SELFI enhanced: robustness to model misspecification and Euclid forecast

Euclid-France Theory and Likelihood workshop
Institut d'Astrophysique de Paris

Florent Leclercq

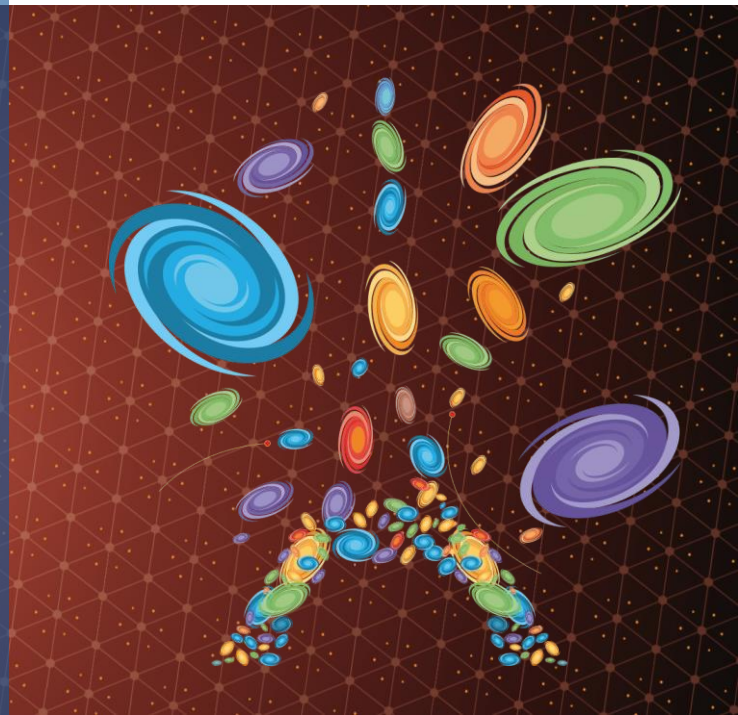
www.florent-leclercq.eu

Institut d'Astrophysique de Paris
CNRS & Sorbonne Université

In collaboration with the Aquila Consortium

www.aquila-consortium.org

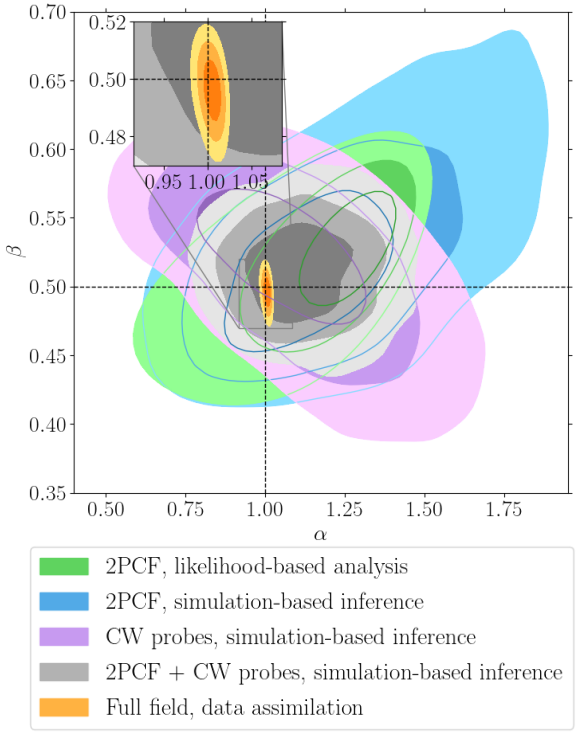
28 November 2022



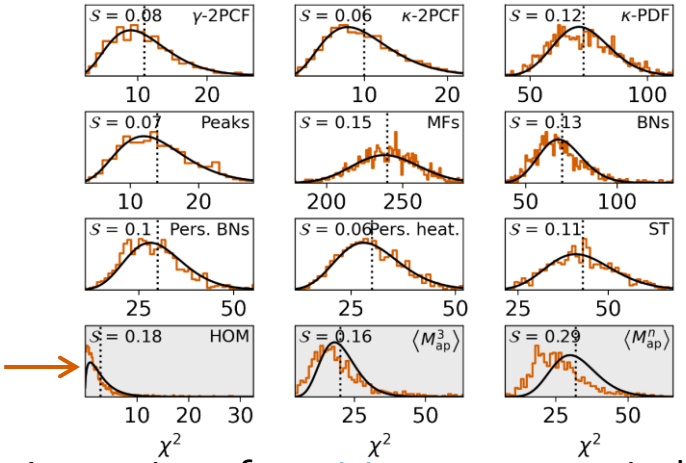
Why I decided to go "likelihood-free" for Galaxy Clustering Additional Probes (GC:AP)

Note: likelihood-free inference ≈ simulation-based inference ≈ implicit likelihood inference

- A question of accuracy: first, avoid biases.



- The HOWLS group also found that some WL additional probes have a non-Gaussian distribution.



- A question of precision: can numerical forward models be used to push further than $k \gtrsim 0.15\ h/\mathrm{Mpc}$? The full field contains much more information.

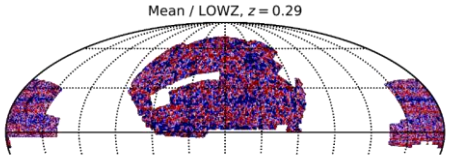
FL & Heavens, 2103.04158

HOWLS-KP paper 1, Ajani *et al.*, in prep.

The issue of model misspecification in Bayesian inference and in simulation-based inference (SBI)

- Model misspecification arises when model differs from actual data-generating process.
- Field-based inference techniques have a successful track record at handling model misspecification, e.g. automatically reporting unknown data contaminations.

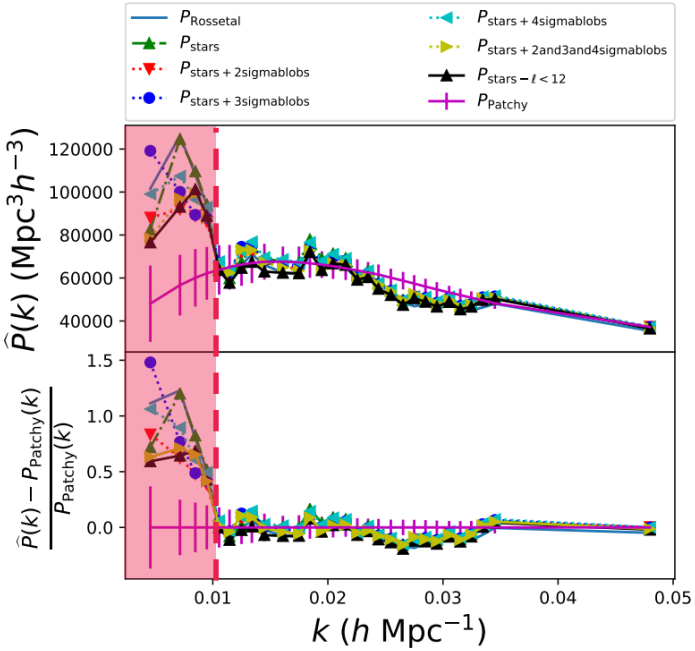
Map of unknown foreground contaminant



- Model misspecification: a major challenge particularly for approaches that marginalise over latent variables, such as simulation-based inference (SBI).

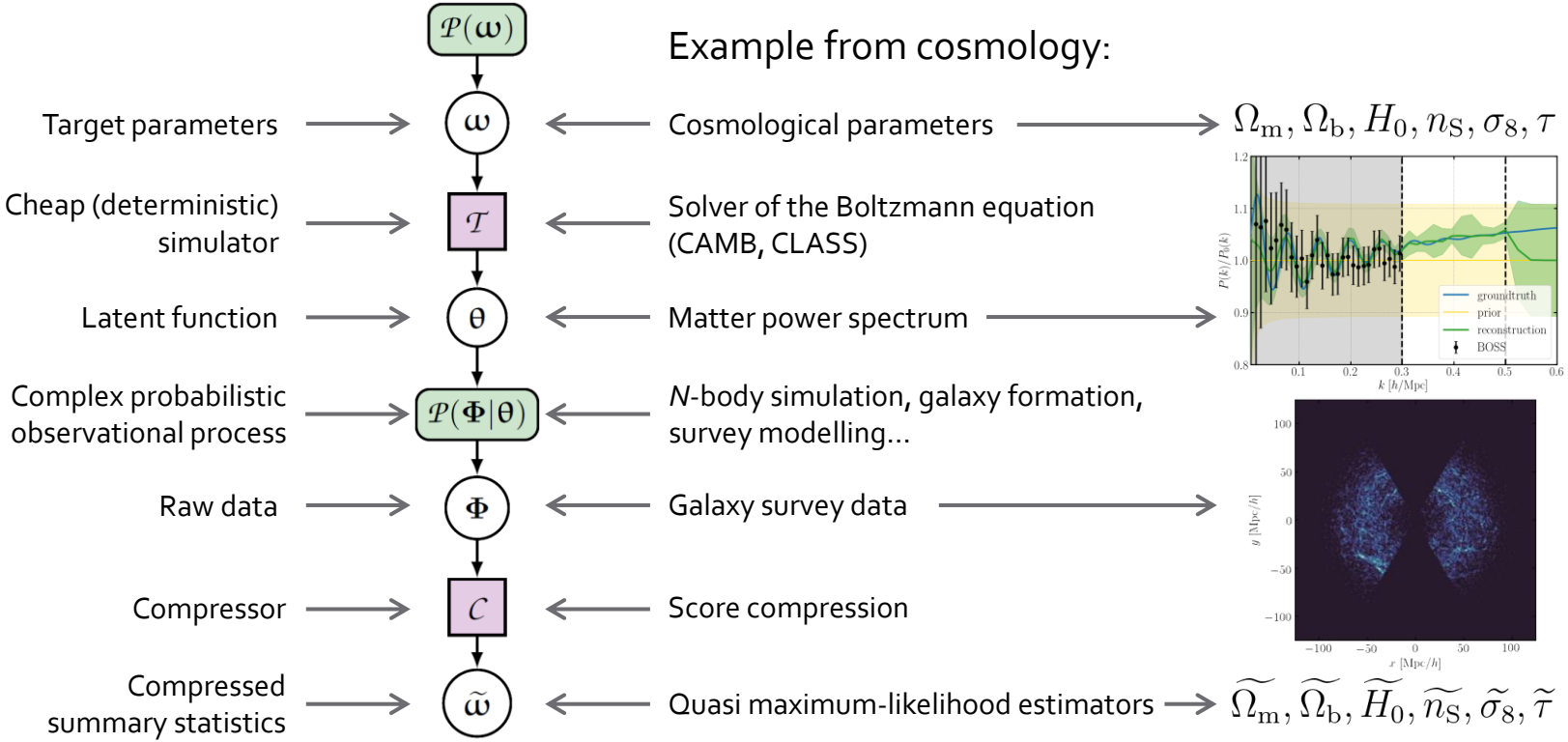
Porqueres, Ramanah, Jasche & Lavaux, 1812.05113
Lavaux, Jasche & FL, 1909.06396

- Typical cosmological example: the galaxy *power spectrum* at large scales.

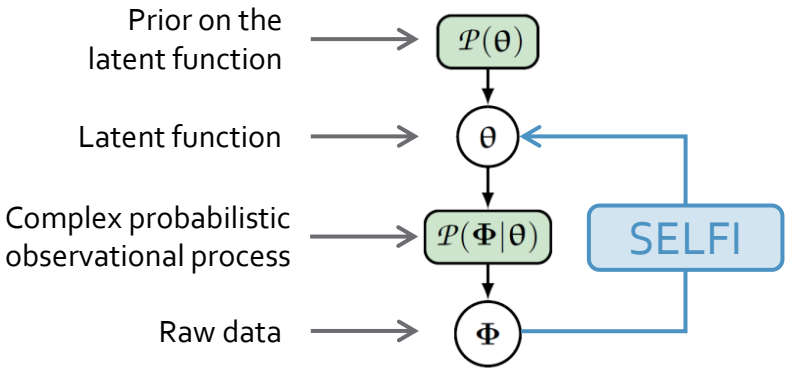


Kalus, Percival *et al.*, 1806.02789

A general class of Bayesian hierarchical models (BHMs): Complex observations of a latent function controlled by top-level parameters

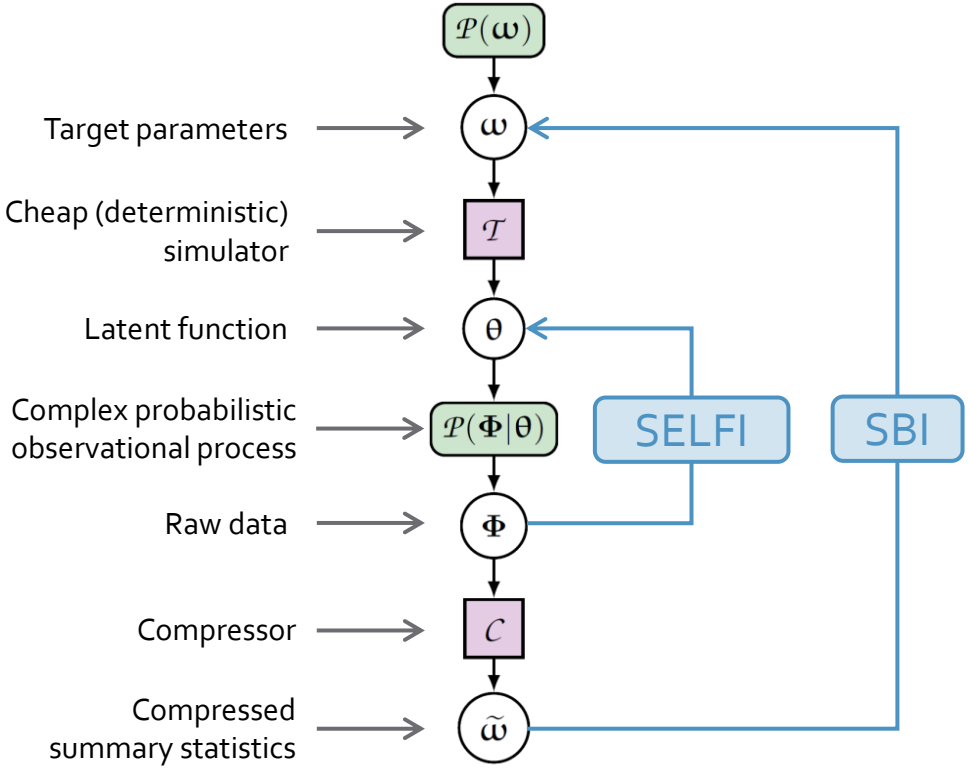


Key idea: a two-step SBI process that recycles simulations



1. Inference of the latent function θ, to check for model misspecification:
   - SELFI algorithm

Key idea: a two-step SBI process that recycles simulations

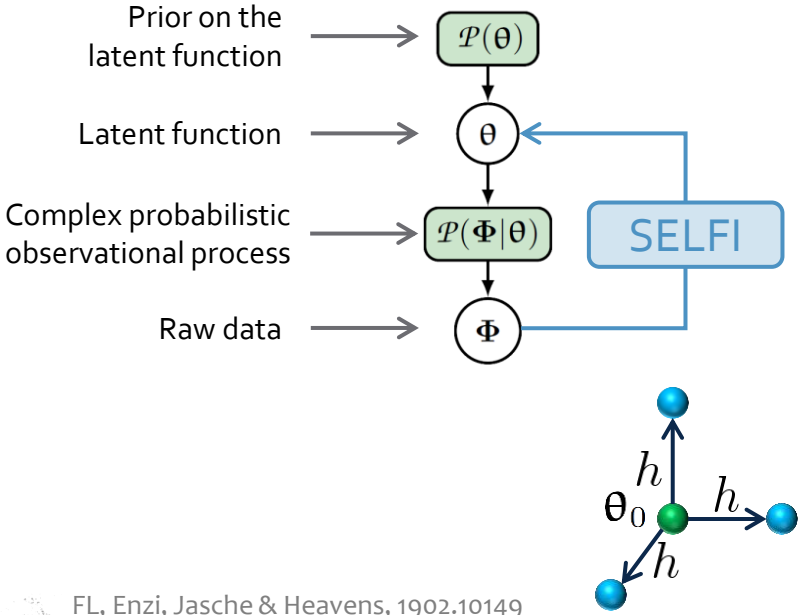


1. Inference of the latent function θ, to check for model misspecification:
   - SELFI algorithm

2. Simulation-based inference of ω:
   - Approximate Bayesian Computation (ABC), Likelihood-Free Rejection Sampling
   - Density/ratio estimation (DELFI / NRE)
   - Bayesian optimisation (BOLFI)
   - others…

Important: the simulations necessary for step 1. are recycled for data compression, which is required for step 2.

Latent function inference: the SELFI approach (*Simulator Expansion for Likelihood-Free Inference*)



- We aim at inferring the latent function θ, which usually contains most/all of the information on ω.

  (initial power spectrum in cosmology, prey/predator population functions in ecology)
- This requires doing SBI 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:

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

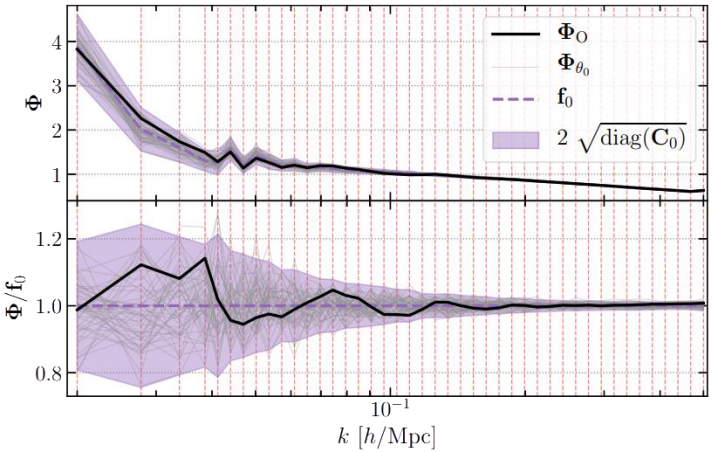
SELFI-2 (second order): coming soon!

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

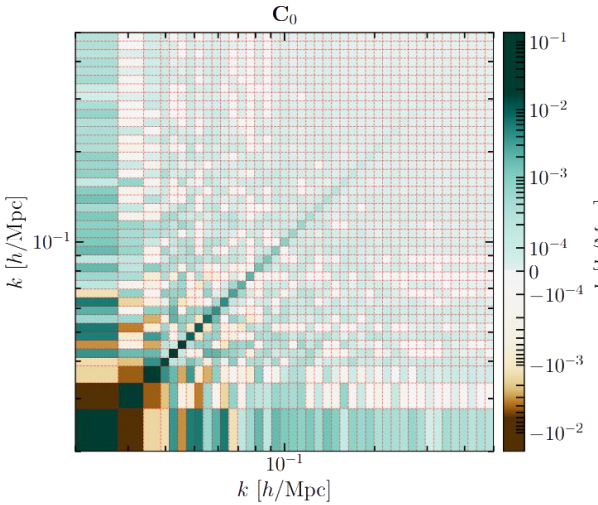
FL, Enzi, Jasche & Heavens, 1902.10149

Galaxy Clustering Additional Probes pipeline: diagnostics of the linearised black-box data model

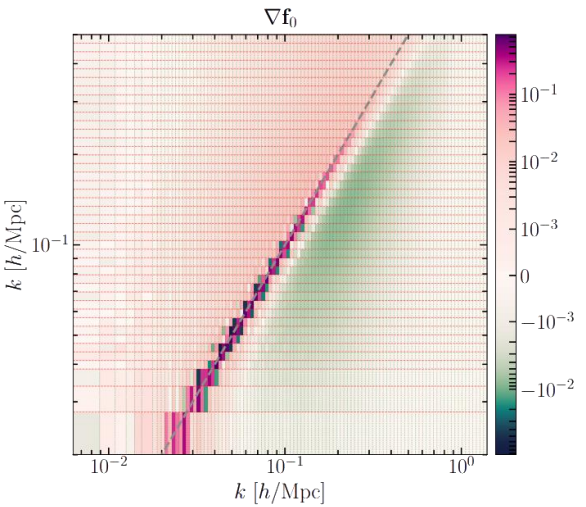
Black-box mean



Black-box covariance



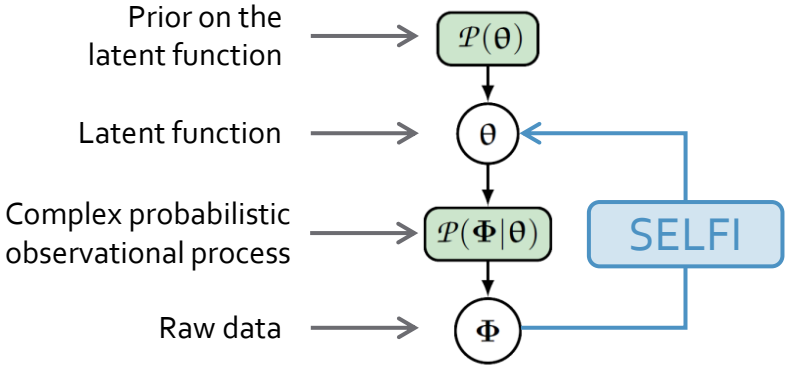
Black-box gradient



- Using only here the (final, non-linearly evolved) power as summary statistics.
- Any AP can go in the data vector, since we need the simulations anyway!

FL, Enzi, Jasche & Heavens, 1902.10149

Latent function inference: the SELFI approach (*Simulator Expansion for Likelihood-Free Inference*)



- Linearisation of the black-box data model:

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

- Further assume:
  - Gaussian prior: $\mathcal{P}(\boldsymbol{\theta}) = \mathcal{G}(\boldsymbol{\theta}_0, \mathbf{S})$
  - Gaussian effective likelihood: $\mathcal{P}(\boldsymbol{\Phi}|\boldsymbol{\theta}) = \mathcal{G}\left[\mathbf{f}(\boldsymbol{\theta}), \mathbf{C}_0\right]$

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

expansion point — observed summaries

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

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

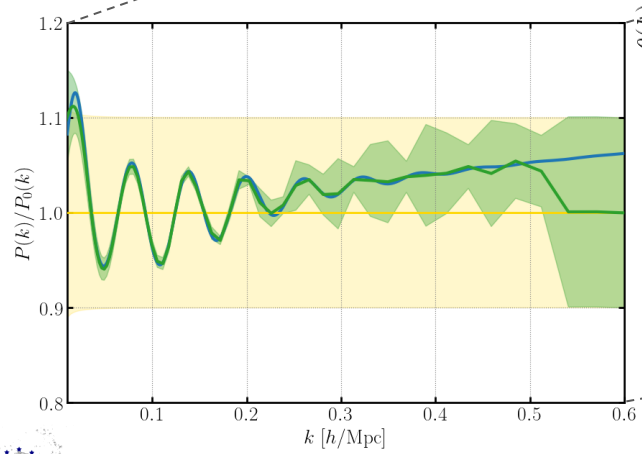
covariance of summaries — gradient of the black-box — prior covariance

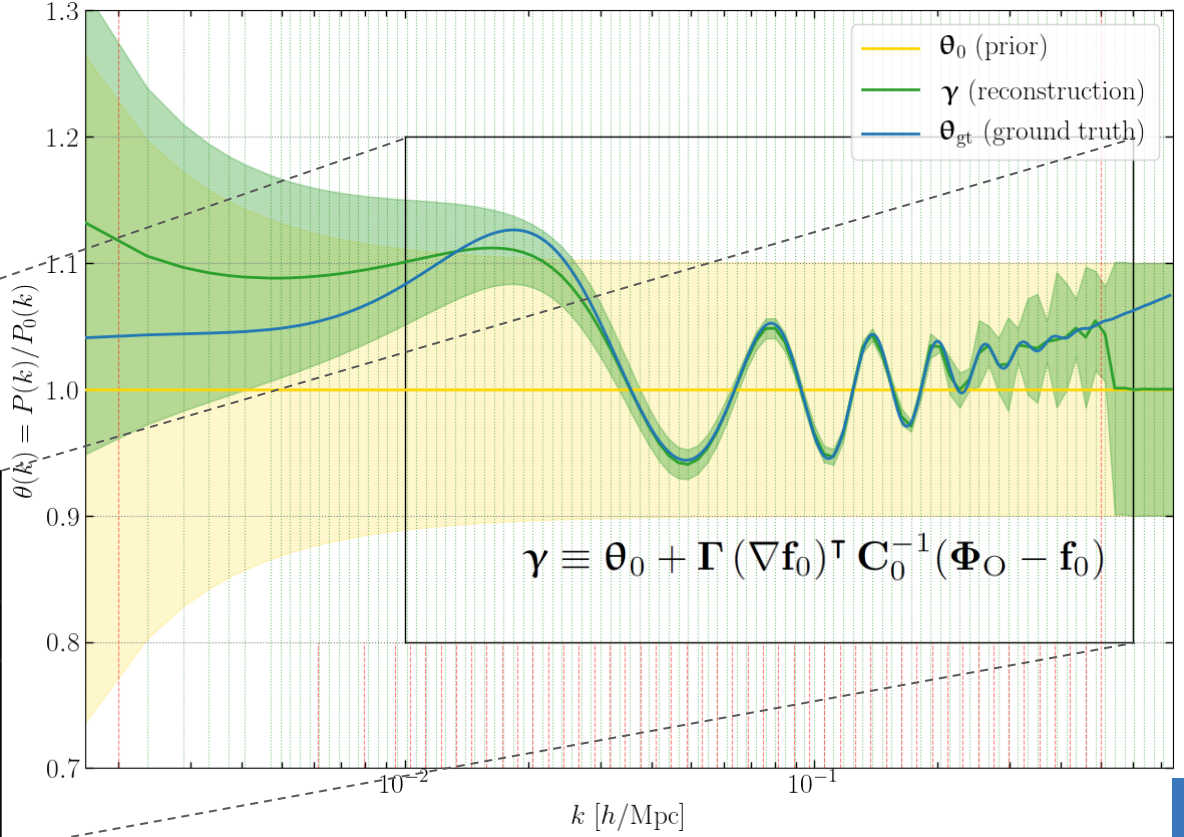
- $\mathbf{f}_0, \mathbf{C}_0$ and $\nabla \mathbf{f}_0$ can be evaluated through simulations only.
- The number of required simulations is fixed *a priori* (contrary to MCMC).
- The workload is perfectly parallel.

FL, Enzi, Jasche & Heavens, 1902.10149

SELFI-1 Euclid forecast (cosmic variance limit)

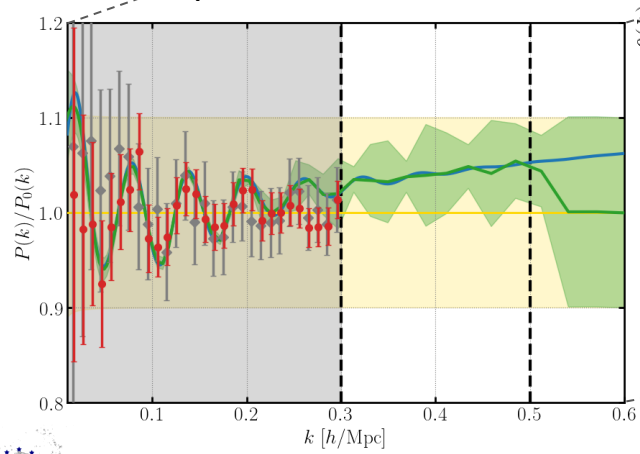
- $V = (3780\ \mathrm{Mpc}/h)^3$ (volume of the Euclid flagship simulation)
- Gaussian random field data model; 6,060 simulations
- 100 parameters are simultaneously inferred

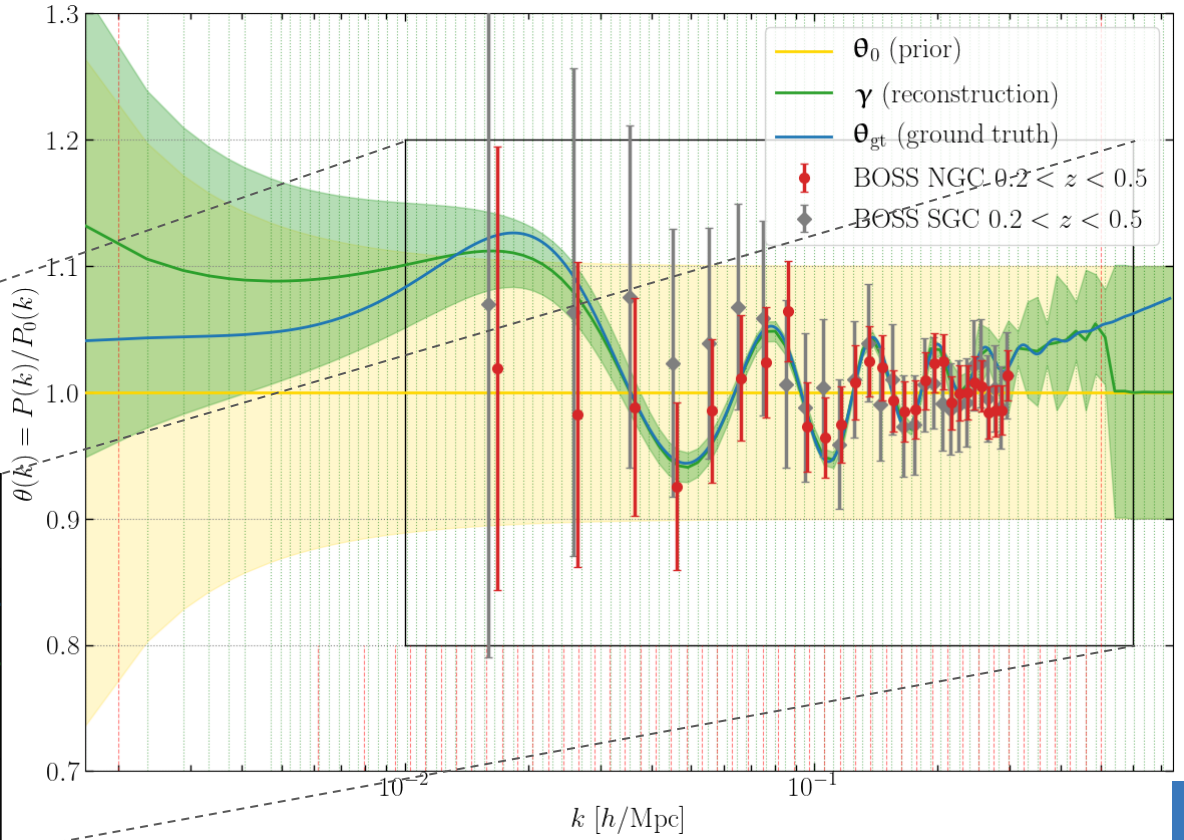




SELFI-1 Euclid vs BOSS

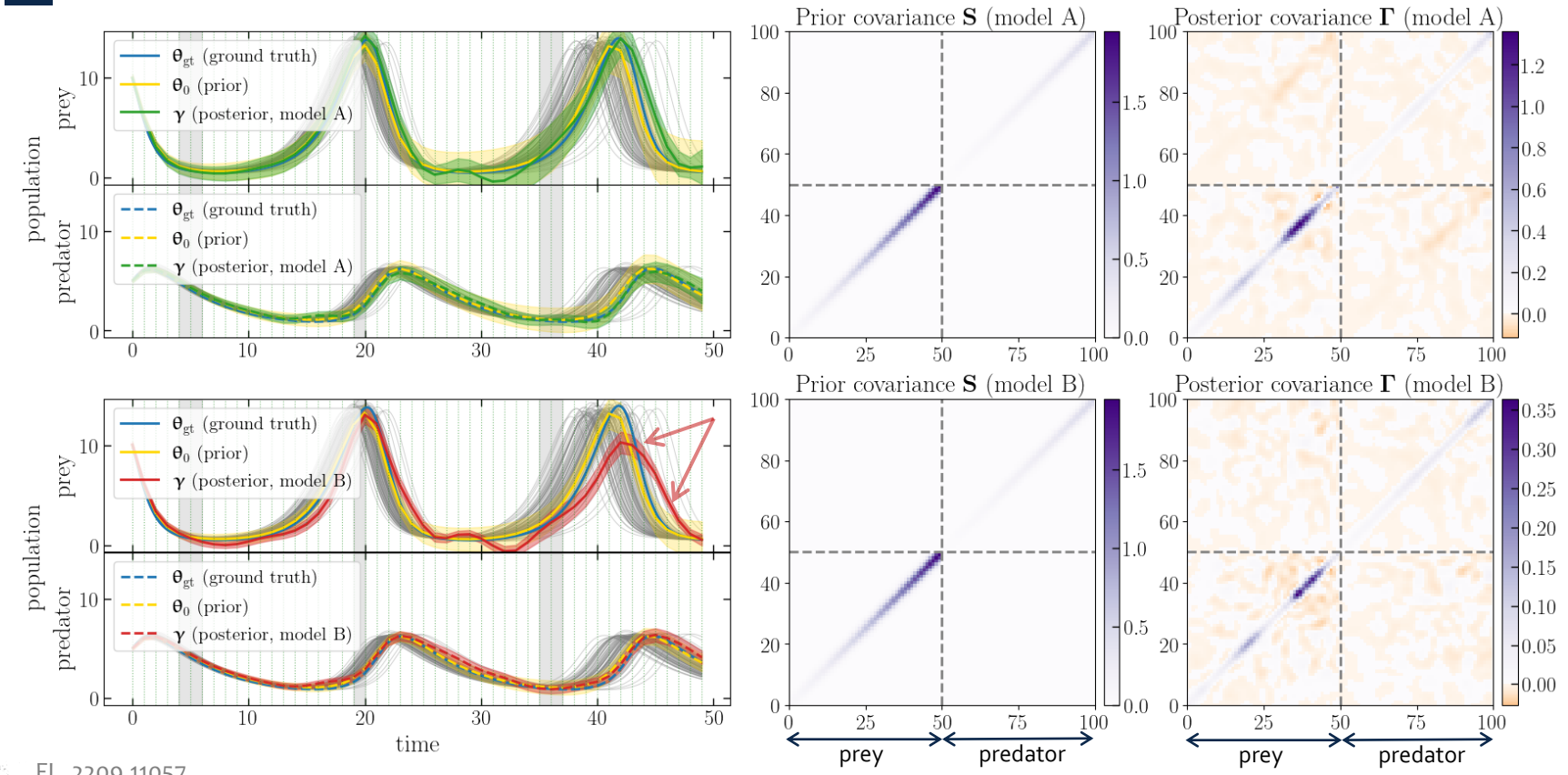
- Numerical data models allow using the galaxy power spectrum as summary statistics up to at least $k \gtrsim 0.5\ h/\mathrm{Mpc}$ safely
- $N_{\mathrm{modes}} \propto k^3$: 5 times more modes are used in the analysis.





Data points from Beutler *et al.*, 1607.03149

SELFI with model misspecification: inference of latent population functions of a prey-predator model



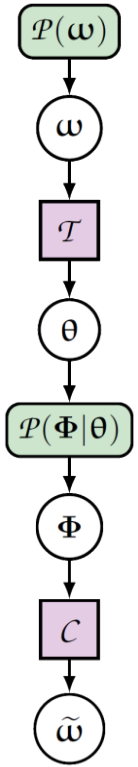
FL, 2209.11057

Check for model misspecification and data compression for SBI

- Qualitatively: the shape of the reconstructed $\boldsymbol{\theta}$ is useful as a check for model misspecification (independent theoretical understanding).
- Quantitatively: we can use the Mahalanobis distance between the reconstruction $\boldsymbol{\gamma}$ and the prior distribution $\mathcal{P}(\boldsymbol{\theta})$:

$$d_{\mathrm{M}}(\boldsymbol{\gamma}, \boldsymbol{\theta}_0|\mathbf{S}) \equiv \sqrt{(\boldsymbol{\gamma}-\boldsymbol{\theta}_0)^\intercal\, \mathbf{S}^{-1} (\boldsymbol{\gamma}-\boldsymbol{\theta}_0)}$$

- In the example:
  - $d_{\mathrm{M}}(\boldsymbol{\gamma}, \boldsymbol{\theta}_0|\mathbf{S}) \approx 5.35$ for model A
  - $d_{\mathrm{M}}(\boldsymbol{\gamma}, \boldsymbol{\theta}_0|\mathbf{S}) \approx 12.54$ for model B
  - $\langle d_{\mathrm{M}}(\mathcal{T}(\boldsymbol{\omega}), \boldsymbol{\theta}_0|\mathbf{S})\rangle \approx 9.43$ in fiducial simulations



- The score function $\nabla_{\boldsymbol{\omega}} \hat{\ell}_{\boldsymbol{\omega}0}$ is the gradient of the log-likelihood at fiducial point $\boldsymbol{\omega}_0$ in parameter space.
- A quasi maximum-likelihood estimator for the parameters is

$$\mathcal{C}(\boldsymbol{\Phi}) = \tilde{\boldsymbol{\omega}} \equiv \boldsymbol{\omega}_0 + \mathbf{F}_0^{-1}\left[(\nabla_{\boldsymbol{\omega}}\mathbf{f}_0)^\intercal \mathbf{C}_0^{-1}(\boldsymbol{\Phi} - \mathbf{f}_0)\right]$$

Fisher matrix: $\mathbf{F}_0 = (\nabla_{\boldsymbol{\omega}}\mathbf{f}_0)^\intercal \mathbf{C}_0^{-1} \nabla_{\boldsymbol{\omega}}\mathbf{f}_0$

$$\nabla_{\boldsymbol{\omega}}\mathbf{f}_0 = \nabla\mathbf{f}_0 \cdot \nabla_{\boldsymbol{\omega}}\mathcal{T}_0$$

Already computed for SELFI

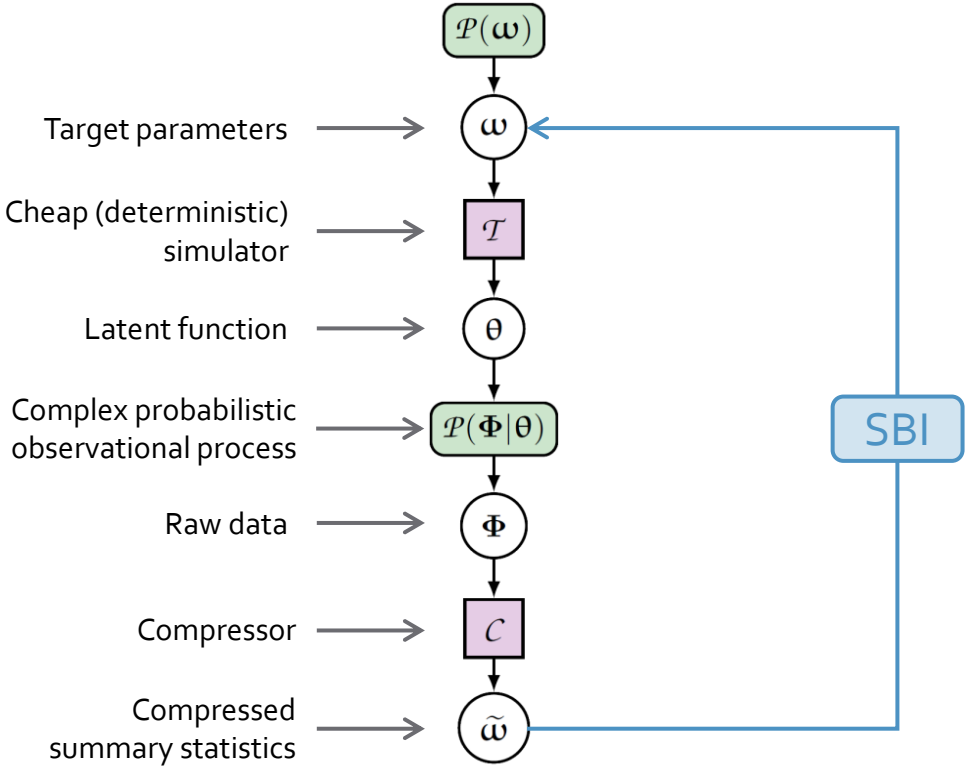
Cheap via finite differences

- Score compression is optimal in the sense that it preserves the Fisher information content of the data.

Alsing & Wandelt, 1712.00012

FL, 2209.11057

Simulation-based inference of top-level target parameters

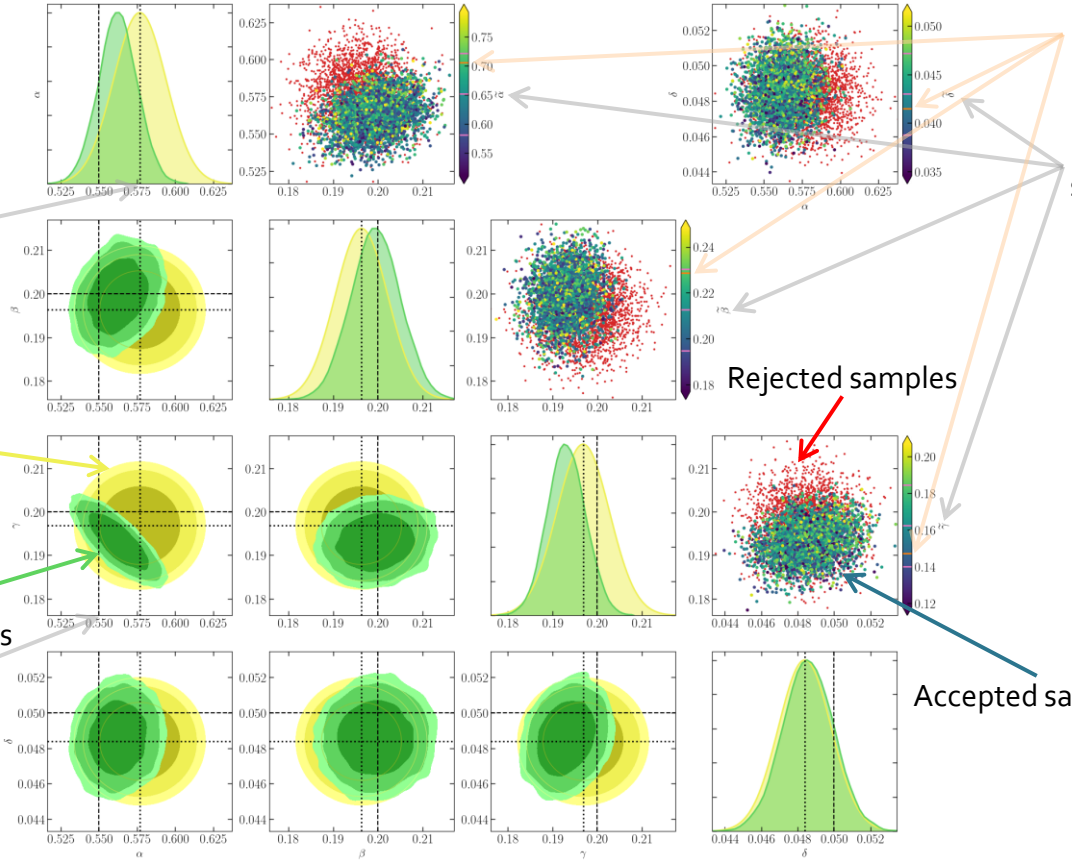


- Any SBI algorithm can be used to obtain the posterior $\mathcal{P}(\boldsymbol{\omega}|\tilde{\boldsymbol{\omega}}_{\rm O})$.
- Final inference:
  - does not depend on the assumptions made to check for model misspecification,
  - is unbiased (only more conservative) in case data compression is lossy.
- Non-parametric approaches can use the Fisher-Rao distance between simulated summaries $\tilde{\boldsymbol{\omega}}$ and observed summaries $\tilde{\boldsymbol{\omega}}_{\rm O}$:

$$d_{\rm FR}(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\omega}}_{\rm O}) \equiv \sqrt{(\tilde{\boldsymbol{\omega}} - \tilde{\boldsymbol{\omega}}_{\rm O})^{\intercal}\, \mathbf{F}_0 (\tilde{\boldsymbol{\omega}} - \tilde{\boldsymbol{\omega}}_{\rm O})}$$

FL, 2209.11057

Prey-predator model: inference of target population parameters using Likelihood-Free Rejection Sampling



Compressed observed summaries

Compressed simulated summaries

Fiducial point for score compression

Prior contours

Posterior contours

Ground truth

Rejected samples

Accepted samples: $d_{\mathrm{FR}}(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\omega}}_{\mathrm{O}}) < \varepsilon$
$\varepsilon = 2$

FL, 2209.11057

Conclusion: the statistical framework is in place for the GC:AP pipeline

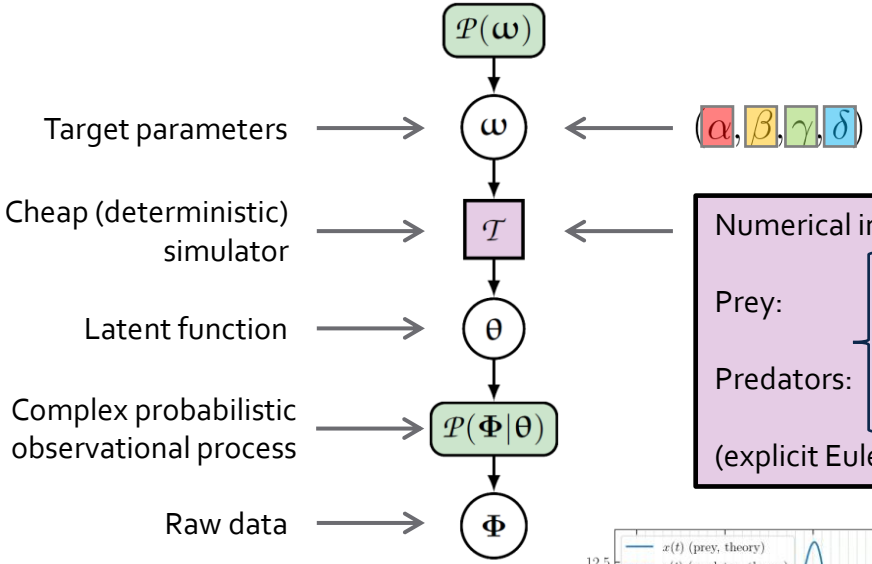
- A novel two-step simulation based Bayesian approach, combining SELFI and SBI, to tackle the issue of model misspecification for a large class of BHMs.
- Advantages of the first step (SELFI):
  - Even if the inference is in high dimension, the simulator remains a black-box.
  - The number of simulations is fixed *a priori* by the user.
  - The computational workload is perfectly parallel.
  - The linearised data model is trained once and for all independently of the data vector (amortisation).
- Advantages of the second step (SBI):
  - SELFI quantities provide a score compressor for free.
  - General advantages of SBI with respect to likelihood-based methods are preserved.
  - Inference does not depend on the assumptions made to check for model misspecification.

➢ A computationally efficient and easily applicable framework to perform SBI of BHMs while checking for model misspecification.

pySELFI is publicly available at https://pyselfi.florent-leclercq.eu.

Additional slides

A prey-predator model with observational effects



$(\alpha, \beta, \gamma, \delta)$

Intrinsic reproduction rate of prey

Mortality rate of prey due to predators encountered

Reproduction rate of predators due to preys eaten

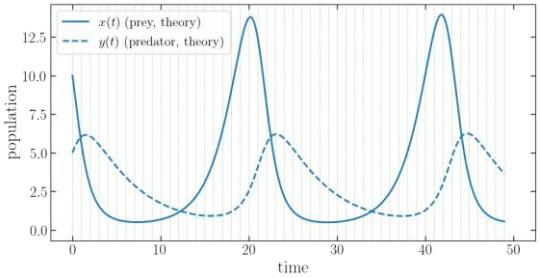
Intrinsic mortality rate of predators

Numerical integration of the Lotka-Volterra equations:

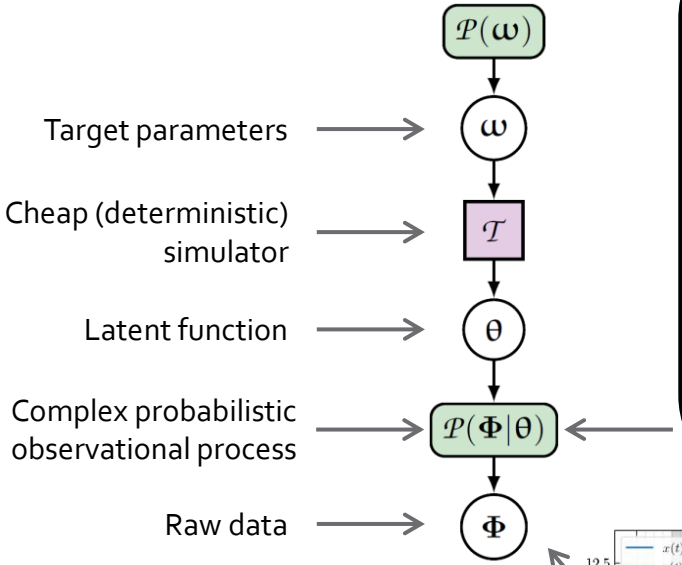
Prey: $\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - \beta xy$

Predators: $\frac{\mathrm{d}y}{\mathrm{d}t} = \delta xy - \gamma y$

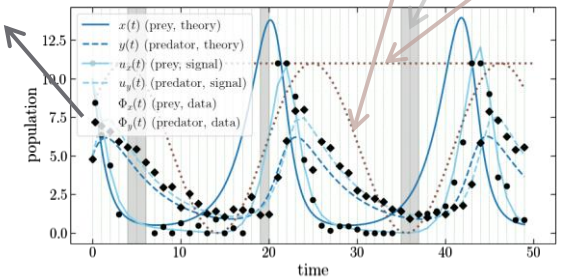
(explicit Euler method)

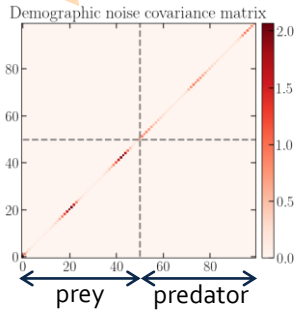


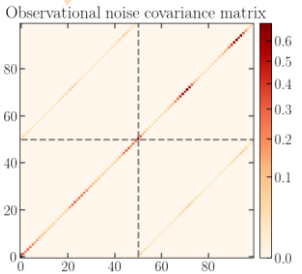
A prey-predator model with observational effects



- Model A (correct):
  - *Signal*: a delayed non-linear observation of the true functions
    $$\begin{cases} s_x(t_{i+1}) &= e_x(t_i) \quad [x(t_i) - p\,x(t_i)y(t_i) + q\,x(t_i)^2] \\ s_y(t_{i+1}) &= \quad\quad y(t_i) + p\,x(t_i)y(t_i) - q\,y(t_i)^2 \end{cases}$$
  - *Noise*: additive Gaussian noise with zero mean
    $$\begin{cases} u_x(t) = s_x(t) + n_x^{\mathrm{D}}(t) + n_x^{\mathrm{O}}(t) \\ u_y(t) = s_y(t) + n_y^{\mathrm{D}}(t) + n_y^{\mathrm{O}}(t) \end{cases}$$
    $$n_x^{\mathrm{D}}(t) \curvearrowleft \mathcal{G}[0, r\,x(t)] \qquad \begin{pmatrix} n_x^{\mathrm{O}}(t) \\ n_y^{\mathrm{O}}(t) \end{pmatrix} \curvearrowleft \mathcal{G}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, s\begin{pmatrix} \frac{y(t)}{x(t)y(t)} & t\sqrt{x(t)y(t)} \\ t\sqrt{x(t)y(t)} & x(t) \end{pmatrix}\right]$$
    $$n_y^{\mathrm{D}}(t) \curvearrowleft \mathcal{G}[0, r\,y(t)]$$
  - *Censoring*: a mask and a threshold
- Model B (misspecified): assumes the true functions are directly observed, does not account for observational noise nor threshold.



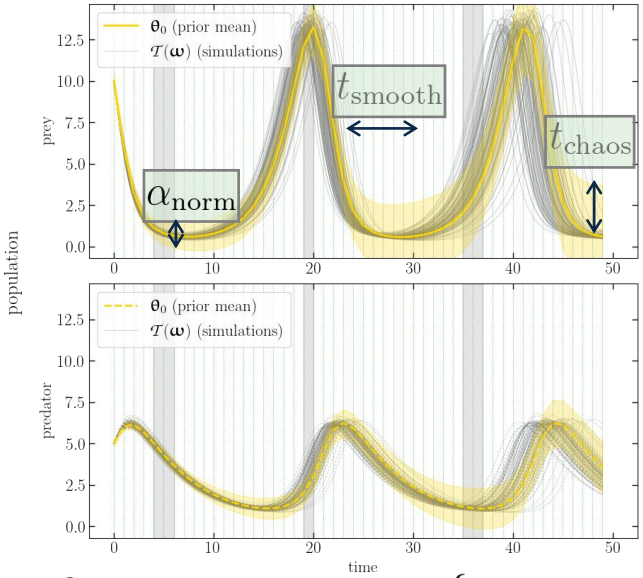




A family of priors for population functions in prey-predator systems

Assumptions:

1. The population functions $\boldsymbol{\theta}$ are Gaussian-distributed.
2. They are strongly constrained to live close to $\boldsymbol{\theta}_0 = \mathcal{T}(\boldsymbol{\omega}_0)$.
3. $x(t)$ and $y(t)$ are smooth functions of time.
4. The uncertainty on $x(t)$ and $y(t)$ grows with time.



➢ Gaussian prior:

mean: $\boldsymbol{\theta}_0$

covariance: $\mathbf{S} \equiv \alpha_{\mathrm{norm}}^2 \mathbf{K} \circ \mathbf{V}$ (Overall prior uncertainty)

- The 3 free hyperparameters $\{\alpha_{\mathrm{norm}}, t_{\mathrm{smooth}}, t_{\mathrm{chaos}}\}$ can be optimised using simulations.

$$(\mathbf{K}_z)_{ij} \equiv \left[-\frac{1}{2}\left(\frac{t_i - t_j}{t_{\mathrm{smooth}}}\right)^2\right] \quad \text{(Smoothness of the population function)}$$

$$\mathbf{K} \equiv \begin{pmatrix} \mathbf{K}_x & 0 \\ 0 & \mathbf{K}_y \end{pmatrix} \qquad \mathbf{V} \equiv \begin{pmatrix} x_0\,\mathbf{u}\mathbf{u}^\intercal & 0 \\ 0 & y_0\,\mathbf{u}\mathbf{u}^\intercal \end{pmatrix} \qquad (\mathbf{u})_i \equiv 1 + \frac{t_i}{t_{\mathrm{chaos}}}$$

(Chaotic behaviour of the system)