



Constraints on local Primordial non-Gaussianities with large-scale structures

Laboratoire d'Annecy-le-Vieux de Physique Théorique

Marina S. Cagliari

Collaborators:

E. Castorina, M. Bonici, D. Bianchi, M. Barberi-Squarotti, K.
Pardede, G. D'Amico, A. Bairagi, B. Wandelt

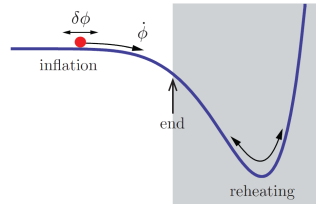
GGI - 09/29/2025

Inflation

Inflation is a **key ingredient of Λ CDM**.

It solves problems and makes predictions:

- large scales causally connected in the past,
- observed Universe is close to flat,
- spectral index,
- almost adiabatic fluctuations.



What is the dynamics of inflation?

Baumann (2009)

Inflation and f_{NL}

Local PNG, f_{NL} , give an insight into the inflation dynamics,

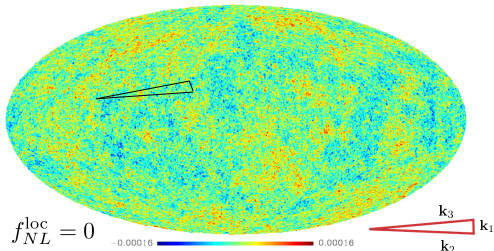
$$\lim_{k_1 \rightarrow 0} B_\zeta(k_1, k_2, k_3) = 4 f_{\text{NL}} P_\zeta(k_1) P_\zeta(k_3).$$

They are zero in **single-field inflation models**, and $f_{\text{NL}} \sim 1$ in **multi-field models**:

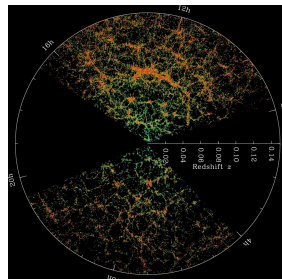
- detection $\rightarrow$ rule out single field models,
- non-detection of $f_{\text{NL}} \sim 1 \rightarrow$ constraints on multi-field models.

To learn something we have to aim for $\sigma_{f_{\text{NL}}} \lesssim 1$.

Maldacena (2003)
Creminelli and Zaldarriaga (2004)

f_{NL} probes

$$\sigma_{f_{\text{NL}}}^{\text{CMB}} \sim 5$$



$$\sigma_{f_{\text{NL}}}^{\text{LSS}} \sim 10$$

Planck+ (2019), SDSS, Castorina+ (2019), Cabass+ (2022), D'Amico+ (2022), Chaudisson+ (2024)

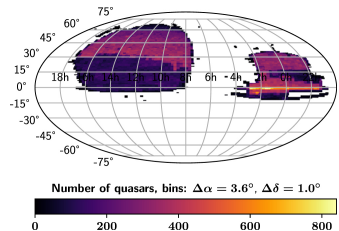
Old and new surveys

The new-generation surveys are here:

- DESI DR1 data public,
- Euclid DR1 data public in one year and a half.

Use **old-generation surveys** to learn how to **extract information from the new data**.

- Results from eBOSS DR16 quasar sample:
 - 343 708 QSOs,
 - $0.8 < z < 2.2$



DESI Collaboration (2016)
Euclid Collaboration (2024)
Ross+ (2020)
Lyke+ (2020)

How does f_{NL} effect LSS?

For $f_{\text{NL}} \neq 0$ the variance of the short-scale density modes is affected by the large scales,

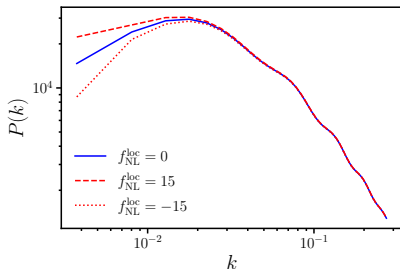
$$\phi = \phi_s + \phi_I$$

$$\langle \delta_s^2 \rangle^{1/2} = \sigma(1 + 2 f_{\text{NL}} \phi_I).$$

The coupling of ϕ_I and ϕ_s produce a **scale dependent** term in the **bias**:

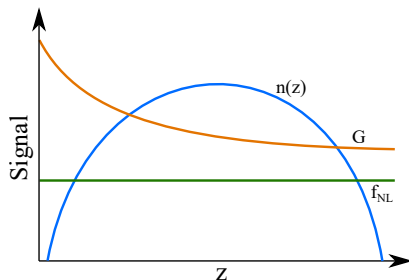
$$b_1 \rightarrow b_1 + f_{\text{NL}} b_\phi k^{-2},$$

with $b_\phi = b_1 - p$.



Dalal+ (2008)
Slosar+ (2008)

Optimal Weights



$$\delta_g \approx b_\phi \phi_p + b_1 \delta_m(z)$$

Tegmark+ (1998), Castorina+ (2019)

$$P(k) \propto \langle \delta^w(k) \delta^w(-k) \rangle .$$

We use a **unique optimal galaxy weighting scheme** for constraining f_{NL} from LSS observations:

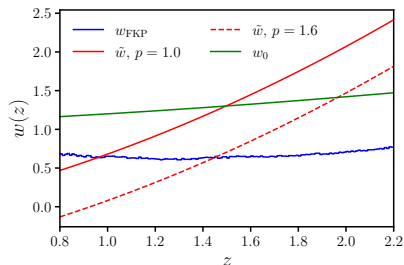
$$P(k) \propto \langle \tilde{\delta}(k) \delta_0(-k) \rangle ,$$

where z -dependent weights were applied to the galaxy field,

$$\tilde{w}(z) = (b(z) - p) ,$$

$$w_0(z) = D(z) \left(b(z) + \frac{f(z)}{3} \right) .$$

Optimal Weights



$$\delta_g \approx b_\phi \phi_p + b_1 \delta_m(z)$$

Tegmark+ (1998), Castorina+ (2019)

$$P(k) \propto \langle \delta^w(k) \delta^w(-k) \rangle .$$

We use a **unique optimal galaxy weighting scheme** for constraining f_{NL} from LSS observations:

$$P(k) \propto \langle \tilde{\delta}(k) \delta_0(-k) \rangle ,$$

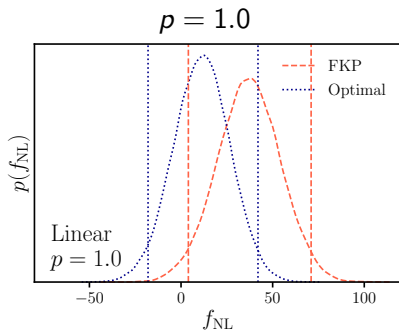
where z -dependent weights were applied to the galaxy field,

$$\tilde{w}(z) = (b(z) - p) ,$$

$$w_0(z) = D(z) \left(b(z) + \frac{f(z)}{3} \right) .$$

f_{NL} posterior

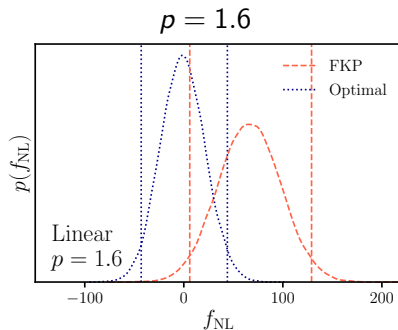
Cagliari+ (2023)



$$\text{FKP: } \sigma_{f_{\text{NL}}} \sim 17$$

$$\text{Optimal: } \sigma_{f_{\text{NL}}} \sim 15$$

11% improvement!



$$\sigma_{f_{\text{NL}}} \sim 31$$

$$\sigma_{f_{\text{NL}}} \sim 21$$

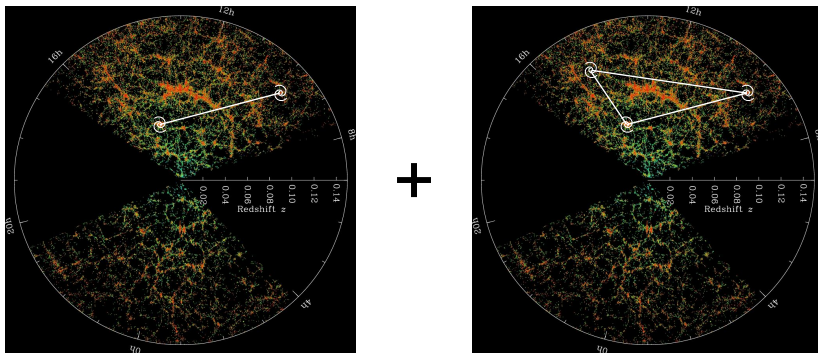
29% improvement!

How to improve the $P(k)$ results?

- Use larger surveys not shot-noise-dominated, like DESI and Euclid. However, larger surveys pose challenges in managing the **large-scale systematic effects**.
- Combine the 2-point statistics (power spectrum) with higher-order statistics, e.g. the bispectrum. In this case, we extract **more information from data we already have** and will be available in the future.

DISCLAIMER: We still need larger surveys!

Bispectrum: 3-point information

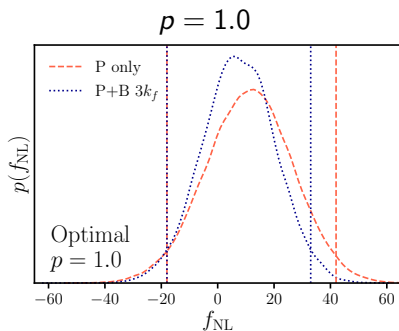


$$P_g(k) \propto \langle \delta_g(k) \delta_g(-k) \rangle,$$

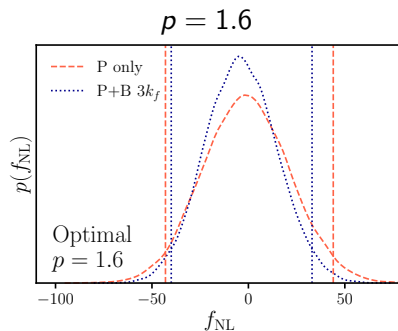
$$B_g(k_1, k_2, k_3) \propto \langle \delta_g(k_1) \delta_g(k_2) \delta_g(k_3) \rangle.$$

f_{NL} posterior

Cagliari+ (2025)

P only: $\sigma_{f_{\text{NL}}} \sim 15$ P+B: $\sigma_{f_{\text{NL}}} \sim 13$

13% improvement!

 $\sigma_{f_{\text{NL}}} \sim 21$ $\sigma_{f_{\text{NL}}} \sim 19$

10% improvement!

Can field-level information help?

Memory problems plague ML-based field-level inference.

Divide the survey in patches

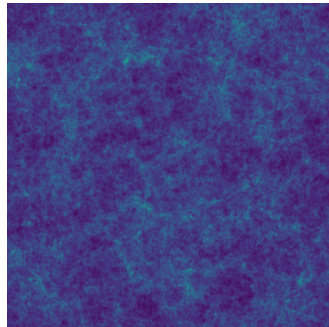
↓ CNN

Small scales: patch summary

Large scales: $P(k)$ and $B(k)$ of the box

↓ Combine

Get better constraints?



Bairagi & Wandelt (2025)

Can field-level information help?

Memory problems plague ML-based field-level inference.

Divide the survey in patches

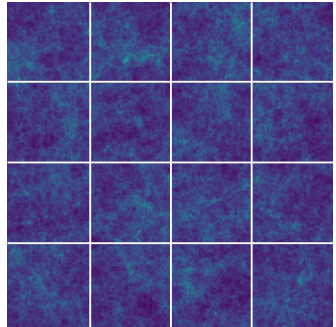
↓ CNN

Small scales: patch summary

Large scales: $P(k)$ and $B(k)$ of the box

↓ Combine

Get better constraints?



Bairagi & Wandelt (2025)

A Fisher information study



$$F_{ij} = \left(\frac{\partial \mathbf{x}}{\partial \theta_i} \right)^T \mathbf{C}^{-1} \left(\frac{\partial \mathbf{x}}{\partial \theta_j} \right) \longrightarrow \sigma_i = \sqrt{(F^{-1})_{ii}},$$

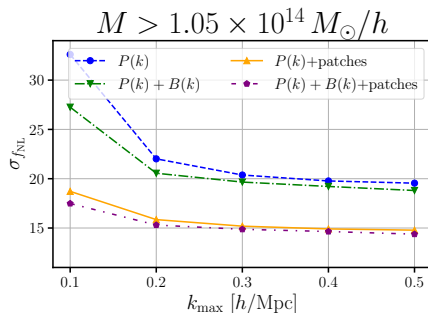
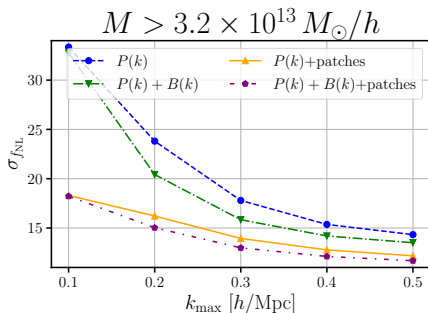
where $\mathbf{x}$ is the estimator and $\mathbf{C}$ its covariance matrix.

- I compute the Fisher information content of $\mathbf{x} = \{P, B, \text{patches}\}$ in **simulated boxes** at $z = 1$.
- I use the halo catalogues of the Quijote-PNG simulations.
- I divide the $1 (\text{Gpc}/h)^3$ simulation into $125^3 (\text{Mpc}/h)^3$ patches with resolution $8 \text{ Mpc}/h$.

Villaescusa-Navarro+ (2020)
Coultion+ (2023)

Forecast on f_{NL}

Cagliari+ (in prep.)



Conclusions

- I presented different ways to improve our current LSS analyses for f_{NL} :
 - Optimal $P(k)$: 29% improvement w.r.t. FKP $\rightarrow$ survey with double volume,
 - $P + B$: 38% improvement w.r.t. FKP $P(k) \rightarrow$ almost a survey with three time the current volume.
- Combining standard summary statistics with ML field-level data compression can improve constraints and reduce the modelling for the analyses.

Future:

- local PNG analysis for Euclid DR1.

Thanks for your attention!

Recent work on interlopers

2504.06919

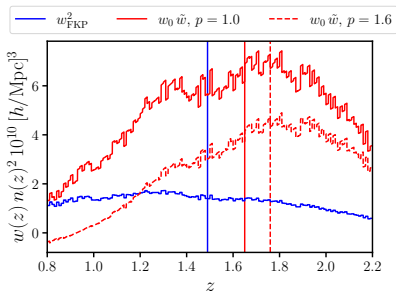
2509.02414

Contacts:

marina.cagliari@lapth.cnrs.fr

[github/mcagliari](https://github.com/mcagliari)

Effective redshift

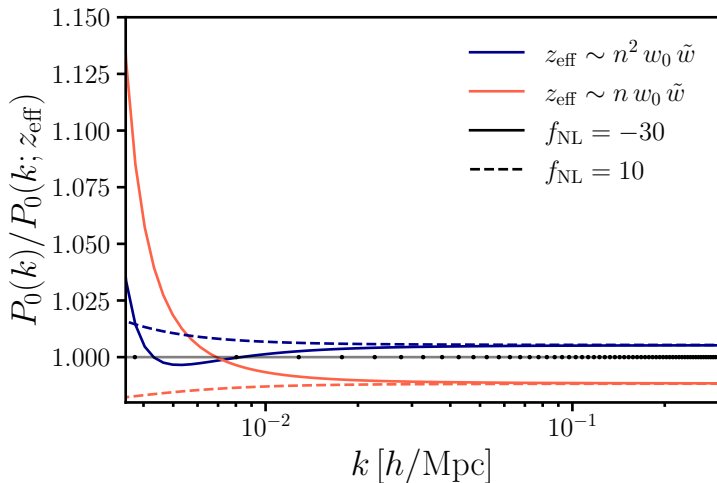


To not lose large-scale information along the line-of-sight we make a **full volume analysis**.

$$z_{\text{eff}} = \frac{\sum_{i=1}^{N_{\text{QSO}}} z_i n_i w_{c,i}^2 w_{\text{FKP},i}^2 \tilde{w}_i w_{0,i}}{\sum_{i=1}^{N_{\text{QSO}}} n_i w_{c,i}^2 w_{\text{FKP},i}^2 \tilde{w}_i w_{0,i}}.$$

$$z_{\text{eff}} = \begin{cases} 1.49 & \text{FKP,} \\ 1.65 & p = 1.0, \\ 1.76 & p = 1.6. \end{cases}$$

How good is the z_{eff} approximation?



Power spectrum

The signal is at low k and at high k dominated by the z error $\rightarrow$ linear theory is enough.

$$P_{\text{QSO}}(k, \mu, z_{\text{eff}}) = G(k, \mu; \sigma_{\text{FoG}})^2 (b_{\text{tot}}(k) + f \mu^2)^2 P_m(k) + N,$$

$$b_{\text{tot}}(k) = b_{\text{QSO}} + f_{\text{NL}}(b_{\text{QSO}} - p)\tilde{\alpha}(k).$$

$P_m(k)$ is computed in Planck2018 cosmology.

Window & Integral Constraint: the recipe

We are only interested in the monopole:

$$P_{\text{QSO}}^{(0)}(k) = Q_{k\,k_p}^{(0)} P_{k_p}^{(0)} - \frac{1}{5} Q_{k\,k_p}^{(2)} P_{k_p}^{(2)} + \frac{1}{9} Q_{k\,k_p}^{(4)} P_{k_p}^{(4)} + \dots$$

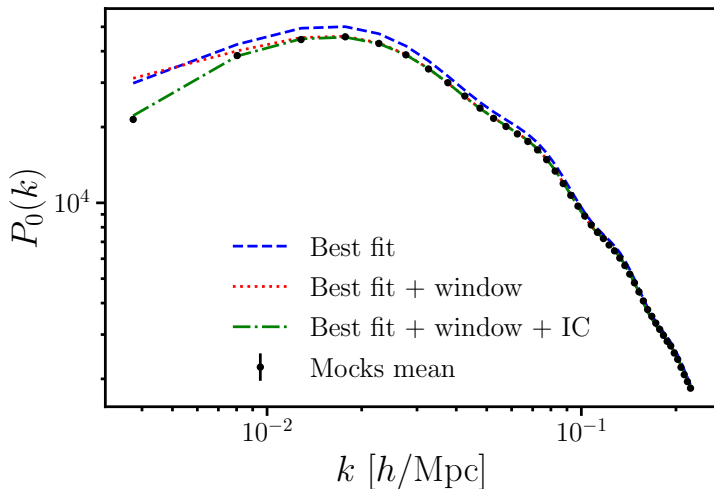
This needs to be corrected due to integral constraint effects:

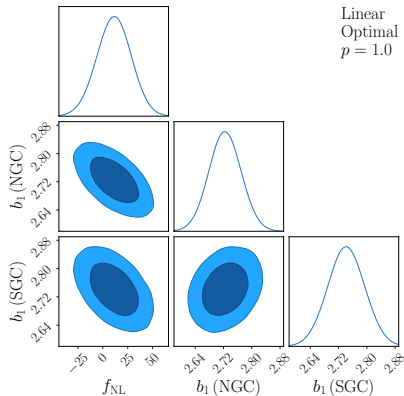
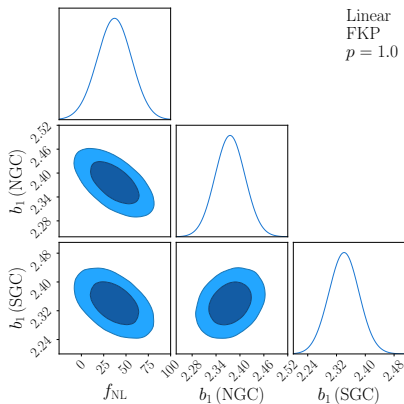
$$P_{\text{QSO}}^{(0),\text{IC}}(k) = P_{\text{QSO}}^{(0)}(k) - P_{\text{QSO}}^{(0)}(0) |W_0(k)|^2 - P_{\text{QSO}}^{(0)}(k) W_{\text{RIC}}(k).$$

$P_{\text{QSO}}^{(0),\text{IC}}(k)$ is what enters in the likelihood.

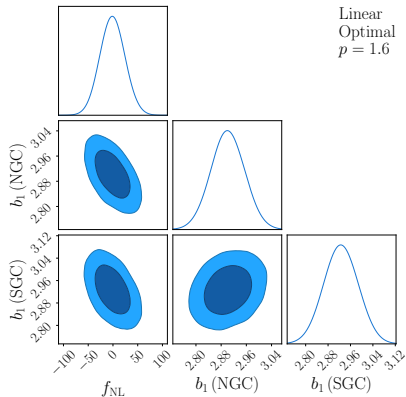
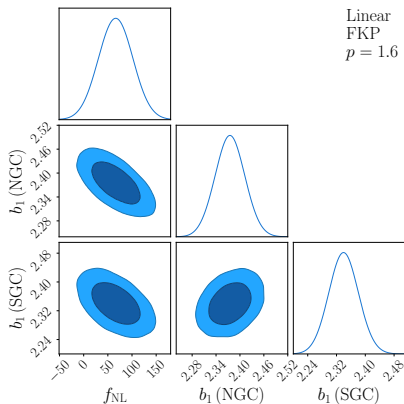
Wilson+ (2017)
De Mattia and Ruhlmann-Kleider (2019)

Everything together



Bias and f_{NL} correlation: $p = 1.0$ 

Bias and f_{NL} correlation: $p = 1.6$



Three-level bispectrum

$$B_{\text{QSO}}(k_1, k_2, k_3) = \mathcal{Z}_1(k_1)\mathcal{Z}_1(k_2)\mathcal{Z}_1(k_3)B(k_1, k_2, k_3) + \\ + 2\mathcal{Z}_1(k_1)\mathcal{Z}_1(k_2)\mathcal{Z}_2(k_1, k_2)P(k_1)P(k_2) + 2 \text{ perm},$$

where also $B(k_1, k_2, k_3)$ depends on the matter $P(k)$.

The window function and GIC are applied to the matter $P(k)$ s that enter the bispectrum model. The RIC is applied similarly to the $P(k)$ model:

$$B_{\text{QSO}}^{(0), \text{IC}}(k_1, k_2, k_3) = B_{\text{QSO}}^{(0), \text{GIC}}(k_1, k_2, k_3) - \\ - B_{\text{QSO}}^{(0)}(k_1, k_2, k_3) W_{\text{RIC}}^B(k_1, k_2, k_3).$$

Window approximation

