



# Dealing with systematic effects: the issue of robustness to model misspecification



Florent Leclercq

[www.florent-leclercq.eu](http://www.florent-leclercq.eu)

Institut d'Astrophysique de Paris  
CNRS & Sorbonne Université

In collaboration with Tristan Hoellinger (IAP)  
and the Aquila Consortium

[www.aquila-consortium.org](http://www.aquila-consortium.org)

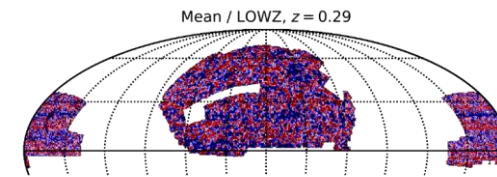
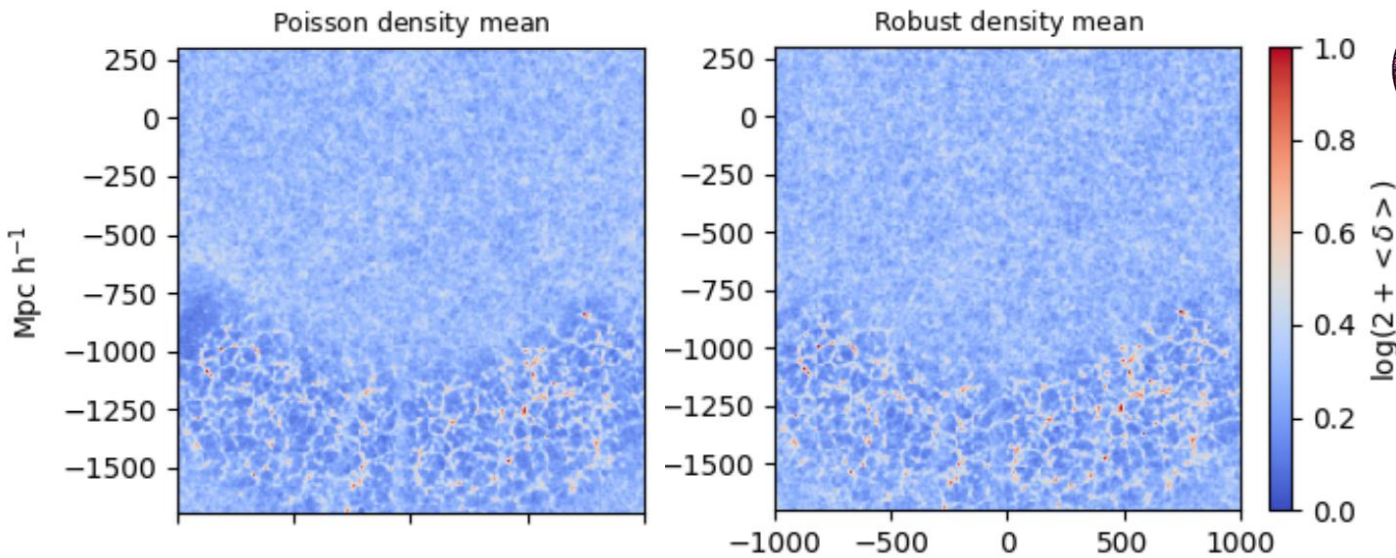
**28 November 2023**



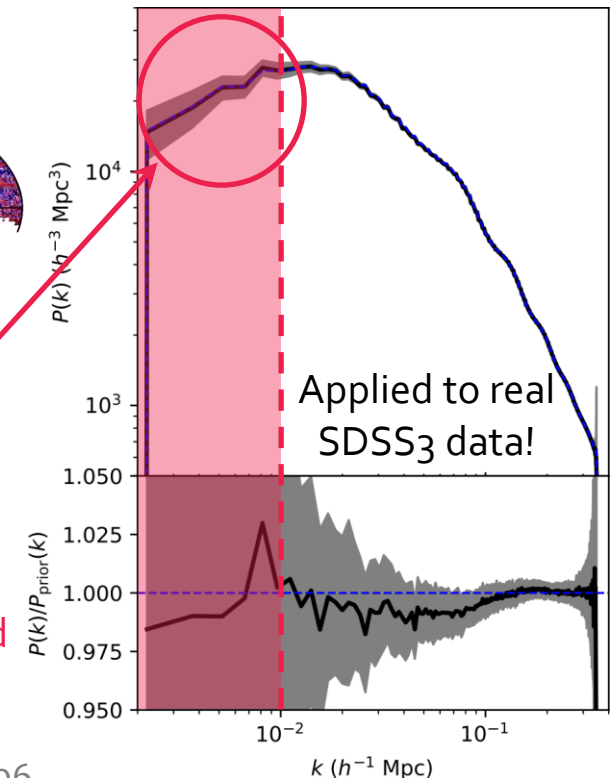
# Model misspecification and unknown systematics with an explicit field-level likelihood

- [Model misspecification](#) is a long-standing problem for Bayesian inference: when the model differs from the actual data-generating process, posteriors tend to be biased and/or overly concentrated.
- This issue is particularly critical for cosmological data analysis in the presence of [systematic effects](#).

- In cosmology, we are sometimes unable to formulate **any** model that fits the data in some regimes.
- Machine-aided report of unknown systematic effects is possible with an [explicit field-level likelihood](#) (BORG):



No apparent contamination, even well beyond the turn-over



Porqueres, Ramanah, Jasche & Lavaux, 1812.05113

Lavaux, Jasche & FL, 1909.06396



Florent Leclercq

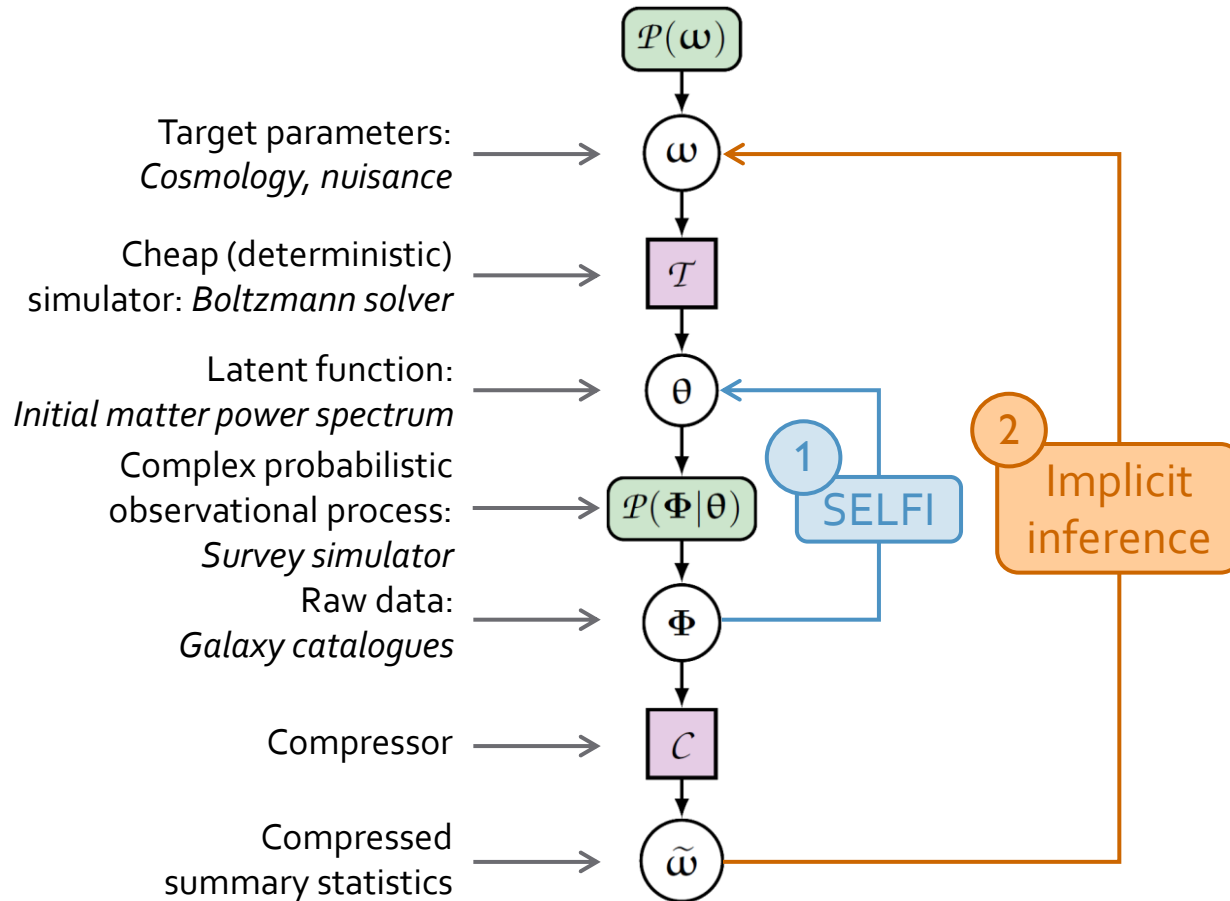
Dealing with systematic effects: the issue of robustness to model misspecification

28/11/2023

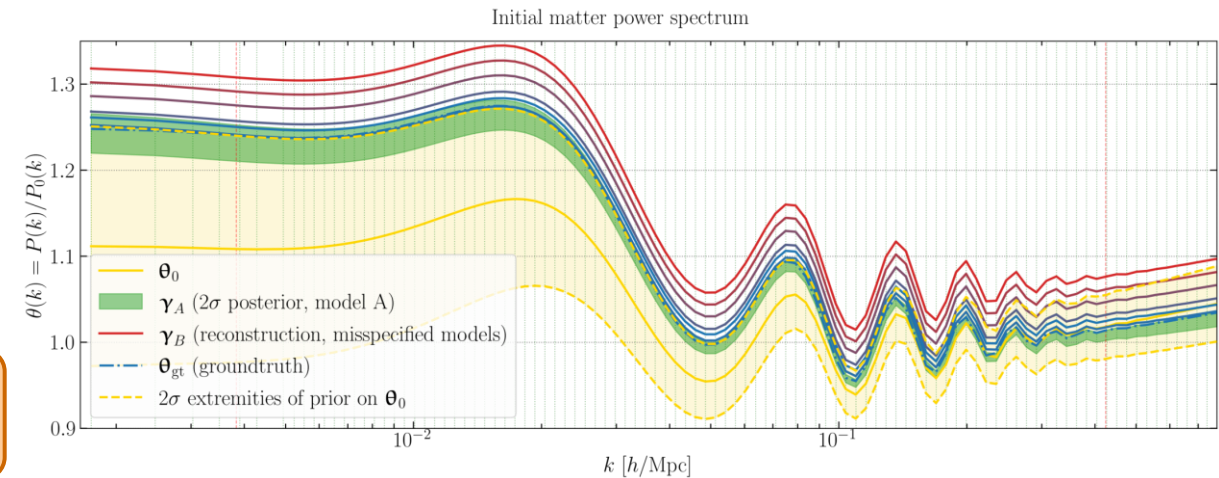
2

# Dealing with systematics in implicit inference problems: SELFI as a first step

- What about [implicit inference](#) problems? We propose a two-step framework:



- [SELFI](#) (*Simulator Expansion for Likelihood-Free Inference*) allows for initial matter power spectrum inference from black-box models.  
FL, Enzi, Jasche & Heavens, 1902.10149
- One can utilise the initial matter power spectrum to check for systematics.



Hoellinger & FL, in prep.

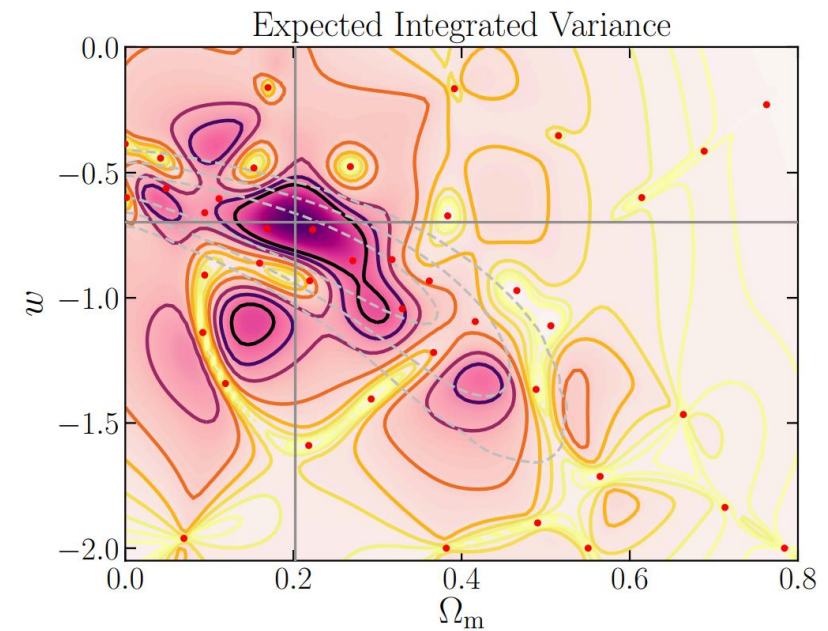
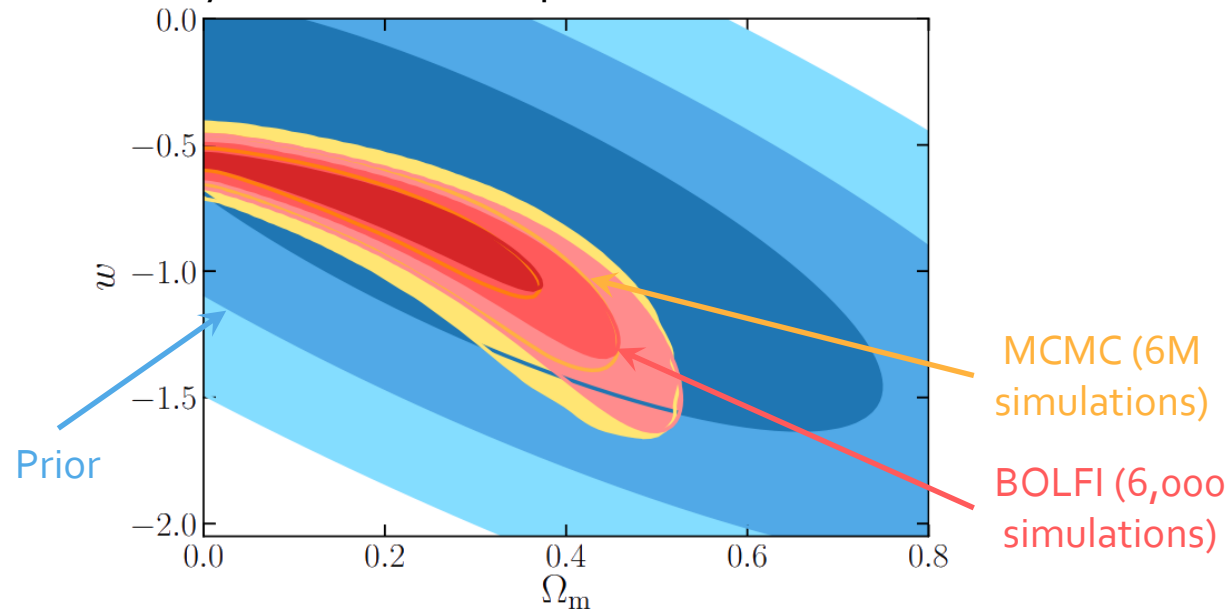
- After a model is selected, the problem can be solved by [implicit inference techniques](#).
- Bonus: the simulations used for [step 1](#) can be recycled to write a free [score compressor](#) for [step 2](#).  
FL, 2209.11057



# Dealing with expensive simulators in implicit inference problems: BOLFI

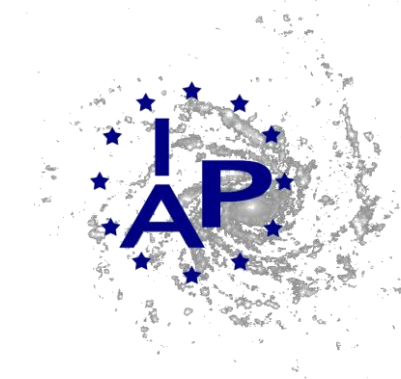
- The simulator will typically be extremely expensive ( $N$ -body simulation, halo finding, complex observational effects). We can typically afford  $O(10,000)$  evaluations.
- Emulation of the data model is not the only option.
- [BOLFI](#) (*Bayesian Optimisation for Likelihood-Free Inference*) uses an acquisition function to place expensive simulations in the parameter space.
- The optimal acquisition function for implicit inference can be derived: the [Expected Integrated Variance](#).

Re-analysis of the JLA supernovae data:



## References:

- [Leclercq 2018, 1805.07152](#), *Bayesian optimisation for likelihood-free cosmological inference*
- [Leclercq et al. 2019, 1902.10149](#), *Primordial power spectrum and cosmology from black-box galaxy surveys*
- [Leclercq 2022, 2209.11057](#), *Simulation-based inference of Bayesian hierarchical models while checking for model misspecification*
- Hoellinger & Leclercq, in prep.



<https://pyselfi.florent-leclercq.eu/>: publicly available implementation of SELFIE  
<https://aquila-consortium.org>

