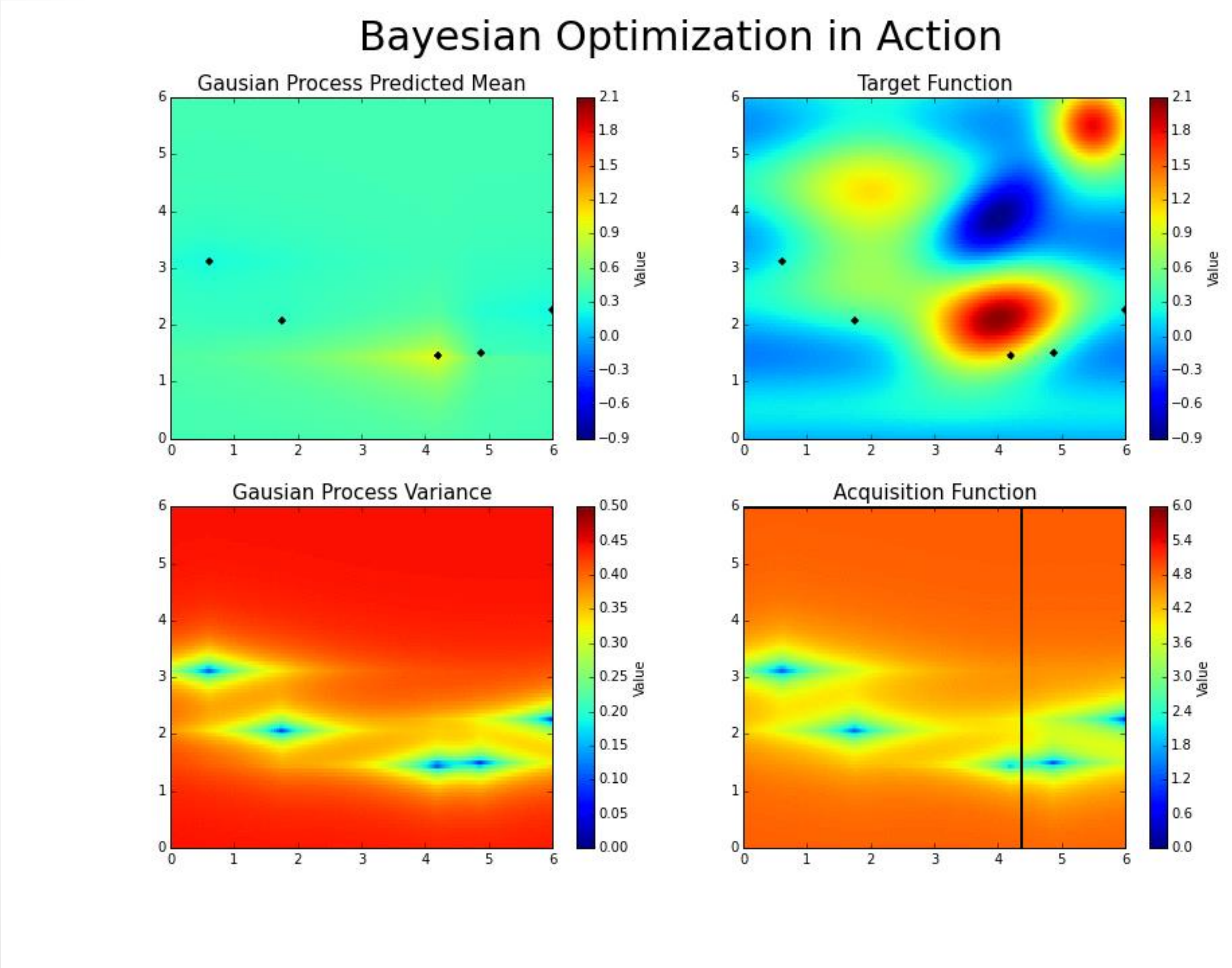
4. "LFRS aims at equal accuracy for all regions in parameter space"

- The **acquisition function** finds a compromise between exploration (trying to find new high-likelihood regions) & exploitation (giving priority to regions where the distance to the observed data is already known to be small)
- **Bayesian optimisation** (decision making under uncertainty) can then be used



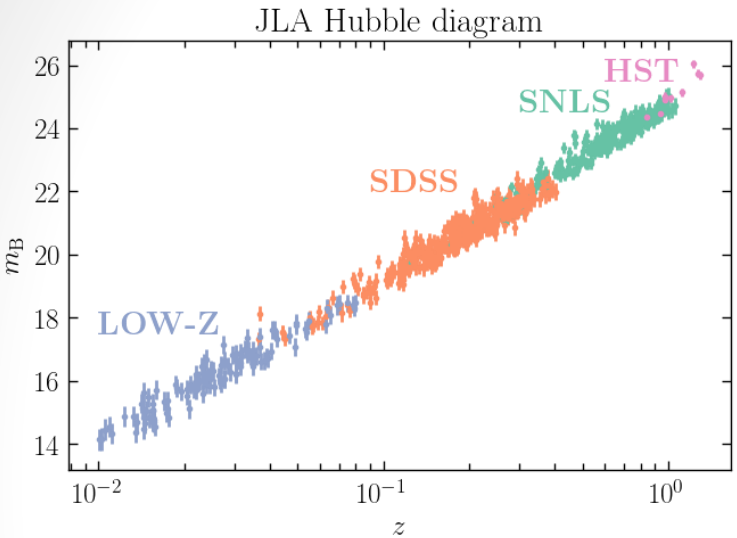
Gutmann & Corander JMLR 2016, arXiv:1501.03291

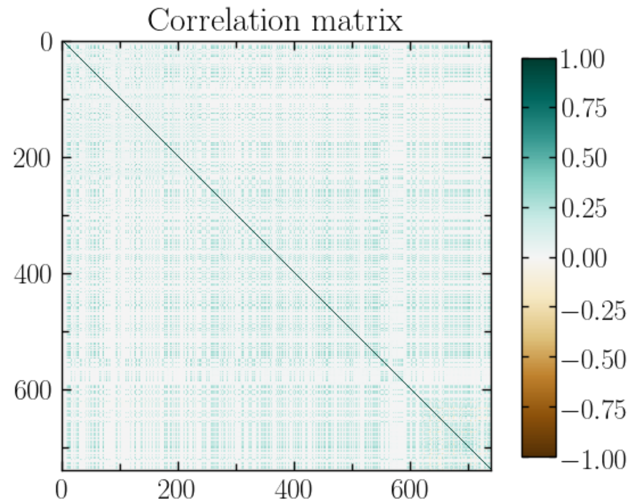
BOLFI: Data acquisition



F. Nogueira, https://github.com/fmfn/BayesianOptimization

BOLFI: Re-analysis of the JLA supernova sample





Betoule et al. 2014, arXiv:1401.4064

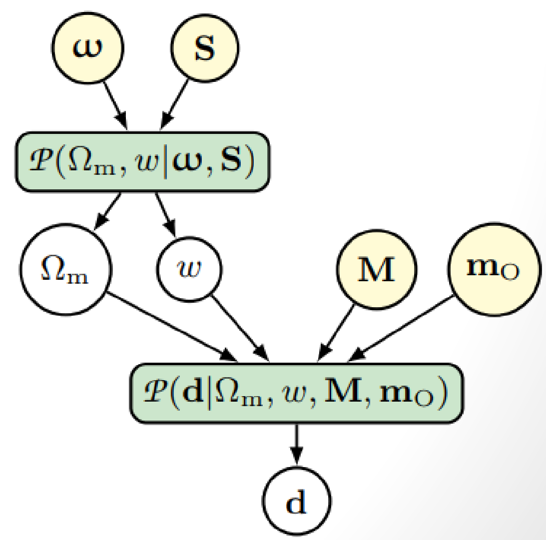
- 6-parameter model:
  2 cosmological parameters + 4 nuisance parameters

$$m_\mathrm{B} = 5\log_{10}\left[\frac{D_\mathrm{L}(z)}{10\ \mathrm{pc}}\right] + \widetilde{M}_\mathrm{B}(M_\mathrm{stellar}, M_\mathrm{B}, \delta M) - \boxed{\alpha} X_1 + \boxed{\beta} C$$

$$\widetilde{M}_\mathrm{B}(M_\mathrm{stellar}, M_\mathrm{B}, \delta M) = \boxed{M_\mathrm{B}} + \boxed{\delta M}\,\Theta\left(M_\mathrm{stellar} - 10^{10}\mathrm{M}_\odot\right)$$

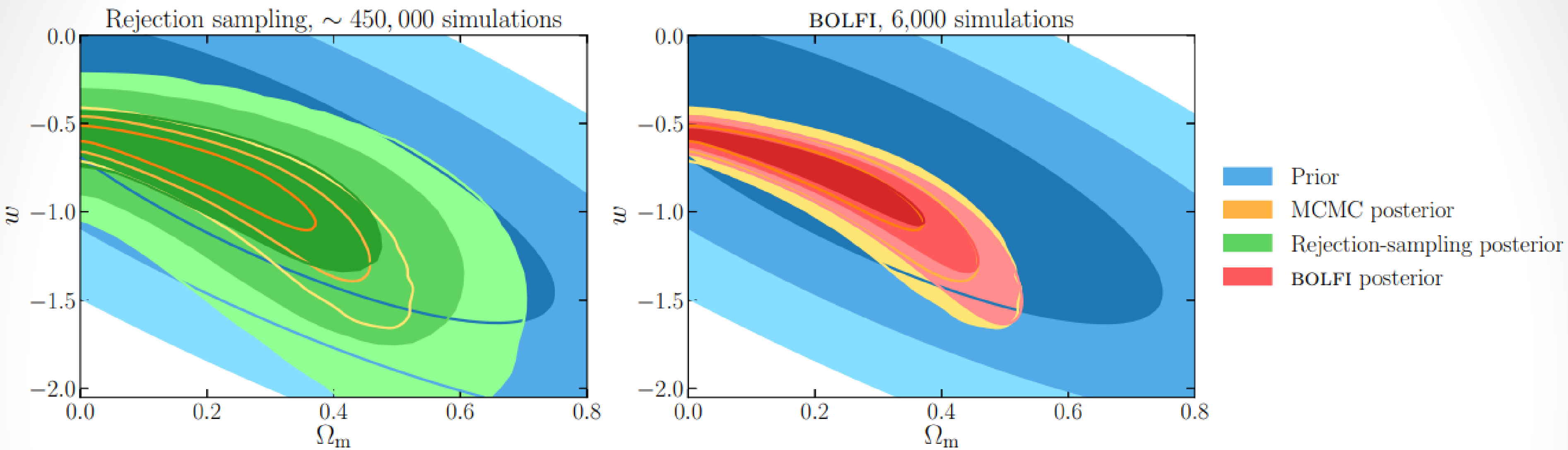
$$D_\mathrm{L}(z) = \frac{(1+z)\,\mathrm{c}}{H_0}\int_0^z \frac{\mathrm{d}z'}{E(z')}$$

$$E(z) \equiv \sqrt{\boxed{\Omega_\mathrm{m}}(1+z)^3 + (1-\boxed{\Omega_\mathrm{m}})(1+z)^{3(\boxed{w}+1)}}$$



FL 2018, arXiv:1805.07152

BOLFI: Re-analysis of the JLA supernova sample



- The **number of required simulations is reduced** by:
  - 2 orders of magnitude with respect to likelihood-free rejection sampling (for a much better approximation of the posterior)
  - 3 orders of magnitude with respect to exact Markov Chain Monte Carlo sampling

FL 2018, arXiv:1805.07152

Standard acquisition functions are suboptimal

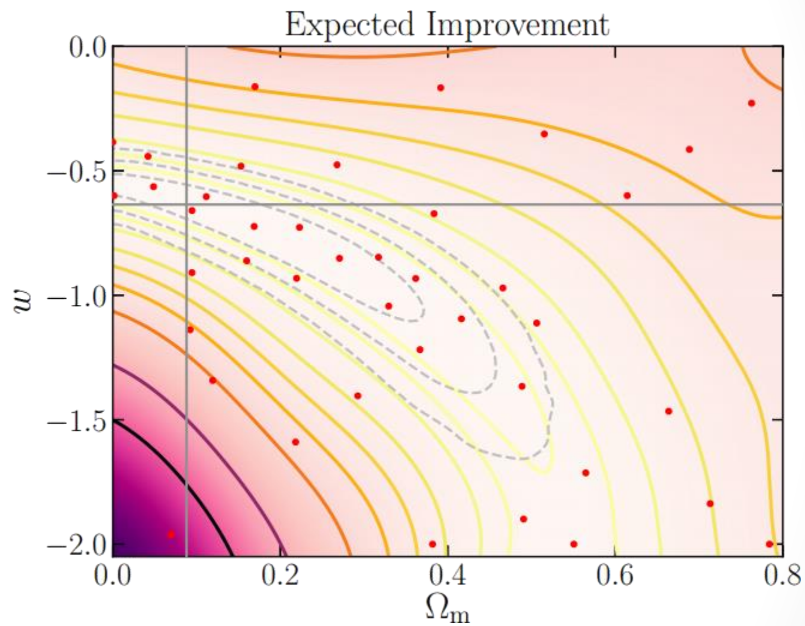
- Goal for Bayesian optimisation: find the optimum (assumed unique) of a function
- Example of acquisition function : the **Expected Improvement**

Gaussian cdf

Gaussian pdf

$$\mathrm{EI}(\boldsymbol{\theta}_\star) \equiv \underbrace{\sigma(\boldsymbol{\theta}_\star)}_{\text{Exploration}}\ \underbrace{\left[z\Phi(z) + \phi(z)\right]}_{\text{Exploitation}}$$

$$z \equiv \frac{\min(\mathbf{f}) - \mu(\boldsymbol{\theta}_\star)}{\sigma(\boldsymbol{\theta}_\star)}$$



- Drawbacks:
  - Do not take into account prior information
  - Local evaluation rules
  - Too greedy for ABC

e.g. Brochu, Cora & de Freitas 2010, arXiv:1012.2599

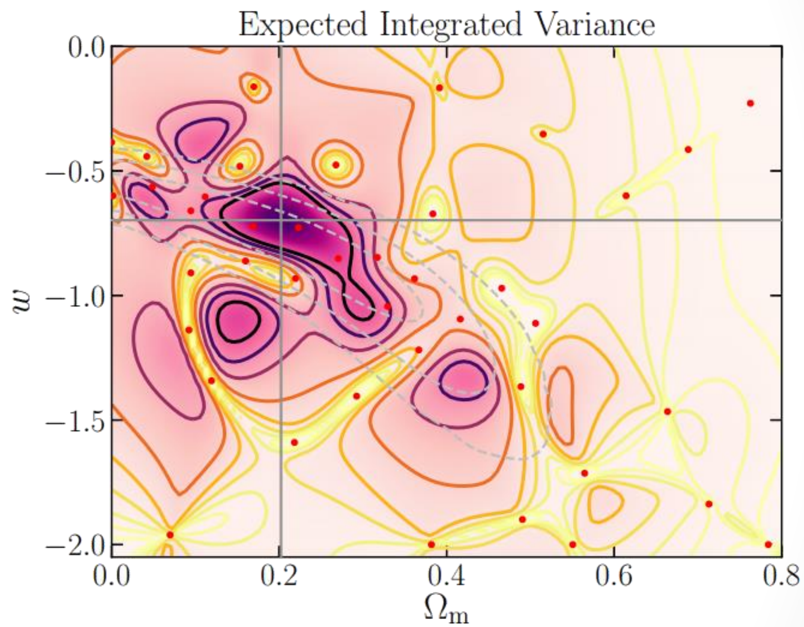
FL 2018, arXiv:1805.07152

The optimal acquisition function for ABC

- Goal for ABC: minimise the expected uncertainty in the estimate of the approximate posterior over the future evaluation of the simulator
- The optimal acquisition function : the **Expected Integrated Variance**

$$\mathrm{EIV}(\boldsymbol{\theta}_\star) = \underset{\text{Integral}}{\int} \frac{\overset{\text{Prior}}{\mathcal{P}(\boldsymbol{\theta})^2}}{4} \underbrace{\exp\left[-\mu(\boldsymbol{\theta})\right]}_{\text{Exploitation}} \underbrace{\left[\sigma^2(\boldsymbol{\theta}) - \tau^2(\boldsymbol{\theta}, \boldsymbol{\theta}_\star)\right]}_{\text{Exploration}} \, \mathrm{d}\boldsymbol{\theta}$$

$$\tau^2(\boldsymbol{\theta}, \boldsymbol{\theta}_\star) \equiv \frac{\mathrm{cov}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_\star)}{\sigma^2(\boldsymbol{\theta}_\star)}$$



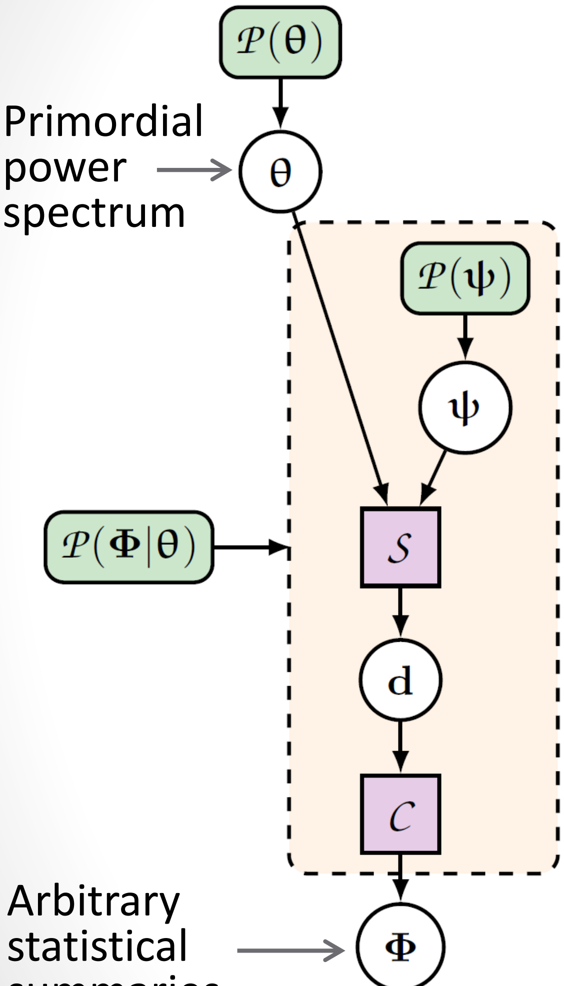
- Advantages:
  - Takes into account the prior
  - Non-local (integral over parameter space): more expensive… but much more informative
  - Exploration of the posterior tails is favoured when necessary
  - Analytic gradient

Järvenpää et al. 2017, arXiv:1704.00520 (expression of the EIV in the non-parametric approach)
FL 2018, arXiv:1805.07152 (expression of the EIV in the parametric approach)

The “number of parameters” route: SELFI

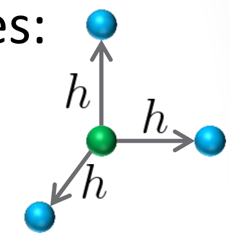
Simulator Expansion for Likelihood-Free Inference

SELFI: Method



- Gaussian prior + Gaussian effective likelihood
- Linearisation of the black-box around an expansion point + finite differences:

$$\hat{\boldsymbol{\Phi}}_\theta \approx \mathbf{f}_0 + \nabla\mathbf{f}_0 \cdot (\theta - \theta_0)$$



The posterior is Gaussian and analogous to a Wiener filter:

expansion point

observed summaries

$$\boldsymbol{\gamma} \equiv \theta_0 + \boldsymbol{\Gamma}\,(\nabla\mathbf{f}_0)^\intercal\,\mathbf{C}_0^{-1}(\boldsymbol{\Phi}_\mathrm{O} - \mathbf{f}_0)$$

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

prior covariance

covariance of summaries

gradient of the black-box

FL, Enzi, Jasche & Heavens 2019, arXiv:1902.10149

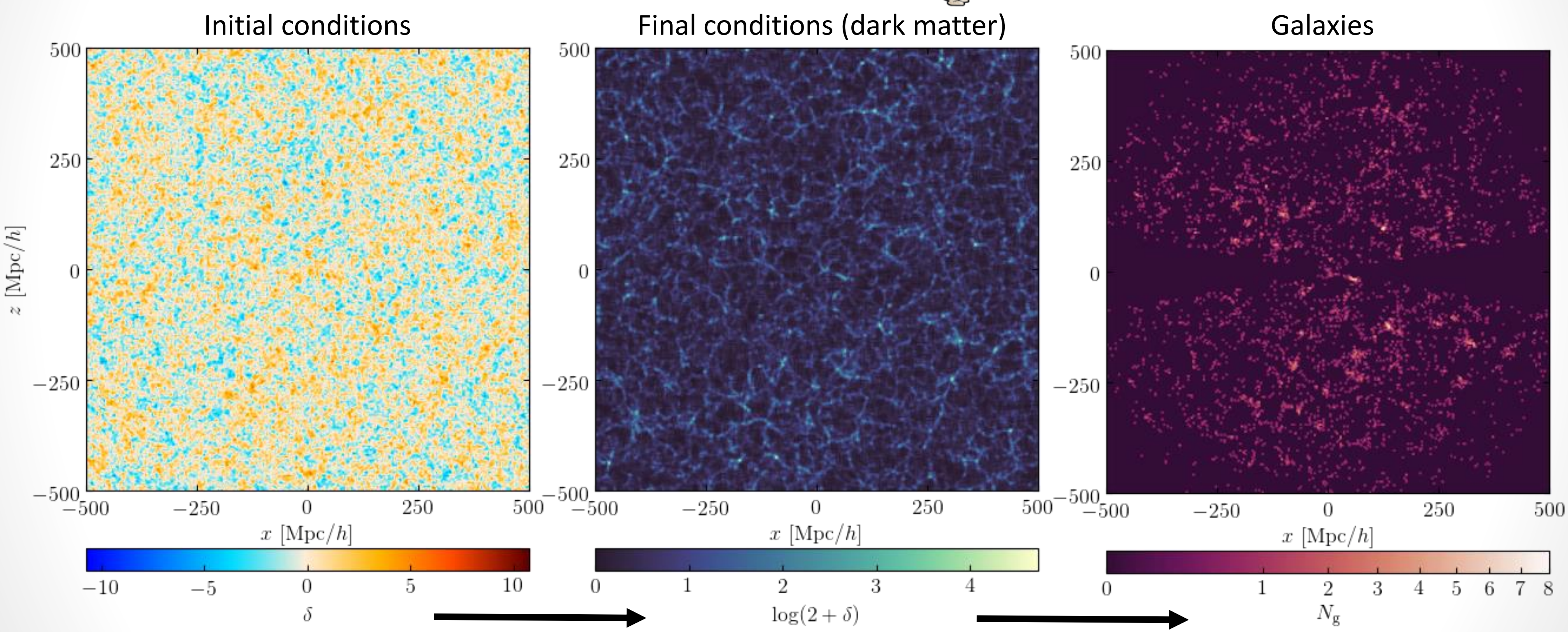
$\mathbf{f}_0, \mathbf{C}_0$ and $\nabla\mathbf{f}_0$ can be evaluated through simulations only

A black-box: Simbelmynë

I'm happy to explain the name during the reception...

Publicly available code:

https://bitbucket.org/florent-leclercq/simbelmyne/

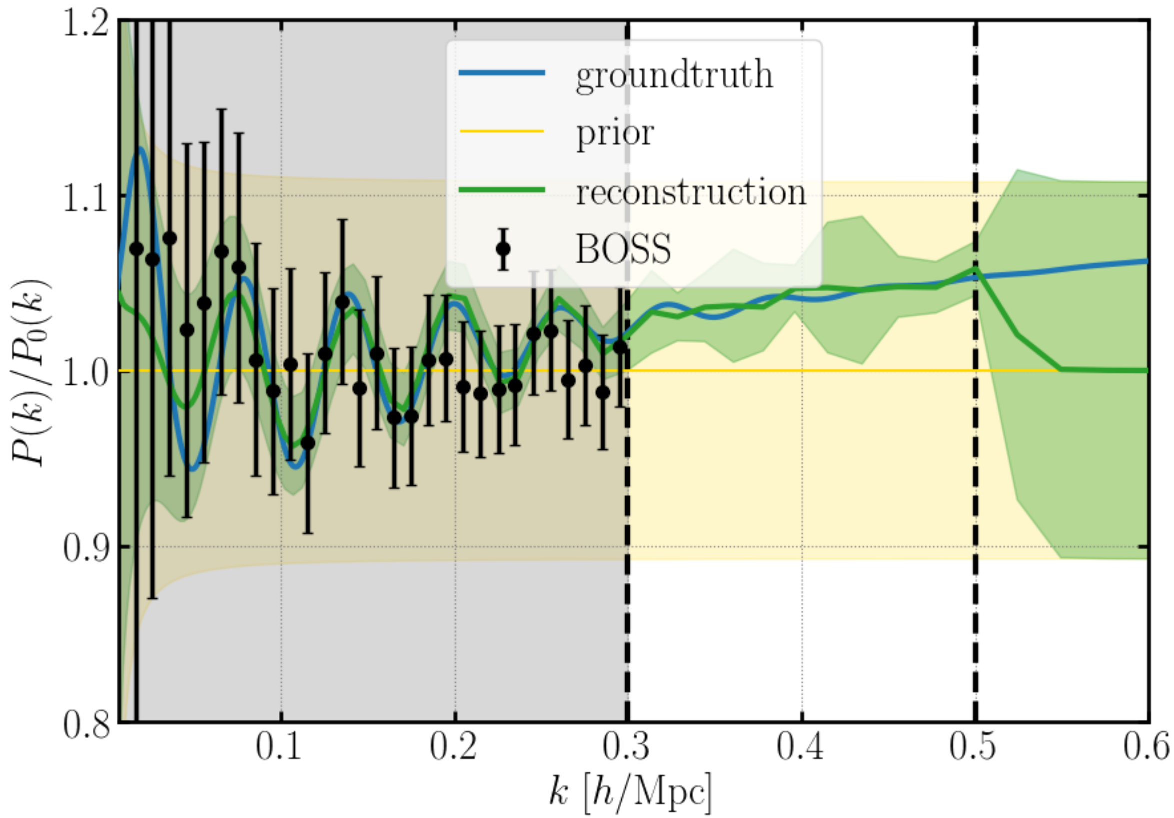


Dark matter simulation with COLA

Tassev, Zaldarriaga & Eisenstein 2013, arXiv:1301.0322

Survey simulation: Redshift-space distortions, galaxy bias, selection effects, survey geometry, instrumental noise

SELFI + Simbelmynë: Proof-of-concept

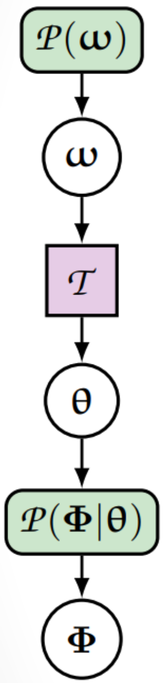


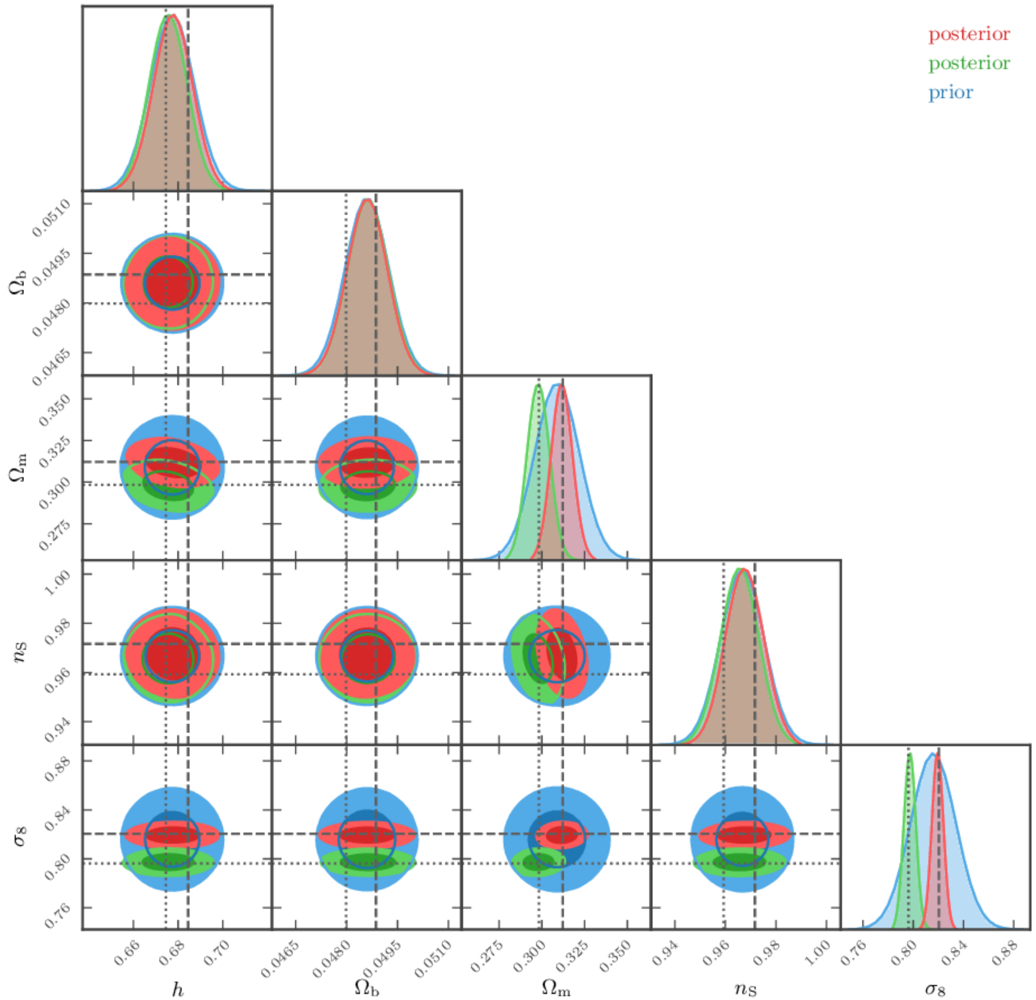
100 parameters are simultaneously inferred from a black-box data model

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

FL, Enzi, Jasche & Heavens 2019, arXiv:1902.10149

SELFI + Simbelmynë: Proof-of-concept





- Robust inference of cosmological parameters can be easily performed *a posteriori* once the linearised data model is learned
- pyselfi will be made publicly available soon

FL, Enzi, Jasche & Heavens 2019, arXiv:1902.10149

Concluding thoughts

- Goal: developing and using algorithms 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!
- The "number of simulations route" (BOLFI):
  - The number of simulations is reduced by several orders of magnitude.
  - The optimal acquisition function can be derived: the Expected Integrated Variance.
- The "number of parameters route" (SELFI):
  - High-dimensional likelihood-free problems can be addressed.
  - The computational workload is fixed a priori and perfectly parallel.