Simulation-based large-scale structure inference

Florent Leclercq
Imperial College Research Fellow
Imperial Centre for Inference and Cosmology

January 23rd, 2018

In collaboration with:

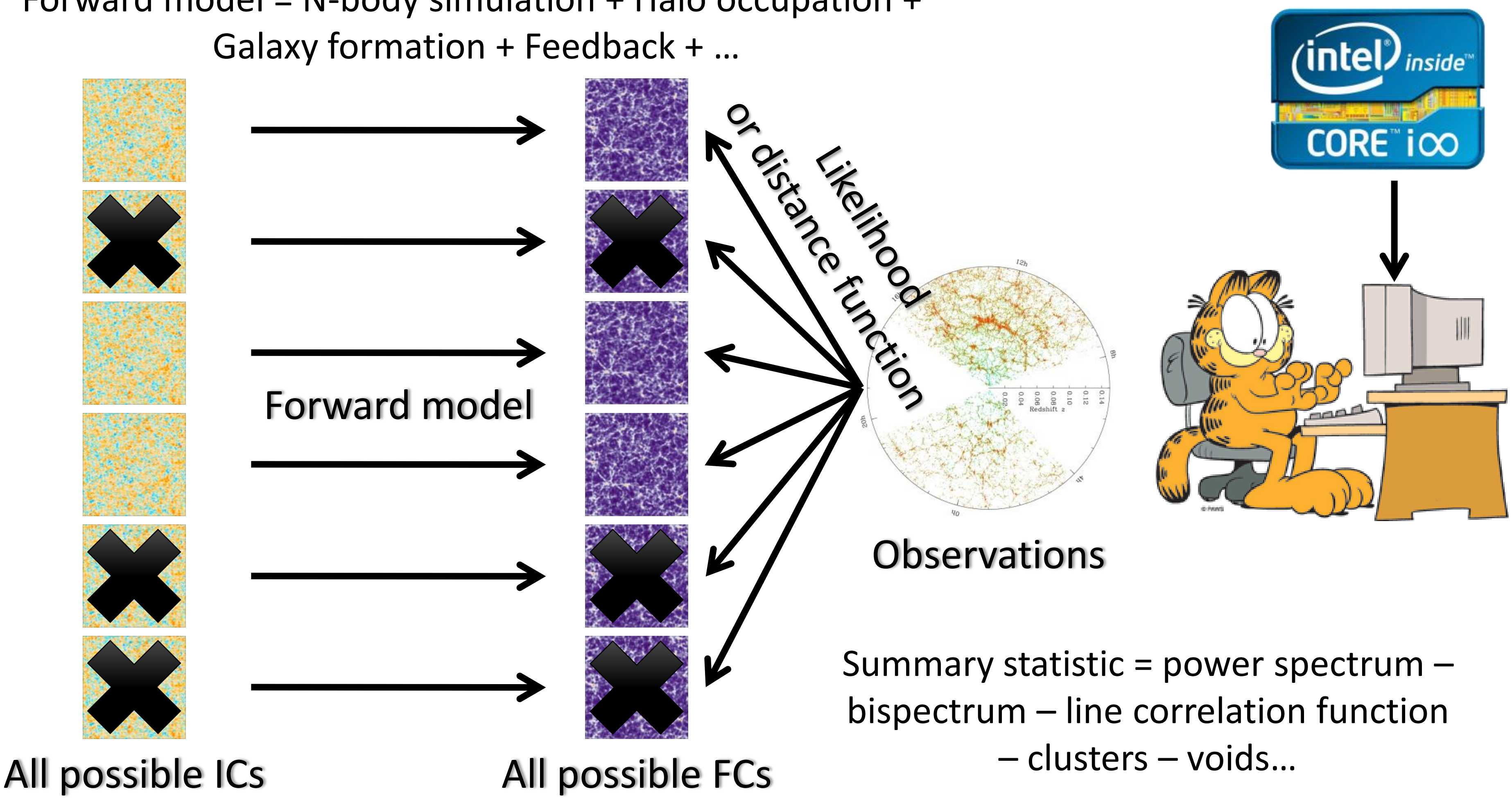
Wolfgang Enzi (MPA),
Baptiste Faure (École polytechnique),
Jens Jasche (ExC Garching/U. Stockholm)

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

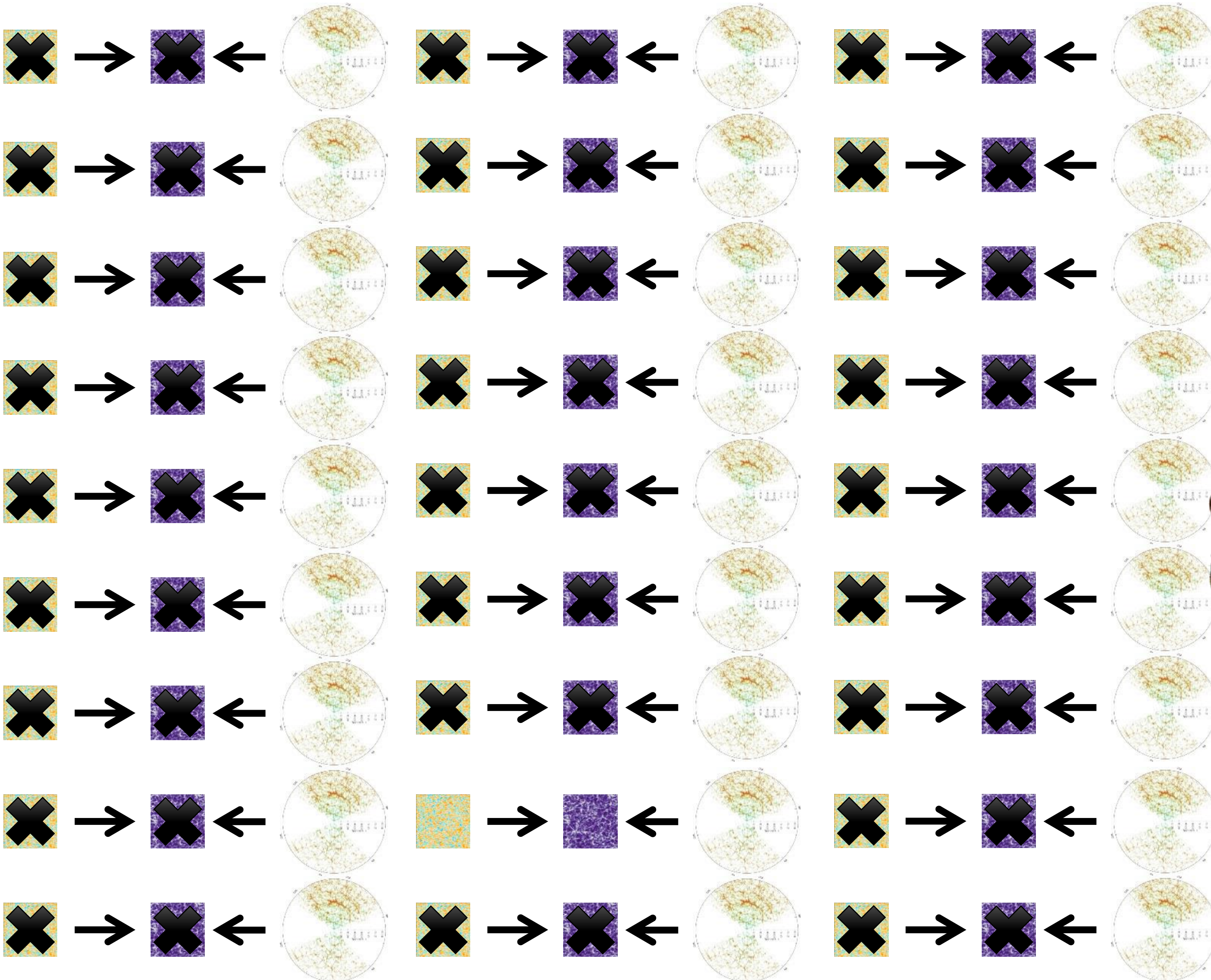
Bayesian forward modeling: the ideal scenario

Forward model = N-body simulation + Halo occupation + Galaxy formation + Feedback + …



Summary statistic = power spectrum – bispectrum – line correlation function – clusters – voids…

Bayesian forward modeling: the challenge



$d \approx 10^7$

Likelihood-Free Large-scale Structure Inference

Approximate Bayesian Computation (ABC)

- Statistical inference for models where:
  1. The likelihood function is intractable
  2. Simulating data is possible

- **General idea**: find parameter values for which the distance between simulated data and observed data is small

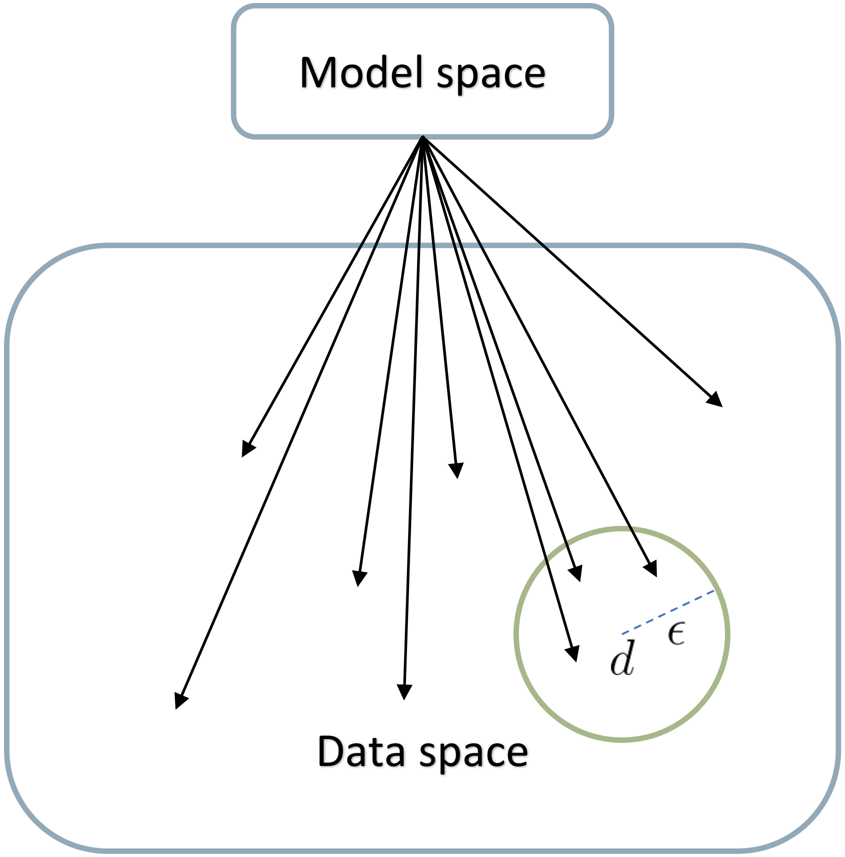
$$p(\theta|d) \Longrightarrow p(\theta|\tilde{d}) \quad \text{where } \mathrm{d}(\tilde{d}(\theta), d) \text{ is small}$$

- **Assumptions**:
  - Only a small number of parameters are of interest
  - But the process generating the data is very general: a noisy non-linear dynamical system with an unrestricted number of hidden variables

Likelihood-free rejection sampling

- 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)

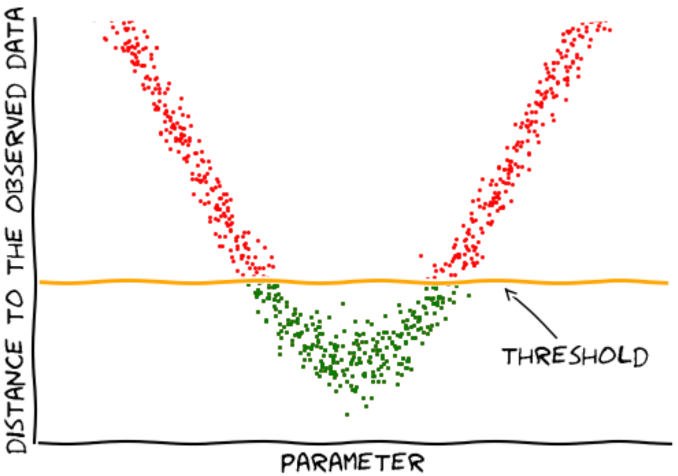
Why is likelihood-free rejection so expensive?

1. It rejects most samples when ϵ is small

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

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

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



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

Proposed solution

Bayesian optimisation for likelihood-free inference (BOLFI)

1. It rejects most samples when ϵ is small

   ➡ **Don't reject samples: learn from them!**

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

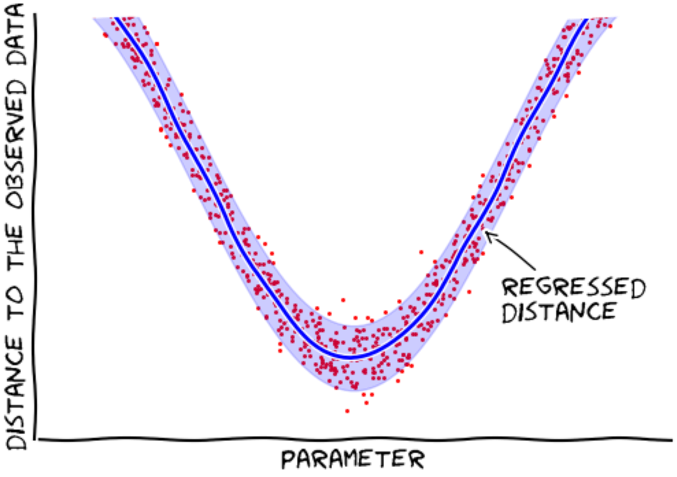
   ➡ **Model the distances, assuming the average distance is smooth**

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

   ➡ **Use Bayes' theorem to update the proposal of new points**

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

   ➡ **Prioritize parameter regions with small distances to the observed data**



Related work in cosmology:

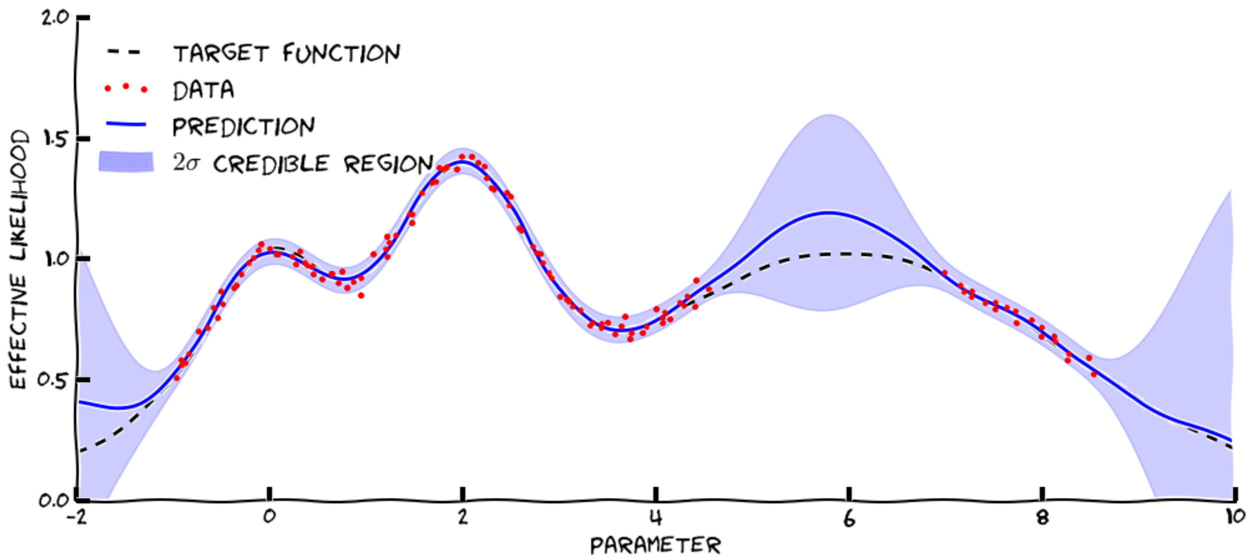
Alsing & Wandelt 2017, arXiv:1712.00012
(data compression for ABC)

Alsing, Wandelt & Feeney 2018, arXiv:1801.01497
(density estimation for ABC – DELFI)

Enzi, Jasche & FL 2018, to be submitted
(ABC with linear expansion of the effective likelihood)

Gutmann & Corander JMLR 2016, arXiv:1501.03291

Regressing the effective likelihood (points 1 & 2)



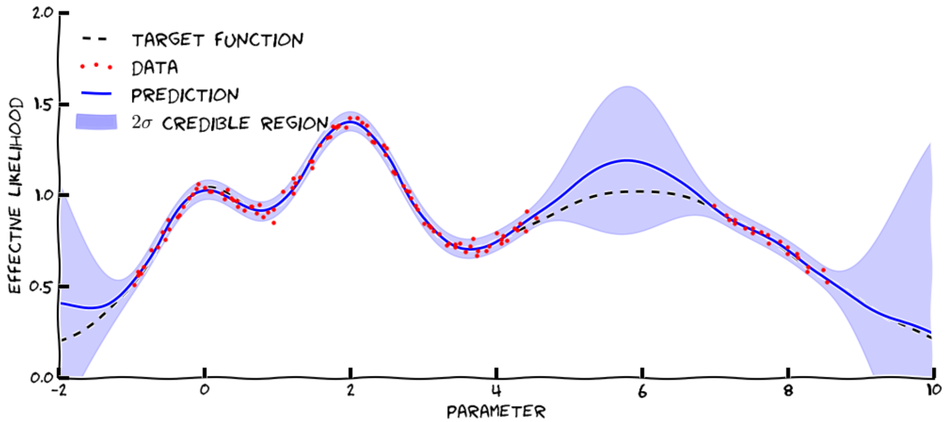
1. "It rejects most samples when ϵ is small"

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

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

- Model the conditional distribution of distances given this training set

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:

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

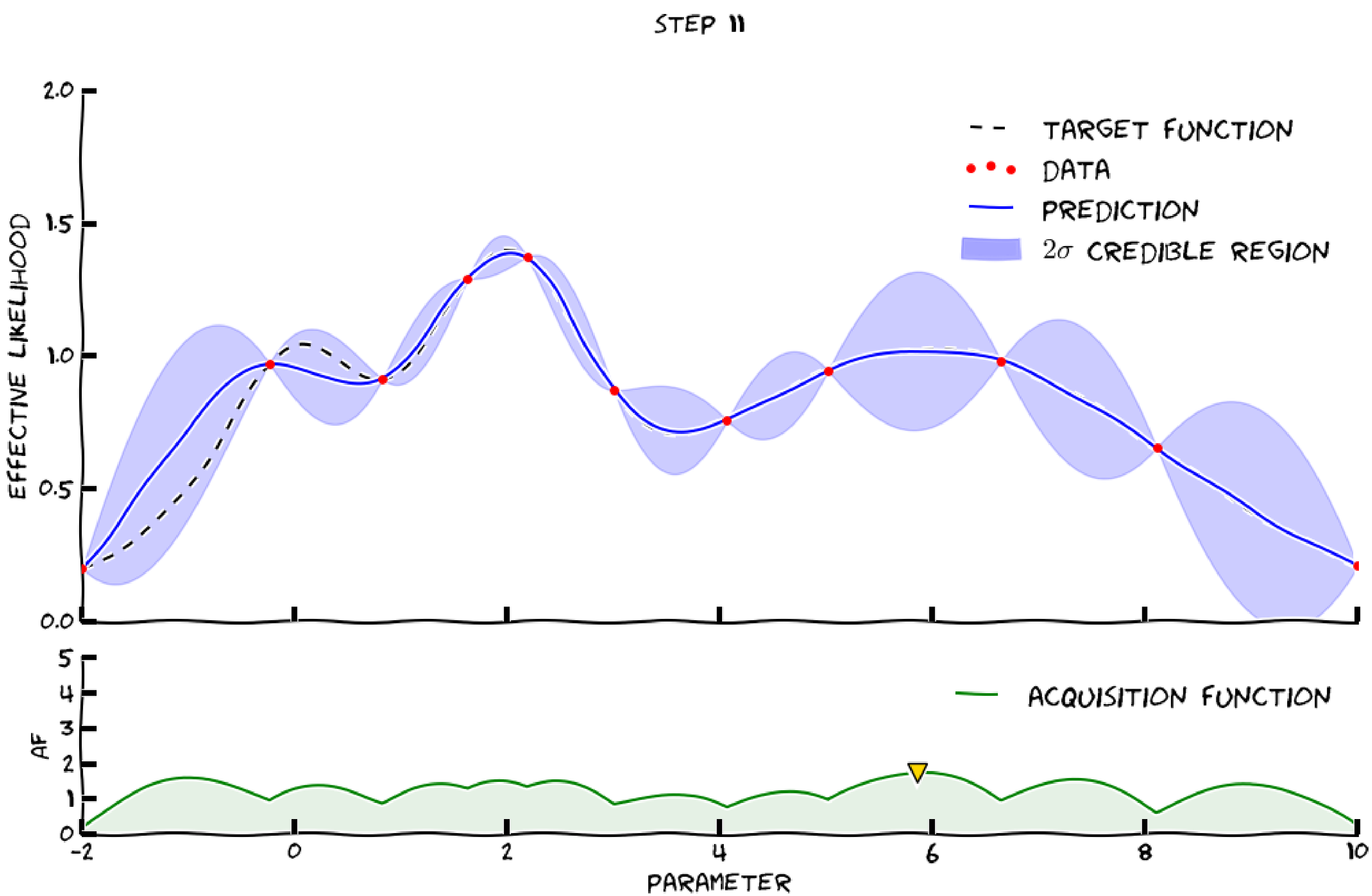
$$\alpha(\mathbf{x}_\star) = \mu(\mathbf{x}_\star) + K(\mathbf{x}_\star,\mathbf{x}_m)^\intercal K^{-1}(\mathbf{x}_m,\mathbf{x}_n)(\mathbf{f}-\mu(\mathbf{X}))_n$$

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

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

Rasmussen & Williams 2006

Data acquisition



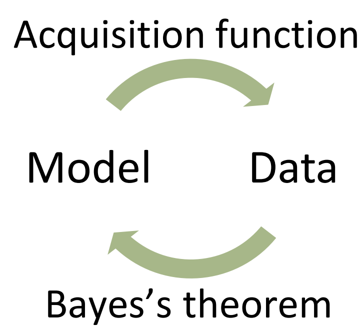
Data acquisition (points 3 & 4)

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

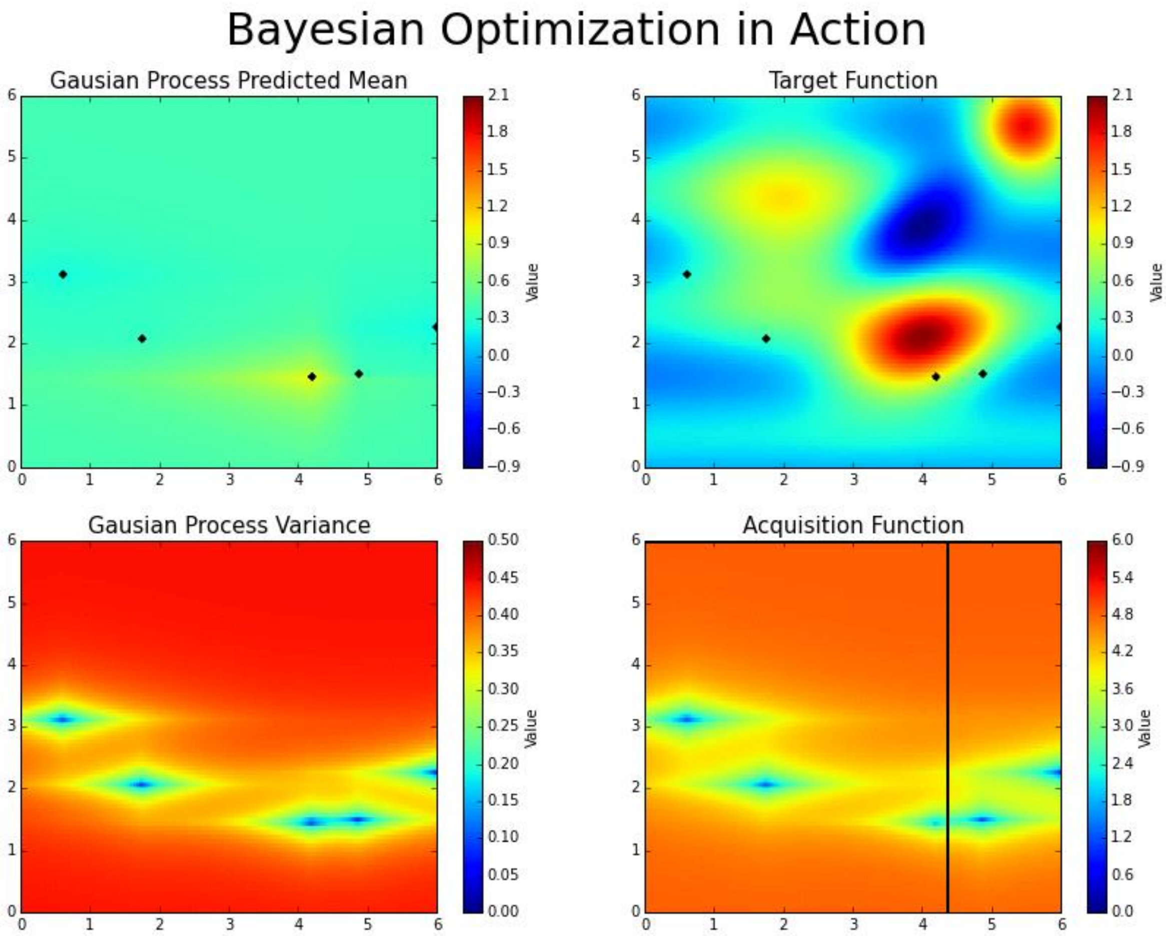
- Samples are obtained from sampling an **adaptively-constructed proposal distribution**, using the regressed effective likelihood

4. "It 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



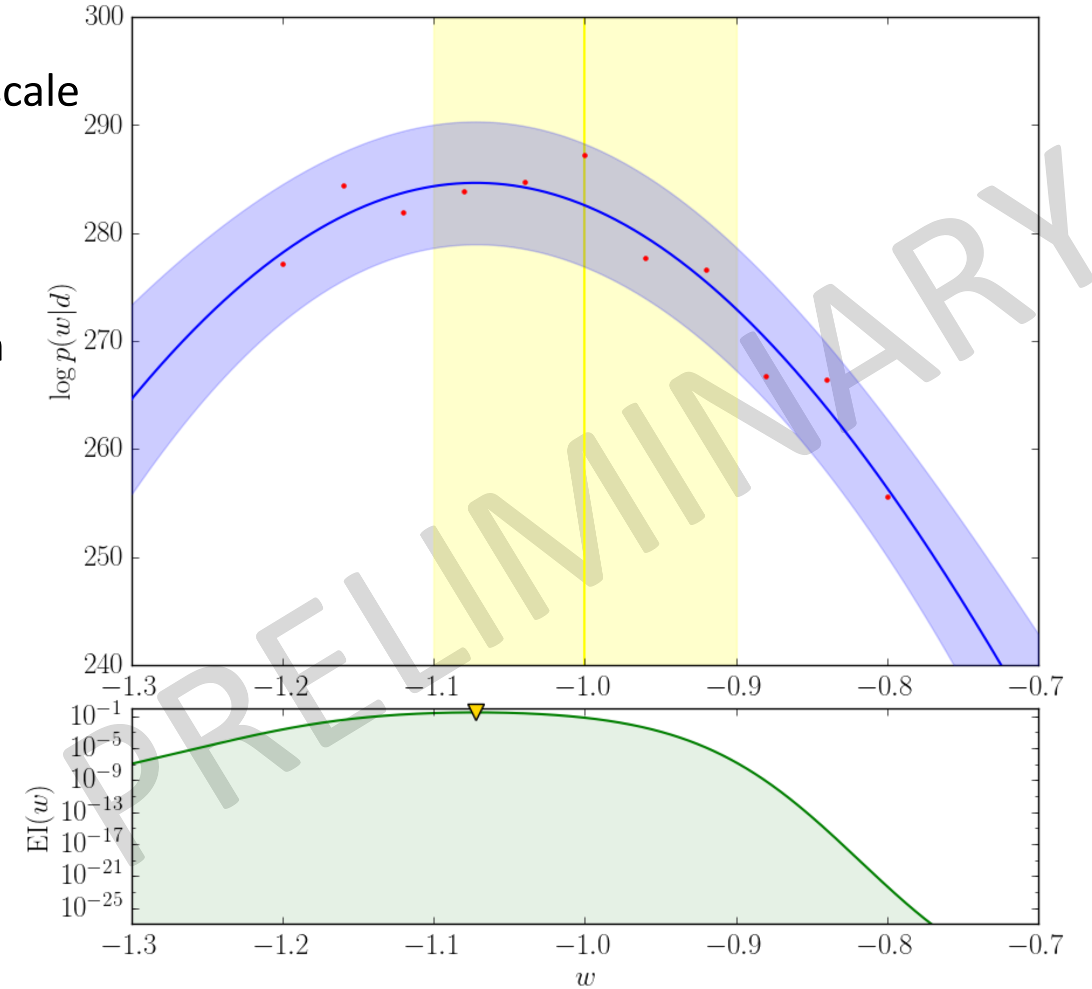
In higher dimension…



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

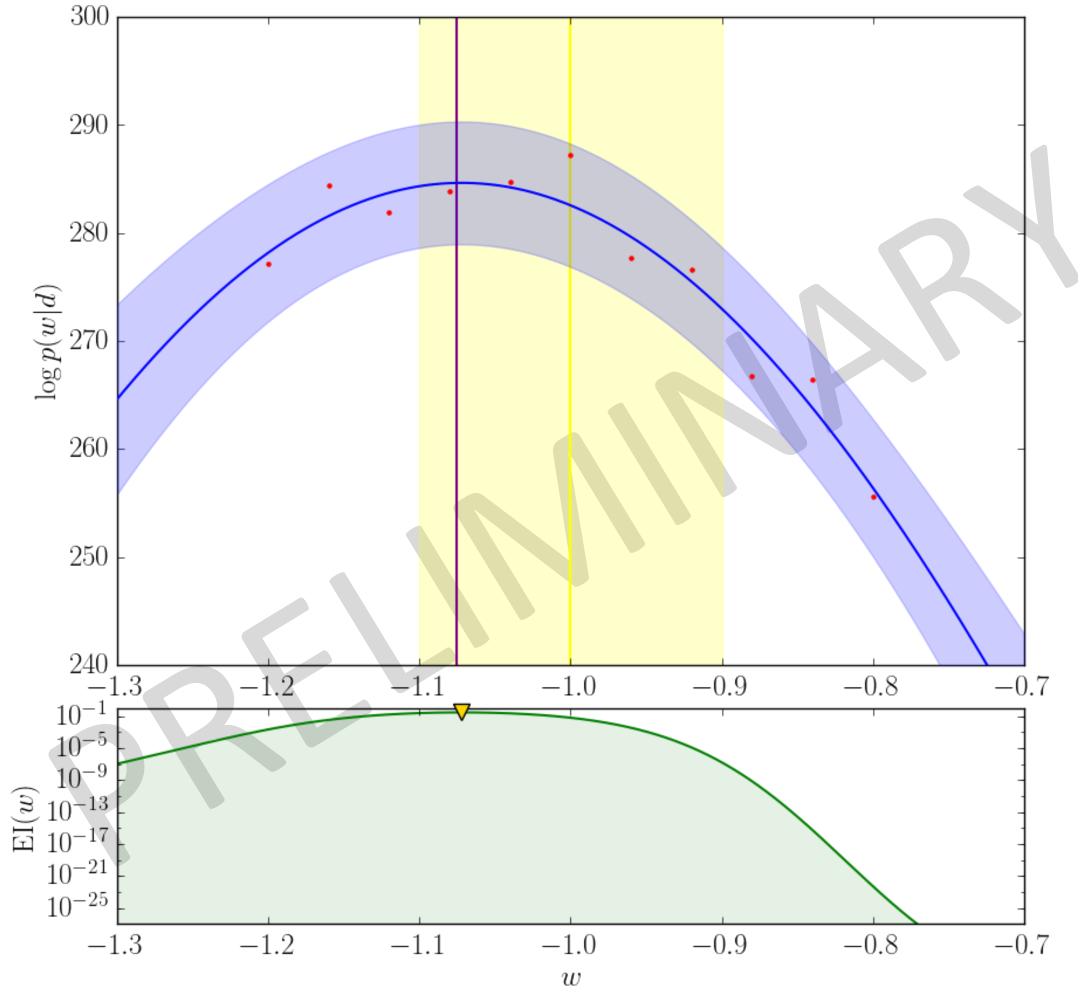
Likelihood-free large-scale structure inference

- 1100 large-scale structure simulations using COLA
- ≈10^7 hidden variables



FL, Jasche & Enzi (in prep.)

Likelihood-free large-scale structure inference



This proof-of-concept has been performed completely blindly.

FL, Jasche & Enzi (in prep.)

Summary

- A likelihood-free method for models where the likelihood is intractable but simulating is possible:
  Regression of the distance + Bayesian optimisation
  - Number of required simulations reduced by several orders of magnitude.
  - The approach will allow to **ask targeted questions to cosmological data**, including all relevant physical and observational effects.
- Optimisation of the data model using **tCOLA + sCOLA**
  - Enormous parallelisation potential for dark matter simulations.
  - Further speed-up expected for realistic synthetic observations.