



Fisica

CORSO DI DOTTORATO DI RICERCA IN

XXXV

CICLO DEL CORSO DI DOTTORATO

*A new model for three-point statistics to probe Galaxy
Clustering in the nonlinear regime*

Massimo Guidi

Nome e cognome del dottorando

_____ firma

Prof. Enzo Branchini

Supervisor

_____ firma

Dr. Alfonso Veropalumbo

Co-supervisor

_____ firma

Prof. Giorgio Matt

Coordinatore

_____ firma



Ph.D. in Physics
35th Cycle

Ph.D. Thesis

A new model for three-point statistics to
probe Galaxy Clustering in the nonlinear
regime

Ph.D. candidate: Massimo Guidi

Supervisor: Prof. Enzo Branchini

Co-supervisor: Dr. Alfonso Veropalumbo

Acknowledgements

This doctoral Thesis is the result of three intense years, and I could not have accomplished it without the help and support of many people. I must extend my heartfelt thanks to my supervisor, Enzo Branchini, who I feel so fortunate to have met. He has guided me and helped me grow as a scientist. His professionalism and dedication to scientific research have been a constant source of motivation, and I have felt his trust throughout these years. This was a crucial factor in the success of this Thesis. I must also extend my gratitude to Alfonso Veropalumbo, whose support and competence have been invaluable; I am deeply grateful for his mentorship during these years. His contribution was essential to this Thesis. Furthermore, I must thank Alexander Eggemeier, a person of exceptional generosity and expertise, whose help was also vital to this Thesis.

I must also extend my thanks to Elena and Raul, who I had the pleasure of meeting during these years, and I am grateful for their support. I would also like to express my gratitude to Giulia, an unexpected acquaintance who added a different and brighter colour to my days and has surprised me with her deep understanding of me. I am thankful for your unwavering support. A special mention must also go to Marco, who was a constant source of support and with whom I spent a large portion of these years.

I am grateful to have shared this journey with Filippo, Simone, Elisabetta, Francesca, Romana, Vittoria, Ivano, Matteo, Alessio, Sacha, Paolo, Leonardo, Lorenzo, and Andrea. I must also thank Cintia, a deeply sensitive person who has made these years so much richer. I must also extend my thanks to Claudia, Pierluigi, Alessandro, and Ariel, who have made this journey even more enjoyable and unexpected. Thanks to Lorenzo, a lifetime friend.

Last but not least, I must acknowledge my family: my mother, Raffaella, my father, Umberto, my aunt, Serena, and my brother Simone. You all serve as beacons in my life, and although I may not have dedicated as much time to you as I would have liked, I have poured much of your support into this Thesis. This Thesis is dedicated to you.

Abstract

This doctoral Thesis investigates the modelling of the three-point correlation function of the Large Scale Structure of the universe. This represents a pioneering examination of the three-point correlation function in perturbative theory at the next-to-leading order. The objective of this modelling is to examine the clustering properties of dark matter perturbations, halo distributions, and galaxy distributions, with the goal of bridging the gap between configuration space and its Fourier space counterpart. The methodology employed is based on the 2D-FFTLLog algorithm, which inputs the multipoles of a bispectrum model and converts them into multiples of the three-point correlation function. The predictions of this novel modelling approach have been compared to measurements derived from simulated data in real space, covering typical redshift ranges of current and future large-scale surveys. In the first place, the focus has been on the improvement of the modelling of the three-point statistics of only matter perturbations. Secondly, in conjunction with two-point statistics, the predictions of the next-to-leading order three-point correlation function model of the galaxy distribution have been compared in the context of the determination of galaxy bias parameters. The results demonstrate a marked improvement in the agreement with the new model in terms of constraining second-order bias in the small-scale regime, with the agreement with established bias relations measured from N-body simulations used as a mean of model selection.

Keywords: cosmology: large-scale structure of Universe – theory – methods: statistical

Contents

List of Figures	ix
List of Tables	xvii
1 Introduction	3
1.1 The Λ CDM model	3
1.2 The Λ CDM model and the current debate	5
1.3 Motivations for this Thesis	7
2 Background cosmology	11
2.1 Friedmann-Lemaître-Robertson-Walker universe	11
2.1.1 The Friedmann equations	11
2.1.2 Cosmological redshift	13
2.1.3 Hubble's law	14
2.2 Energy forms in a Friedmann-Lemaître-Robertson-Walker Universe	15
2.2.1 Radiation	17
2.2.2 Matter	18
2.2.3 Cosmological constant	18
2.3 The inflationary paradigm	19
3 Perturbation Theory	25
3.1 Primordial perturbations	26
3.2 Gravitational instability	28
3.2.1 Fluid equations	28
3.2.2 Linear growth	30
3.2.3 Standard perturbation theory	31
3.2.4 Lagrangian perturbation theory	32
3.3 From matter to galaxies	34

3.3.1	Halo model	34
3.3.2	Perturbative bias	39
4	Matter and galaxy clustering statistics	43
4.1	Correlation functions of density fluctuations	43
4.1.1	Statistical description	44
4.1.2	Two-point statistics	45
4.1.3	Higher-order statistics	47
4.2	Modeling clustering statistics from perturbation theory	49
4.2.1	Modelling two- and three- point statistics in Fourier space: the matter case	50
4.2.2	The effective field theory approach for modelling for the matter power spectrum and bispectrum: the matter case	56
4.2.3	Modelling two- and three- point statistics in Fourier space: the galaxy case	58
5	Modelling and measuring two- and three- point statistics in configuration space	67
5.1	Configuration space vs. Fourier space	67
5.2	Perturbative model for the two-point correlation function	69
5.3	Perturbative model for the three-point correlation function	70
5.4	Measuring the three-point correlation function	72
5.5	Modelling the three-point correlation function in the small-scale regime: the 2D-FFTLog algorithm	75
5.5.1	2D-FFTLog technique	76
5.5.2	2D-FFTLog calibration	79
5.6	Model vs. simulated data: evaluating statistical errors	80
5.6.1	Numerical covariance	80
5.6.2	Theoretical covariance	81
5.7	The three-point correlation function pipeline	83
6	Testing the next-to-leading order model for the matter	87
6.1	3PCF matter models	87
6.2	The DEMNUni simulations	88
6.2.1	Dark matter clustering measurements	89
6.2.2	Gaussian covariance	90
6.3	Comparing models to data	91
6.4	Results	92

6.4.1	Comparisons among 3PCF models	93
6.4.2	3PCF measurements	96
6.4.3	3PCF models vs. data: overview	96
6.4.4	3PCF models vs. data: the BAO scale	99
6.4.5	3PCF models vs. data: quantitative analysis	101
7	Constraining galaxy bias with the Euclid Spectroscopic Survey	107
7.1	Modeling galaxy bias and clustering statistics	107
7.2	Dataset	108
7.2.1	Measurements	109
7.2.2	Gaussian covariance	110
7.3	Results	110
7.3.1	Clustering measurements	110
7.3.2	Parameter inference	112
7.3.3	Bias constraints from the next-to-leading order 2PCF and leading order 3PCF alone	114
7.3.4	Bias constraints from the joint 2PCF + 3PCF analysis	115
8	Bias loop corrections in configuration space: a joint 2PCF + 3PCF analysis	121
8.1	Models	121
8.2	Dataset	122
8.2.1	Measurements	123
8.2.2	Covariance	123
8.3	Results	124
8.3.1	Clustering measurements	124
8.3.2	Parameter inference	124
8.3.3	Bias constraints from 2PCF and 3PCF measurements alone	127
8.3.4	Bias constraints from the joint 2PCF and 3PCF analysis: next-to- leading order models	131
	Conclusions	135
A	Appendix	141
A.1	Theoretical covariance: two- and three- point correlation functions	141
A.2	Bias loop corrections to the Bispectrum: kernels	144
A.3	Bias loop corrections to the Bispectrum: numerical integration	144
	References	157

List of Figures

1.1	The content of the universe in various forms of energy in the Λ CDM model, as measured by CMB observations and large-scale structure surveys (1; 2; 3; 4; 5). Source: Max-Planck-Institute for Astrophysics Garching and Pixabay.	4
1.2	The Hubble tension, including the first standard sirens measurement following GW170817 (6), Planck (3) and Hubble Space Telescope (HST) with GAIA DR2 (7)). Blue stars correspond to measurements of H_0 in the local universe with calibration based on Cepheids. Red dots refer to derived values of H_0 from the CMB, assuming Λ CDM. Green crosses are direct measurements of H_0 with standard sirens. Forecasts are CMB Stage IV (8), standard sirens (9) and distance ladder with full GAIA and HST (10). Figure from (11)	6
1.3	Marginalised posterior distribution in the $\sigma_8 - \Omega_m$ plane, comparing the 3×2 pt analyses from KiDS-1000 with BOSS and 2dFLenS (12), with the 3×2 pt analysis from DES Y1 (6), and the CMB constraints from Planck Collaboration (3). Figure from (12)	6
2.1	The CMB anisotropy as a function of ℓ and the angular scale from Planck's data (3).	21
3.1	Description of how non-linearities in the gravitation evolution, in the biasing, overlap to the initial condition set by inflation.	25
3.2	Schematisation of how the scale factor re-enters in the Hubble radius after the inflationary epoch. The red line refers to the scale factor and the green line to the Hubble radius in different epochs. Picture from (13).	28
3.3	Illustration of a two-dimensional Gaussian density field $\delta(R)$ smoothed on different scales $R : [0.4, 1, 2]h^{-1}\text{Mpc}$, respectively <i>top-left</i> , <i>top-right</i> , <i>bottom-left</i> . The <i>bottom right</i> panel refers to $R = 2h^{-1}\text{Mpc}$ along with the spherical collapse threshold. Picture from (14).	37

4.1	Figurative description of the three-point moment in terms of the unconnected (first four diagrams, from <i>left to right</i>) and connected (last diagram) parts. Picture from (15)	48
4.2	Slices of thickness $50h^{-1}\text{Mpc}$ of a mock galaxy distribution for SDSS (left) and a realisation of a Rayleigh-Lèvy flight (right). They have the same two-point correlation function, and they exhibit visible differences. Figure from (16)	48
4.3	Tree-level power spectrum diagram. P_{11} represents the diagrammatic term corresponding to P_L	51
4.4	One-loop power spectrum diagrams. Left diagram refers to $P_{13,m}$, right to $P_{22,m}$	52
4.5	Power spectrum at one-loop and single contributions, assuming ΛCDM cosmology from Planck (3).	53
4.6	Tree-level bispectrum diagram. $B_{211,m}$ represents the diagrammatic term corresponding to $B_{\text{tree-level},m}$	54
4.7	One-loop bispectrum diagrams. From right to left: $B_{222,m}$, $B_{321-I,m}$, $B_{321-II,m}$, $B_{411,m}$	55
4.8	Diagrammatic representation for the n -th multipoint propagator. The vertexes correspond to the n incoming momenta. The blob vertex corresponds to the diagrammatic representation of a vertex in a renormalised perturbation theory. The represented diagrams show the multipoint propagator involved in a generic correlation function up to two-loop plus higher order corrections.	63
4.9	Power spectrum at one-loop in the bias loops multipoint propagator approach. The possible arising terms at one-loop are represented by the combination of 1-loop and tree-level <i>one</i> -point propagators (left diagram) connected by one internal line, and two tree-level <i>two</i> -point propagators connected by two internal lines (right diagram). The tree-level power spectrum is recovered by two tree-level <i>one</i> -point propagators (left diagram) connected by one internal line.	63

4.10	Bispectrum at one-loop in the bias loops multipoint propagator approach. The possible arising terms at one-loop are represented by the combination of three <i>one</i> -point propagators (left diagram, two at three-level, one at one-loop) connected by two internal lines, three tree-level <i>two</i> -point propagators connected by three internal lines (middle diagram) and the same but connecting tree-level one <i>one</i> -point, <i>two</i> -point, <i>three</i> -point propagator (right diagram). The three-level bispectrum is recovered by three <i>one</i> -point propagators at three-level connected by two internal lines (left diagram).	64
5.1	Figurative description of the algorithmic approach presented in (17). The primary galaxy is positioned at X , and around it, spherical shells are used to divide the secondary galaxies into radial bins. The coefficients in Eq. 5.23 are then calculated through these radial bins.	74
5.2	Different computation of ζ for equilateral computation with different grid parameters using 1D-FFTLog. Left: plot of different evaluations. Middle: ratio of previous computations. Right: residuals of previous evaluations. . .	79
5.3	Different computation of ζ for equilateral computation with different grid parameters using 2D-FFTLog. Left: plot of different evaluations. Middle: residuals of previous computations. Right: difference between previous evaluations compared to theoretical error assuming theoretical covariance. .	80
5.4	Flow chart describing the strategy assumed in this Thesis.	83
6.1	Difference between LO-1D, considered as a benchmark, and LO-2D 3PCF models, $\Delta\zeta = \zeta^{\text{LO-2D}} - \zeta^{\text{LO-1D}}$ (blue curve), as a function of the triangle configurations identified by the side lengths. Top and central panels show the model predictions at $z = 1.05$ and at $z = 0.49$. The grey band represents the $1\text{-}\sigma$ Gaussian uncertainty. The bottom panel shows the different sides of the triangles as a function of the Triangle Index, using logarithmic y-axis. The boxed plot represents, using logarithmic x- and y- axes, a zoomed-in area of the bottom panel including only separation distances $r_{12} \leq 17.5h^{-1}\text{Mpc}$, providing a closer look at the region of very small scales.	93
6.2	Same as figure 6.1 showing the difference between 3PCF models LO-1D and NLO-SPT (red curves).	94

- 6.3 Relative difference between LO-1D, considered as a benchmark, and LO-2D 3PCF models, $\Delta\zeta = (\zeta^{\text{LO-2D}} - \zeta^{\text{LO-1D}})/\zeta^{\text{LO-1D}}$ (blue curve), as a function of the triangle configurations identified by the side lengths. Top and central panels show the relative errors between model predictions at $z = 1.05$ and at $z = 0.49$. The grey band represents the $1-\sigma$ propagated Gaussian uncertainty. The bottom panel shows the different sides of the triangles as a function of the Triangle Index, using logarithmic y-axis. The boxed plot represents, using logarithmic x- and y- axes, a zoomed-in area of the bottom panel including only separation distances $r_{12} \leq 17.5h^{-1}\text{Mpc}$, providing a closer look at the region of very small scales. 95
- 6.4 Measurements of 3PCF from the DEMNUni snapshots at $z = 0.49$ (top plots) and $z = 1.05$ (bottom plots). In each figure for different redshifts, the upper panel shows the 3PCF amplitude multiplied by the size of the sides of the triangles as a function of the triangle configuration, indicated on the X-axis. The bottom panel shows the signal-to-noise, assuming Gaussian errors. Different colours and symbols are used for different triangle types, as indicated in the labels. In our classification scheme, as described in the text, it is possible for a triangle to have multiple classifications. For instance, a triangle classified as squeezed BAO would also fall into the category of small-scale triangles. First and second panels refer to $z = 0.49$, while third and fourth panels $z = 1.05$. Bottom plot shows the sides of the triangles as a function of the Triangle Index, ranging from $r_{\min} = 17.5h^{-1}\text{Mpc}$ and $r_{\max} = 132.5h^{-1}\text{Mpc}$. The bin-width is $\Delta r = 5h^{-1}\text{Mpc}$ 97
- 6.5 Comparison between normalised absolute values of residuals of SPT-NLO and LO-2D models. The LO-2D residuals are displaced in the X-axis, while the SPT-NLO case is displaced in the Y-axis. The left plot refers to $z = 0.49$, while the right plot refers to $z = 1.05$. Different symbols focus on BAO scales (red triangles), small scales (green rhombuses) and small BAO scales (yellow stars). 98

- 6.6 *Top:* 3PCF as a function of the two side lengths r of a set of isosceles triangles, with the third side of length $\delta r = 22.5h^{-1}\text{Mpc}$. The 3PCF measurements are shown as black dots, and the error bars indicate the 1σ Gaussian uncertainty. The continuous red curve and the dashed blue ones show the predictions of the NLO-SPT of the LO-2D models, respectively. Left and right panels illustrate the results at $z = 0.49$ and at $z = 1.05$. *Bottom:* normalised residuals for $z = 0.49$ (left), $z = 1.05$ (right), as defined in Eq. 6.11, for the three models depicted by the same colours and shapes of the top panel. 100
- 6.7 Best fit values for the four NLO-EFT free parameters (left to right) from the reduced χ^2 analysis, as a function of r_{\min} . The red continuous band refers to $\eta_{\min} = 2$, the blue dotted band to $\eta_{\min} = 0$. Top and bottom panels show the results for the $z = 0.49$ and $z = 1.05$ snapshots, respectively. 101
- 6.8 Cumulative reduced chi-square χ_r^2 in the range $[r_{\min}, 112.5]h^{-1}\text{Mpc}$, as a function of r_{\min} for the case $\eta_{\min} = 2$. continuous red and dashed blue curves refer, respectively, to the NLO-SPT and LO-2D models. The grey band represents a 99.7% confidence level assuming Gaussian statistics. The left and the right panels refer, respectively, to redshift $z = 0.49$ and $z = 1.05$ 102
- 6.9 Same as Fig. 6.8 fore the case $\eta_{\min} = 0$. Model predictions for NLO-EFT, NLO-SPT and LO-2D are shown with green dot-dashed, reds continuous and blue dashed curves, respectively. 103
- 6.10 Cumulative chi-square difference between NLO-SPT and NLO-EFT models divided by the NLO-SPT degree of freedom. Dotted lime line refers to $z = 1.05$, the solid green one refers to $z = 0.49$. From left to right, respectively, the case $\eta_{\min} = 2$ and $\eta_{\min} = 0$. The grey band represents a 99.7% confidence level coming from the propagation of errors on Eq. 6.10. 104
- 7.1 Measurements of 2PCF extracted from four Flasghip snapshots. For each panel, the upper and the bottom plots refer to the measurement and the signal-to-noise ratio (S/N). *Top-left, top-right, bottom-left, bottom-left* plots refer respectively to $z = 0.90, 1.19, 1.53, 1.78$, the bin-width is $\Delta r = 5h^{-1}\text{Mpc}$. . 111
- 7.2 Measurements of 3PCF extracted from four Flasghip snapshots. *Top-left, top-right, middle-left, middle-left* panels show measurements and signal-to-noise ratio (S/N), respectively at $z = 0.90, 1.19, 1.53, 1.78$. *Bottom* panel shows the sides of the triangles as a function of the Triangle Index, ranging from $r_{\min} = 5 h^{-1}\text{Mpc}$ to $r_{\max} = 145 h^{-1}\text{Mpc}$. The bins-width is $\Delta r = 10h^{-1}\text{Mpc}$ 112

- 7.3 Contour plots of next-to-leading order 2PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (blue, green, red). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 1.78, 1.53, 1.19, 0.90$. The maximum scale for 2PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$ 115
- 7.4 Contour plots of leading order 3PCF fixing $\eta_{\min} = 3$ and varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (blue, green, red). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 0.90, 1.19, 1.53, 1.78$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73 116
- 7.5 Contour plots of leading order 3PCF fixing $r_{\min} = 20, 30, 30, 30h^{-1}\text{Mpc}$ varying $\eta_{\min} = 3, 2, 1$ (blue, orange, purple). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 1.78, 1.53, 1.19, 0.90$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73 117
- 7.6 Contour plots from a 3PCF-only (yellow) and a joint 2PCF and 3PCF analysis (brown). The minimum scales for the 3PCF are the one listed in Tab. 7.5. For the 2PCF, it is fixed to $r_{\min} = 20h^{-1}\text{Mpc}$. The maximum scale both for 3PCF and 2PCF are the maximum clustering measured lengths, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73 119
- 8.1 Average (black circle) of 2PCF extracted from measuring 2PCF from 298 mock catalogues (thin grey lines). Error bars correspond to the diagonal of the numerical covariance extracted from the 298 mocks. The bin is $\Delta r = 5 h^{-1}\text{Mpc}$ 125
- 8.2 *Top panel*: mean average (black circle) of 3PCF extracted from measuring 3PCF from 298 mock catalogues (thin grey lines). *Middle panel*: signal-to-noise ratio. *Bottom panel*: sides of triangles as a function of the Triangle Index. The bin is $\Delta r = 5 h^{-1}\text{Mpc}$ 125
- 8.3 Contour plots of next-to-leading order 2PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (blue, green, red). The maximum scale for 2PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73. 128

- 8.4 Contour plots of next-to-leading order 3PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (green, red, purple) and the leading order 3PCF at $r_{\min} = 20h^{-1}\text{Mpc}$ (blue) fixing $\eta_{\min} = 3$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black, orange and green lines refer to the bias relation reported in Eqs. 3.72 , 3.73, 4.76. 129
- 8.5 Contour plots of leading and next-to-leading order 3PCF fixing $r_{\min} = 40h^{-1}\text{Mpc}$. The blue contour represents constraints from the leading order 3PCF. Green from the next-to-leading order 3PCF with free parameters. Purple and red from the next-to-leading order 3PCF model using the Lagrangian bias relation specified in the legend. The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black, orange and green lines refer to the bias relation reported in Eqs. 3.72 , 3.73, 4.76. . . 130
- 8.6 Contour plots from the next-to-leading order 2PCF (grey), the next-to-leading order 3PCF (blue) and the joint 2PCF + 3PCF (red) at $r_{\min} = 40h^{-1}\text{Mpc}$ at $\eta_{\min} = 3$. The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73. 131
- 8.7 Contour plots from the joint 2PCF (next-to-leading order) + 3PCF (leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ for $\eta_{\min} = 3, 2, 1$ (brown, green, blue). The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73. 132
- 8.8 Contour plots from the joint 2PCF (next-to-leading order) + 3PCF (next-to-leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ for $\eta_{\min} = 3, 2, 1$ (red, orange, yellow) and the joint 2PCF (next-to-leading order) + 3PCF (leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ at $\eta_{\min} = 3$ (red, orange, yellow). The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73. 133
- 8.9 Inferred bias parameters b_1, b_2 and γ_2 as a function of the minimum scale r_{\min} and parameter η_{\min} . Solid red and dashed blue lines refer respectively to the 3PCF next-to-leading and leading order. 134
- A.1 k_1, k_2 and k_3 - where $k_1 \leq k_2 \leq k_3$ - as a function of the Index. 145

-
- A.1 Computation of individual contributions B_{222} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations. 147
- A.2 Computation of individual contributions B_{321-I} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations. 151
- A.3 Computation of individual contributions B_{321-II} of galaxy bispectrum at the next-to-leading order. For each plot: the *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while the *bottom plot* displays in blue the percentage differences between the two evaluations. 153
- A.4 Computation of individual contributions B_{411} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations. 155

List of Tables

4.1	Divergence conditions for one-loop terms $P_{13,m}$ and $P_{22,m}$	52
6.1	Summary of models used in this paper, methods used to compute them and their degrees of freedom.	88
6.2	Average difference between the absolute values of normalised residuals of SPT-NLO and LO-2D models. Redshifts are listed in columns two to three, while all configurations, BAO, small scales and squeezed BAO scales are listed in rows two to five.	99
6.3	Best fit values for the four parameters of the NLO-EFT model for the squeezed BAO configurations, as depicted in Fig. 6.6.	100
6.4	Signal-to-noise $[S/N]$ at $r_{\min} = 40h^{-1}\text{Mpc}$ and $\eta_{\min} = 0$ and the χ^2 comparison between the model 3PCFs considered in this work, and listed in column 1, and the measurements performed on two snapshots of the DEMNUni simulations at the redshifts listed in column 2.	105
6.5	Same as Tab. 6.4, but $r_{\min} = 20h^{-1}\text{Mpc}$ and $\eta_{\min} = 2$	106
6.6	Reduced χ^2 values obtained when comparing measured and model 3PCF for triangle configurations that encompass the BAO scale (column 3). The models are listed in column 1, and the redshift of the simulation snapshot is shown in column 2. The analysis is described in Sec. 6.4.4.	106
7.1	List of, for each redshift, the number of galaxies N_g and the mean number density \bar{n} in the Flagship galaxy catalogues.	109
7.2	The list of ratios N_R between the random and data datasets	109
7.3	The list of assumed uniform prior ranges.	113
7.4	List of minimum validity scales r_{\min} and η_{\min} or 3PCF models for each snapshot	118
7.5	List of inferred values of bias parameters of the leading order 3PCF model estimated throughout a 3PCF-only and a joint 2PCF and 3PCF analysis. . .	118

7.6 List of ratios between errors on bias parameters of the leading order 3PCF model estimated throughout a 3PCF and a joint 2PCF and 3PCF analysis. . 118

8.1 The list of assumed uniform prior ranges. 126

“Facts are constituted by older ideologies, and a clash between facts and theories may be proof of progress.”

Paul Feyerabend, *Against Method*

1.1 The Λ CDM model

Over the centuries, cosmology has been a profound source of human speculation. Recent experimental observations and theoretical hypotheses have culminated in the formulation of a consolidated model for the universe, known as the Λ CDM model. This model has its roots in the early 20th century, following the observations of Edwin Hubble (18), who discovered that galaxies were moving away from each other, with their velocity increasing over time. Assuming that human observation is not in a preferential point of the universe, this led to the assumption that it is space-time itself that is expanding. Going back in time, this consideration leads to the idea that all matter originated from a single primordial point, referred to as the *Big Bang* singularity. In 1964, Arno Penzias and Robert Wilson (19) discovered evidence of this singularity in the form of cosmic background radiation (CMB), a primordial imprint of photons at decoupling with matter that represents a key milestone for the Big Bang model.

In the wake of Edwin Hubble's groundbreaking observations, Fritz Zwicky discovered that the velocities of galaxies orbiting the center of the Coma cluster were significantly higher than could be explained by the visible mass alone (20). This led to the conclusion that there must be a large quantity of invisible mass, subsequently referred to as dark matter. Subsequent experimental observations over the following years confirmed this hypothesis (21; 22; 23; 24), and theoretical work has attributed the stability of galactic disks to the presence of dark matter (22). Furthermore, observations of anisotropies in the temperature field of the Cosmic Microwave Background (CMB) revealed that cold dark matter was required to account for the fluctuations that led to the formation of large-scale structures. All of this evidence led to the formulation of the idea that a quantity significantly greater than

ordinary matter in the universe is composed of what is known as cold dark matter (CDM) (25; 26).

In the years that followed, an even more revolutionary piece of observational evidence transformed our understanding of the universe. The discovery of accelerated expansion (27; 28) introduced a new form of energy content into the universe: the so-called dark energy. This, which is necessary to explain the dynamics and kinematics of an accelerating expanding universe, is estimated to be the dominant contribution to the universe, accounting for about 70%. The concept of accelerated expansion was incorporated into the standard cosmological model as a cosmological constant. This was introduced by Albert Einstein to have static solutions and was reintroduced to explain accelerated expansion. However, nowadays, the cosmological constant is seen as a way to parameterize our lack of understanding about the nature of accelerated expansion.

The so-called Λ CDM model has thus been confirmed over time by various experimental evidence, such as the CMB (1; 2; 3) and large surveys of the spatial distribution of galaxies in the Universe (4; 5). These observations have provided us with a precise description of the physical content of the universe, albeit very different from the understanding we had only a century ago. Despite the Λ CDM model proving to be a solid model in describing an impressively wide range of observations, it raises some fundamental questions. The nature of dark energy and dark matter still eludes the scientific community, particularly in reconciling the standard cosmological model with the standard model of particle physics.

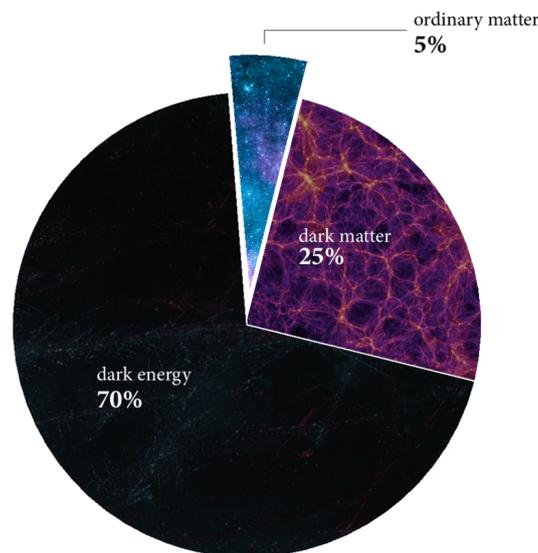


Figure 1.1 The content of the universe in various forms of energy in the LCDM model, as measured by CMB observations and large-scale structure surveys (1; 2; 3; 4; 5). Source: Max-Planck-Institute for Astrophysics Garching and Pixabay.

1.2 The Λ CDM model and the current debate

The cosmological constant, whilst providing an excellent fit for cosmological observations, is currently the subject of much debate regarding its theoretical interpretation. It can be interpreted as a form of vacuum energy, as proposed by (29; 30). However, theoretical predictions stemming from quantum field theory provide a value that is vastly different from that inferred from cosmological observations, with a discrepancy of around 30-120 orders of magnitude. Some theoretical developments have suggested alternative interpretations for the cosmological constant, such as a dynamical source of energy (e.g. (31)) or a modification of Einstein's gravity on large scales (e.g. (32)). Both hypotheses have observable consequences and lead to a modification of the equation of state of dark energy $\omega_{\text{DE}} = p_{\text{DE}}/\rho_{\text{DE}}$ (see Eq. 2.3) - equal to $\omega_{\text{DE}} = -1$ in the Λ CDM model - and of the rate at which fluctuations in the mass density evolve with time.

The experimental observations in cosmology that were carried out over the century and during the first decades of this one have not only exposed our ignorance about the nature of the universe, or at least of its constituents but have also highlighted some tensions in the observations made. The most significant of which has been recently highlighted for the Hubble constant, H_0 , that sets the velocity of the expansion of the universe at the present epoch: there are significant discrepancies between the H_0 value inferred from the CMB and the one obtained from more local measurements of extragalactic objects in the local Universe. As an example, the Planck collaboration found $H_0 = 67.2 \pm 0.60$; km/s/Mpc (3), which is in tension with the value of $H_0 = 74.03 \pm 0.60$; km/s/Mpc (33) obtained through the analysis of Hubble Space Telescope observations using 70 long-period Cepheids in the Large Magellanic Cloud. A joint analysis of CMB and BAO (34) data has supported the value inferred from CMB data alone, lending credibility to those values. These two estimates - the first being direct and the second being indirect - were accompanied by a third estimate recently obtained from the so-called standard sirens (6), which are distance indicators derived from the measurement of gravitational waves. This third estimate, although compatible with the previous two estimates given their wide uncertainty, has further brought attention to this tension.

Another tension that has emerged in recent observational analyses of the Cosmic Microwave Background (CMB) and weak gravitational lensing, as well as galaxy clustering through Large Scale Structure (LSS) surveys, concerns the parameters σ_8 and Ω_m . The discrepancy shown between CMB data and observations from the Kilo Degree Survey (KiDS), Baryon Oscillation Spectroscopic Survey (BOSS), Dark Energy Survey (DES), and 2dF

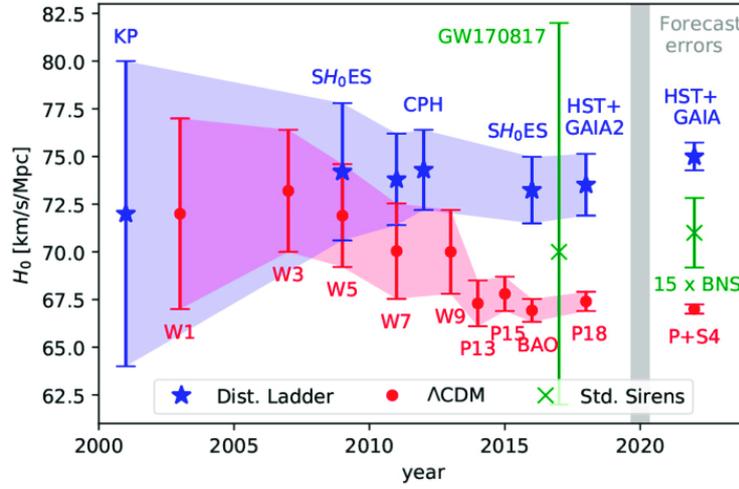


Figure 1.2 The Hubble tension, including the first standard sirens measurement following GW170817 (6), Planck (3) and Hubble Space Telescope (HST) with GAIA DR2 (7)). Blue stars correspond to measurements of H_0 in the local universe with calibration based on Cepheids. Red dots refer to derived values of H_0 from the CMB, assuming Λ CDM. Green crosses are direct measurements of H_0 with standard sirens. Forecasts are CMB Stage IV (8), standard sirens (9) and distance ladder with full GAIA and HST (10). Figure from (11)

Galaxy Redshift Survey is a subject of investigation in the current research debate, despite the tension in question being less pronounced than that of the Hubble constant.

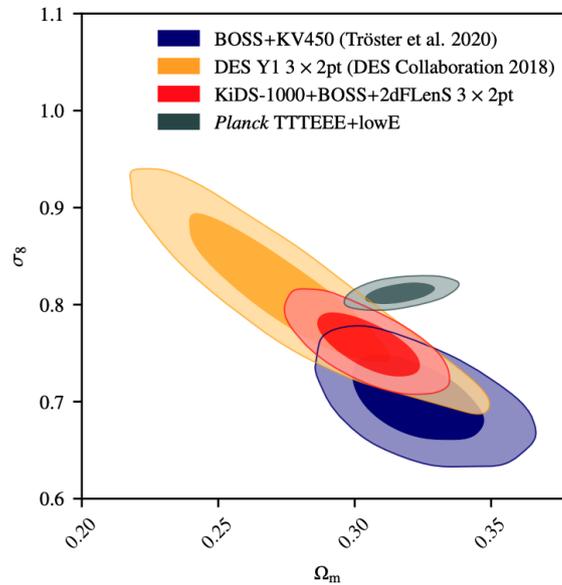


Figure 1.3 Marginalised posterior distribution in the $\sigma_8 - \Omega_m$ plane, comparing the 3×2 pt analyses from KiDS-1000 with BOSS and 2dFLENs (12), with the 3×2 pt analysis from DES Y1 (6), and the CMB constraints from Planck Collaboration (3). Figure from (12)

In Fig. 1.3, the aforementioned tension is shown; the constraints obtained from combined weak lensing analyses of BOSS and KiDS (35) are shown in blue, yellow and red constraints from so-called combined weak gravitational lensing, cosmic shear, galaxy clustering two-point correlation functions (3x2) from DES (36) and KiDS, BOSS and 2dF (12), and in grey, the CMB results from Planck (3). To resolve the tension in the coming years, it will be necessary to reduce uncertainties on cosmological parameters.

The very existence of these tensions along with the uncertainties on the theoretical foundations of the Λ CDM model illustrates the importance of improving observational constraints which, in turns, is the main scientific drive behind the ongoing and soon-to-be started observational campaigns aimed at probing increasingly larger fraction of the Universe. Within this effort, two probes have proved more effective than the others, gravitational lensing and the spatial clustering of extragalactic objects. This Thesis will focus on the latter.

1.3 Motivations for this Thesis

Large-scale galaxy catalogues provide a wealth of cosmological information. The first analyses of the three-dimensional distribution of galaxies date back to the late seventies, although their first use as a cosmological tool came with the CfA Redshift Survey (37). That survey collected the redshifts of around 2,400 galaxies between 1977 and 1982. In the 1990s, the Two-degree-Field Galaxy Redshift Survey (2dF) contained around 220,000 galaxies, leading to the first precise power spectrum measurement (38). A further improvement was achieved by the Sloan Digital Sky Survey (SDSS) (39) with more than one million galaxies, which enabled the determination of Baryonic Acoustic Oscillations (BAO). 1.5 million galaxies around redshift $z = 0.7$ were collected by the BOSS survey, which provided stringent constraints on the Λ CDM model. The next generation of redshift surveys such as Euclid (40; 41), Dark Energy Spectroscopic Instrument (DESI) (42), Large Synoptic Survey Telescope (LSST) (43), Nancy Grace Roman Space Telescope high latitude survey (WFIRST) (44) and Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) (45) aim at significantly improving the state of the art precision in the estimate of cosmological parameters and will allow us to shed light on the nature of the dark components, possibly solving the "tensions" on the cosmological parameters mentioned above.

So far, the two-point statistics have represented the preferred tool to analyse the clustering properties of matter in the Universe and, from that, to constrain its physical properties. Indeed, 2-point statistics is all one needs if the cosmological fields obey Gaussian statistics. However, unlike the CMB case, two-point statistics of large-scale structures are not a sufficient or

complete description for extracting all available information on the gravitational evolution of structures and inevitably lead us to consider higher-order statistics, such as the three-point function. These statistics are particularly important for breaking the degeneracy among cosmological parameters that inevitably affect analyses that consider 2-point statistics only. They also provide a powerful framework for testing inflationary model predictions and constraining the relation between the spatial distribution of the matter, which is mostly dark, and its luminous tracers: the galaxies. For these reasons, three-point statistics have long been considered an important tool in studying the Large Scale Structure of the universe (LSS) (46; 47; 48; 49; 50; 51; 52; 53). To be effective, they require very large datasets, so it is not surprising that the interest in these tools has increased in recent years with the availability of galaxy catalogues with unprecedented size and will further increase in the near future when the next-generation datasets will be available from upcoming spectroscopic galaxy surveys.

At its core, three-point statistics is about counting triplets of different sizes and shapes. To efficiently extract cosmological information, it is mandatory to obtain reliable theoretical predictions on the largest possible number of triangle configurations, which implies including the numerous triangles of small size, i.e. to probe the nonlinear regime in the evolution of the mass density fluctuations (54).

Standard Perturbation Theory (SPT, see (15) for a review) of clustering statistic has proved, so far, an effective way of accessing nonlinear scales in Fourier space. Hence the widespread use of perturbative expansion techniques to investigate the clustering properties of the matter in the universe through the power spectrum and bispectrum statistics (55; 56; 57; 58; 59; 60). Concerning bispectrum, efforts have been made to develop alternative routes by re-summing perturbative contributions in the Eulerian (61; 62; 63), Lagrangian (64) and Effective Field Theory (EFT) approaches (65; 66; 67; 68). The state of 3-point correlation modelling in configuration space is comparatively less advanced. So far, 3PCF models have been developed at the tree level only (49; 53; 69). They have been successfully used to analyse clustering on quasi-linear scales (70; 71; 72; 73; 74), hence limited to a relatively small fraction of available triplets and missing information from the mildly nonlinear scales. As yet, direct modelling in configuration space has not yet been explored due to the complexity of fluid equations in configuration space. This is because modelling 3-point statistics in configuration space is complicated due to the relation with the Fourier space counterpart in which models are provided. The inverse-Fourier transform induces a scale mixing that requires adopting a computationally demanding numerical approach, even in the mildly nonlinear regime. On the other hand, the 3PCF approach offers a significant advantage when dealing with real datasets consisting of galaxy surveys with complicated geometry. In Fourier space, the survey footprint induces mode coupling in Fourier that requires computationally

demanding numerical approaches (75; 76). For 3PCF, the impact of the survey footprint can be efficiently corrected at the estimator level. A second element that has hampered the development of the 3PCF tool has been, until not long ago, the computational cost of the standard estimators counting all triplets in the sample. The situation has changed dramatically since new types of 3PCF estimators capable of reducing computational cost from N^3 to N^2 , have been proposed (17; 77; 78). The aim of this Thesis is to bridge the gap with modelling in Fourier space, achieving the same perturbative order reached in Fourier space. This Thesis is divided as follows. In the first chapters, from Chapter 2 to Chapter 4, the introductory framework is presented, focusing on the theoretical developments in perturbation theory. In Chapter 5, the proposed strategy in this Thesis is presented. In Chapters 6, 7 and 8, the applications of the aforementioned strategy and the results of this Thesis are described.

2

Background cosmology

In this chapter, I present a comprehensive overview of the key characteristics of a Friedmann-Lemaître-Robertson-Walker (FLRW) universe, which is commonly accepted as the metric describing the homogeneous Universe. Specifically, in Sec. 2.1 and Sec. 2.2, I provide a summary of the fundamental principles behind the FLRW equations. Additionally, in Sec. 2.3, I demonstrate the limitations of the Hot Big Bang Theory and how the inflationary paradigm resolves these issues. For further details, see references (79; 80).

2.1 Friedmann-Lemaître-Robertson-Walker universe

2.1.1 The Friedmann equations

The observations in modern cosmology can be largely condensed by invoking the Cosmological Principle, which postulates the homogeneity and isotropy of the Universe at large scales. This principle is encapsulated in the following line element ds^2 ruling the geometry of the so-called background cosmology:

$$ds^2 = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega \right], \quad (2.1)$$

where $a(t)$ is the so-called scale factor, $d\Omega = d\theta^2 + \sin^2(\theta)d\phi^2$ and $c, \theta, d\Omega$ being, respectively, the speed of light and the two angles in a spherical coordinate system.

The parameter k in this equation determines the topology of the FLRW Universe, with $k = 0, +1, -1$ denoting a flat, positively curved, and negatively curved universe, respectively. It is worth mentioning that the flat case corresponds to a conformal Minkowski metric. The scale factor $a(t)$ plays a crucial role in describing the evolution of the Universe, and it is worth noting that the values of $g_{00} = 1$ and $g_{0i} = 0$ allow for a global definition of a time

coordinate. Einstein equations, presented below, play a key role in determining the scale factor dynamics

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}^{p.f}, \quad (2.2)$$

where $T_{\mu\nu}^{p.f}$ represents the stress-energy tensor of a cosmological perfect fluid, considered as an appropriate physical description of universe's expansion and structures' formation

$$T_{\mu\nu}^{p.f} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu}. \quad (2.3)$$

where ρ and p represent, respectively, the energy density and pressure of the fluid. Choosing a cosmic rest frame, i.e. the frame where the average velocity of energy forms in the Universe is zero, $u_\mu = (1, 0, 0, 0)$ and $T_{\mu\nu}^{p.f}$ takes the form

$$T_{\mu\nu}^{p.f} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \quad (2.4)$$

Considering the 00 component and defining the Hubble parameter $H = \frac{\dot{a}}{a}$, the *first Friedmann equation* can be expressed as follows:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G c^2}{3} \rho - k \frac{c^2}{a^2}. \quad (2.5)$$

The latter can be expressed in the following form

$$H^2 = H_0^2 \left(\frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda \right), \quad (2.6)$$

where H_0 indicates the measured Hubble parameter at the current time and the indexes r, m, k, Λ refer to radiation, matter, curvature and cosmological constant density parameters, defined as

$$\Omega_n = \frac{\rho_n}{\rho_c}, \quad (2.7)$$

$$\rho_c = \frac{3H^2}{8\pi G} = 1.8788 \times 10^{-26} h^2 \text{kg m}^{-3}, \quad (2.8)$$

where ρ_c is the so-called *critical density* and ρ_n being determined by the following expansion

$$\rho = \sum_{n=-\infty}^{+\infty} \rho_n a^{-n}. \quad (2.9)$$

Usually, as it will be explained in Sec. 2.2, mainly the case $n = 4, 3, 0$ are considered. Focusing on the spatial component of Einstein's equation in an FLRW metric, it is possible to derive the following expression

$$\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2k\frac{c^2}{a^2} = -c^2\frac{4\pi G}{c^3}(\rho - p). \quad (2.10)$$

Plugging Eq. 2.5 into Eq. 2.10, the latter - commonly dubbed as *second Friedmann equation* - can be expressed as

$$\frac{\ddot{a}}{a} = -c^2\frac{4\pi G}{3}(\rho + 3p). \quad (2.11)$$

The curvature term affects the scale factor dynamics, in fact:

- if $k = 0$, then $(\frac{\dot{a}}{a})^2 = \frac{8\pi G c^2}{3}\rho \geq 0$ and $\dot{a} = 0$ asymptotically,
- if $k = -1$, then $(\frac{\dot{a}}{a})^2 = \frac{8\pi G c^2}{3}\rho + \frac{c^2}{a^2} \geq 0$ and $\dot{a} \neq 0$ during the evolution,
- if $k = +1$, then $(\frac{\dot{a}}{a})^2 = \frac{8\pi G c^2}{3}\rho - \frac{c^2}{a^2} \geq 0$ and this represents the turning point case.

Now from the first law of thermodynamics, it is possible to derive another essential equation

$$dU = \delta Q - PdV, \quad (2.12)$$

where $\delta Q = TdS$. By definition, the universe is a closed system, so $\delta Q = 0$. Using $U = \rho V$ and FLRW metric presented in Eq. 2.1 it is possible to derive the *continuity equation*

$$\dot{\rho} = -3H(\rho + p). \quad (2.13)$$

2.1.2 Cosmological redshift

Let me redirect the attention to other pertinent aspects of the kinematics of the FLRW Universe: the gravitational redshift represents a noteworthy phenomenon in FLRW cosmology. To illustrate this point, let us consider the geodesic equation for a massless particle, specifically, the photon in the FLRW metric:

$$\frac{d^2 x^\alpha}{d\tau^2} + \Gamma_{\mu\nu}^\alpha \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \quad (2.14)$$

It's easy to see that the 00 component of a geodesic equation can be rewritten in the form

$$\frac{dE}{d\tau} + \frac{\dot{a}}{a}E = 0. \quad (2.15)$$

This equation gives us the energy-scale factor relation

$$E \sim \frac{1}{a}. \quad (2.16)$$

So, reminding $E = \frac{hc}{\lambda}$, the redshift relation due to geometry at large scale reads as

$$\frac{\lambda_0}{\lambda(t)} = \frac{a_0}{a(t)} = 1 + z, \quad (2.17)$$

where $z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}}$ is the redshift. By measuring photon redshift, it is possible to know the scale factor evolution from t_0 to t .

2.1.3 Hubble's law

Another cornerstone in modern cosmology is the role of the *Hubble law* and its relation with FLRW cosmology. In 1929 Edwin Hubble observed a recessional motion of astronomical objects explained by Hubble's law:

$$v = H_0 d, \quad (2.18)$$

where v is recession velocity, H_0 is the Hubble constant (for the current debate around its estimate, see a brief overview in Sec. 1) and d is the luminosity distance, defined by

$$F = \frac{L_m}{4\pi d^2}, \quad (2.19)$$

where F is the measured flux, and L_m is the measured luminosity. It is now possible to define the *comoving distance* as the distance between two points measured along a path defined at the present cosmological time:

$$\chi = \int_{t_i}^t c \frac{dt'}{a(t')}, \quad (2.20)$$

where t_i is the time of emission of a photon detected by an observer at the current time t . Considering a source located at the proper comoving distance because of the expansion of the universe, it is possible to define the proper distance in physical units:

$$D(z) = (1 + z)r(z). \quad (2.21)$$

Using Eqs. 2.6, 2.20, the relation between comoving distance and redshift reads as:

$$\begin{aligned} r(z) &= c \int_a^1 \frac{da}{a^2 H_0 \left[\frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda \right]} \\ &= \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m (1+z')^3 + \Omega (1+z')^4 + \Omega_\Lambda}}. \end{aligned} \quad (2.22)$$

where it has been used the convention $a_0 \equiv a(t_0) = 1$. From this expression and using 2.21, the correct formulation of the Hubble law can be expressed as

$$D(z) = \frac{c}{H_0} (1+z) \int_0^{z^*} \frac{dz}{\sqrt{\Omega_m (1+z)^3 + \Omega (1+z)^4 + \Omega_\Lambda}}. \quad (2.23)$$

It is worth stressing that the latter Eq. is a generalisation of Eq. 2.18, which can be recovered in the limit $z \ll 1$. Considering the second-order approximation of Eq. 2.23 it is possible to write the following relation

$$D(z) = \frac{c}{H_0} z \left[1 + \frac{1}{2} (1 - q_0) z + o(z^2) \right], \quad (2.24)$$

$$q_0 \equiv \left[- \left(\frac{1}{H_0} \right)^2 \frac{\ddot{a}}{a} \right]_{t=t_0}. \quad (2.25)$$

The parameter q_0 takes the name deceleration parameter, and $q_0 \geq 0$ identifies a decelerating universe; otherwise, $q_0 \leq 0$ identifies an accelerating universe.

2.2 Energy forms in a Friedmann-Lemaître-Robertson-Walker Universe

Assuming a linear relationship between energy density and pressure, the equation of state is usually expressed in the form

$$p = \omega \rho, \quad (2.26)$$

where w is the parameter defining the form of energy. By recourse to the aforementioned equation, it becomes feasible to re-cast the continuity Eq. (2.13) and the second Friedmann Eq. (2.11) into alternative forms

$$\frac{\dot{\rho}}{\rho} = -3(1+\omega) \frac{\dot{a}}{a}, \quad (2.27)$$

$$\frac{\ddot{a}}{a} = -c^2 \frac{4\pi G}{3} (1+3\omega) \rho. \quad (2.28)$$

It is instructive to separate energy density into hypothesised energy forms expanding ρ as follows

$$\rho = \sum_{n=-\infty}^{n=+\infty} \rho_n a^{-n}, \quad (2.29)$$

where ρ_n s remain fixed to their values at the scale factor $a_0 = 1$ and $n \in \mathbb{N}$. For a perfect fluid, the Eq. 2.26 is related to the expansion in Eq. 2.29 as

$$\omega = \frac{n}{3} + 1. \quad (2.30)$$

Using the definition given in Eq. 2.7, it is possible to obtain an alternative expression for Friedmann Eq. 2.5

$$\left(\frac{H}{H_0}\right)^2 = \sum_{n=-\infty}^{n=+\infty} \Omega_n a^{-n} + (1 - \Omega_{tot}) a^{-2}. \quad (2.31)$$

where

$$\Omega_{tot} \equiv \sum_{n=-\infty}^{n=+\infty} \Omega_n. \quad (2.32)$$

Using 2.29, let me write Eq. 2.27 in terms of the present epoch where $a = 1$, $H = H_0$, $n = 3\omega + 3$ and Ω_n are fixed at their $a = 1$ values

$$\frac{\ddot{a}}{a} = H_0^2 \sum_{n=-\infty}^{n=+\infty} \left(1 - \frac{n}{2}\right) \Omega_n a^{-n}. \quad (2.33)$$

Furthermore, considering Eq. 2.5 and computing it at the current time, it becomes

$$H_0^2 = H_0^2 \left[\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda \right], \quad (2.34)$$

from which the following constraint derives, playing an important relation between cosmological parameters

$$\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1. \quad (2.35)$$

Before focusing on the energy content in an FLRW universe, it is worth stressing two main kinds of horizons commonly encountered:

- *Hubble horizon*: also known as the *Hubble radius*, is the distance beyond which objects in the universe are moving away from each other faster than the speed of light due to the expansion of the universe. It is defined as the distance that light can travel since the beginning of the universe, as determined by the current expansion rate of the universe

and the amount of matter and energy it contains, usually defined by

$$r_H(t) = \frac{c}{H(t)}, \quad (2.36)$$

- *Particle horizon*: the maximum distance from which information or particles can reach an observer at a given time due to the finite speed of light. It is the distance that light could have travelled since the beginning of the universe, considering the expansion of the universe and the amount of matter and energy it contains, and it expressed as

$$r_P(t) = c \eta(a) = c \int_0^a \frac{1}{a' H(a')} \frac{da'}{a'}, \quad (2.37)$$

where $\eta(a)$ is the so-called conformal time.

The particle horizon is different from the Hubble horizon in that it is defined by the information that can be transmitted to an observer rather than the distance beyond which objects are moving away from each other faster than the speed of light. The particle horizon is, therefore, a measure of the amount of the universe that is causally connected to an observer at a given time. At the same time, the Hubble horizon is a measure of the distance beyond which objects are receding from each other too quickly to be observed.

2.2.1 Radiation

Radiation is a form of energy that behaves as a fluid, and its equation of state is given by $\omega = \frac{1}{3}$ ($n = 4$) in the context of the Friedmann equations. Were it possible to completely confine all radiation within static, non-interacting boxes, the radiation would not exhibit the equation of state $\omega = \frac{1}{3}$. As the universe expands, each box would retain the same amount of radiation energy as measured within, thereby precluding any loss of energy. The energy density between the boxes would remain null, and the number density of the boxes would decrease in accordance with the spatial dimensionality of the universe, $n = 3$. Consequently, the boxes would evolve as $\omega = 0$. This means that radiation is a form of energy that cannot be confined to static, non-interacting boxes.

After the inflation era, radiation plays the main contributor to Friedmann Eq. 2.5. When radiation dominates the Friedmann equations, then

$$H^2 = H_0^2 \Omega_m a^{-4}, \quad (2.38)$$

and the solution is

$$a \sim t^{\frac{1}{2}}. \quad (2.39)$$

2.2.2 Matter

Matter represents a form of energy that evolves as $\omega = 0$ ($n = 3$). Particles that display $\omega = 0$ include baryonic matter, dark matter, and black holes that operate as $\omega = 0$ entities in a cosmological context. Should a particle move in a spatial direction, its ω value would rise to $0 < \omega < \frac{1}{3}$. In the event that it moves relativistically, its ω value would approach $\omega \approx \frac{1}{3}$, rendering it a form of radiation.

When particles dominates the Friedmann equation 2.6, then

$$H^2 = \Omega_m H_0^2 a^{-3}, \quad (2.40)$$

and the solution is

$$a \sim t^{\frac{2}{3}}. \quad (2.41)$$

2.2.3 Cosmological constant

The form of energy that evolves according to the equation of state $\omega = -1$ ($n = 0$) in the Friedmann Eq. 2.6 is known as the cosmological constant or Λ . This energy is uniformly distributed in space, with no confinement in any spatial direction. As the universe expands, the cosmological constant must remain isotropic to all rest frames in the universe. This form of energy has been proposed as a crucial component of the universe since the time of Einstein, who originally sought to create a repulsive gravitational component to prevent a universe composed solely of $\omega = 0$ matter particles from collapsing. Over the years, various suggestions have been made to account for possible anomalies in cosmological data through the introduction of a cosmological constant. While these proposals have often been met with skepticism in early studies, supernova data indicate that standard candle supernovae appear dimmer than expected in a universe dominated only by $\omega = 0$ matter (28), (27). These observations are well-explained by the presence of a high-density cosmological constant with $\omega = -1$. Furthermore, current analyses of cosmic microwave background data indicate that a flat universe with $\Omega_{total} = 1$ is comprised of two $\omega = 0$ components that account for approximately 30% of the critical density, with the remaining 70% attributed to the $\omega = -1$ component. This model is consistent with galaxy clustering data and suggests that the $\omega = -1$ cosmological constant constitutes approximately 70% of the universe's energy. As the equation 2.27 shows, a dominating component with $\omega = -1$ exerts gravitationally repulsive pressure, leading to an acceleration of the expansion of the universe. For the flat

case, Eq. 2.27 can be integrated yielding

$$\ln(a) = H_0 t + \ln(a_0), \quad (2.42)$$

where t can be considered the time since $\omega = -1$ energy began to dominate the expansion, when $a = a_0$. The solution is

$$a = a_0 e^{H_0 t}. \quad (2.43)$$

The universe is said to be in a de-Sitter phase, and the Hubble parameter H is static at H_0 . Such a phase is hypothesized to have dominated the early universe in a phase called inflation.

2.3 The inflationary paradigm

The solution of the first Friedmann equation 2.5, using any discussed form of energy except a deSitter universe - i.e. a universe whose physical dynamics is dominated only by cosmological constant -, is a monotone function of cosmological time. This can be shown by extrapolating back in time Eq. 2.41 and Eq. 2.39. In this sense, that extrapolation of the expansion of the universe backward in time yields an infinite density and temperature at a finite time in the past. This singularity indicates that general relativity is not an adequate description of the laws of physics in this regime. The fact that the Universe expands implies that it was denser and warmer in the past. More in general, the Hot Big Bang theory has its problems. A part of this is because this theory needs particular initial conditions; otherwise, it would grossly fail to describe the early and present Universe. let me distinguish the main three issue for which the Hot Big Bang theory find a hard explanation:

- **Flatness problem:** considering the Einstein equations valid until the Planck era, Eq 2.5 can be recast as

$$\Omega - 1 = \frac{8\pi G\rho}{3H^2} = \frac{kc^2}{H^2 a^2}. \quad (2.44)$$

Both in radiation- and matter- dominated eras, it is possible to recognize that since $(\Omega_0 - 1)$ is measured to be of order unity at the present time, the curvature is constrained to be unitary at $\mathcal{O}(10^{-64})$ at Planck time and $\mathcal{O}(10^{-16})$ if we limit to the nucleosynthesis time - i.e the process by which the universe's light elements (primarily hydrogen, helium, and lithium) were formed in the first three minutes after the Big Bang (for details see (81)). The flatness problem is also known as a fine-tuning problem.

- **Entropy problem:** the entropy density is of the order of photon number density; in the present Universe, we have

$$s \sim 10^3 \text{ cm}^{-3}. \quad (2.45)$$

Thus, the estimate for the entropy in the observable part of the Universe, whose size is $R_0 \sim 10^4 \text{ Mpc} \sim 10^{28} \text{ cm}$, is

$$S \sim sR_0^3 \sim 10^{88}. \quad (2.46)$$

This huge dimensionless number is one of the properties of our Universe. The Hot Big Bang theory does not explain why the Universe has such a large entropy. This problem is known as the entropy problem.

- **Horizon problem:** from the epoch of last-scattering, photons free-stream, and basically, they travel undisturbed, providing us with a snapshot of the Universe at $z \sim 1000$, meaning their detection is almost a picture of the primordial universe. The angular size of the Hubble radius at the last scattering is expected to be a few degrees, which implies the existence of several casually disconnected patches on the last scattering surface that we observe now. This means there were $\sim 10^6$ disconnected regions with the volume that now corresponds to our horizon. The horizon problem is also related to the problem of temperature anisotropies, that can be in general expanded in spherical harmonics as follows:

$$\frac{\Delta T}{T}(x_0, \tau_0, \mathbf{n}) = \sum_{\ell m} a_{\ell m}(x_0) Y_{\ell m}(\mathbf{n}) \quad (2.47)$$

where x_0 and τ_0 are our position and the present time, while \mathbf{n} is the direction of the observation, ℓ s are the different multipoles so that

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell' \ell} \delta_{m m'} C_\ell \quad (2.48)$$

where C_ℓ is the so-called CMB power spectrum. Due to homogeneity and isotropy, C_ℓ is neither a function of x_0 nor m . The two-point correlation function (for details, see chapter 5, can be expressed as

$$\left\langle \frac{\delta T(\mathbf{n})}{T} \frac{\delta T(\mathbf{n}')}{T} \right\rangle = \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell P_\ell(\mu = \mathbf{n} \cdot \mathbf{n}') \quad (2.49)$$

where P_ℓ is the Legendre polynomial of order ℓ . We report the CMB temperature fluctuations, as presented in Eq. 2.49, as a function the angular scale or multipole

moments in Fig. 2.1 Let me now consider the last-scattering surface, that is distant

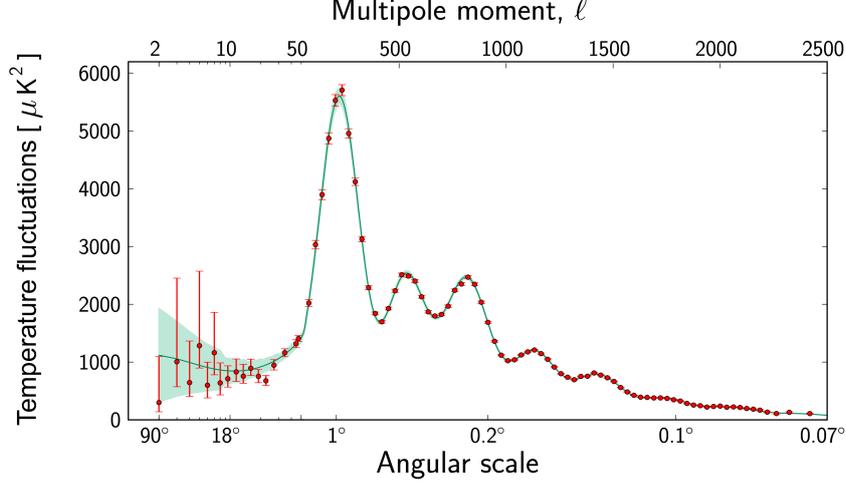


Figure 2.1 The CMB anisotropy as a function of ℓ and the angular scale from Planck's data (3).

from us by the quantity

$$\int_{t_{\text{ls}}}^{t_0} \frac{dt}{a} = \int_{\tau_{\text{ls}}}^{\tau} d\tau = \tau_0 - \tau_{\text{ls}}. \quad (2.50)$$

A given comoving scale, order of the comoving sound horizon at the last-scattering, $\lambda \sim c_s \tau_{\text{ls}}$ - where $c_s \simeq \frac{1}{\sqrt{3}}$ - is therefore projected on the last-scattering surface sky on an angular scale

$$\theta \simeq c_s \frac{\tau_{\text{ls}}}{\tau_0 - \tau_{\text{ls}}} \simeq c_s \frac{\tau_{\text{ls}}}{\tau_0}, \quad (2.51)$$

assuming $\tau \gg \tau_{\text{ls}}$. Since from the last scattering, the Universe's evolution is ruled by the matter component, the angle θ_{HOR} subtended by the sound horizon is

$$\theta_{\text{HOR}} \simeq c_s \left(\frac{T_0}{T_{\text{ls}}} \right)^{\frac{1}{2}} \sim 1^\circ, \quad (2.52)$$

having used $T_{\text{ls}} = 0.3 \text{ eV}$ and $T_0 \sim 10^{-13} \text{ GeV}$, corresponding to the multipole

$$\ell_{\text{HOR}} = \frac{\pi}{\theta_{\text{HOR}}} \simeq 200. \quad (2.53)$$

So, two photons on the last scattering surface separated by an angle larger θ_{HOR} , corresponding to multipoles smaller than ℓ_{HOR} are not in causal connection. But, from Fig. 2.1, on that range, small anisotropies of the same order of magnitude are present making the considered photon appear to be in causal contact.

All of these issues have elegant solutions within the framework of the inflationary paradigm. According to this theory, the early hot cosmological epoch was preceded by an exponential expansion known as inflation. During the inflationary epoch, a small region of the Universe, initially comparable in size to the Planck length, underwent a rapid and tremendous expansion, increasing its size by many orders of magnitude beyond what is observable today.

Allow me to provide a brief explanation of the underlying physics of inflation. Consider a Lagrangian density involving a scalar field:

$$\mathcal{L}_\varphi = -\frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi). \quad (2.54)$$

The energy density and the pressure of this scalar field $\varphi(t)$ can be expressed as a function of their kinetic energy and the potential in the form

$$\rho_\varphi \simeq \frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + V(\varphi), \quad (2.55)$$

$$p_\varphi \simeq \frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - V(\varphi). \quad (2.56)$$

If the potential energy dominates over its kinetic energy, I obtain $p_\varphi \simeq -\rho_\varphi$ providing an exponential phase. The inflaton field, to provide an almost constant energy density, must obey a slow evolution over time, i.e.

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0, \quad (2.57)$$

implies the condition

$$|\ddot{\varphi}| \ll 3|H\dot{\varphi}|. \quad (2.58)$$

This condition can be expressed in terms of the so-called slow-roll parameters that are so required to be small

$$\epsilon \equiv \frac{1}{16\pi G} \left[\frac{V'(\varphi)}{V(\varphi)} \right]^2 \ll 1, \eta \equiv \frac{1}{8\pi G} \frac{V''(\varphi)}{V(\varphi)} \ll 1. \quad (2.59)$$

Inflation ends when the slow-roll conditions are violated, and the inflation field decays into particles in the so-called reheating phase. Throughout the inflationary paradigm, it is possible to model an epoch of exponential expansion for which $\Omega - 1 = \frac{k}{a^2 H^2}$ as the Hubble rate is constant during inflation and the ratio between curvature values at different times t_I, t_f becomes

$$\frac{|\Omega - 1|_{t=t_f}}{|\Omega - 1|_{t=t_I}} = e^{-2N}, \quad (2.60)$$

solving the flatness problem. The expansion epoch due to the inflaton field, in addition to the fact that the entropy of $S \simeq (aT)^3$, also solves the entropy problem. Let me just stress that a large amount of entropy is not produced during inflation but during the non-adiabatic phase transition, which usually takes place during the radiation era. Finally, during the inflationary epoch, the Hubble radius H^{-1} is constant, and during this phase, all the physical scales that have left the Hubble radius can re-enter the Hubble radius in the past. This fact solves the horizon problem. The inflation paradigm also provides us with a mechanism for generating primordial perturbations, as it will be explained in Chapter 3

3

Perturbation Theory

In this Chapter, I introduce the theory that describes the generation and subsequent evolution of scalar density perturbations in a cosmological framework, a process that eventually leads to the formation of the large-scale structure of the Universe (LSS) that we observe in the low redshift Universe. The main elements of the theory that I will use in this Thesis are summarised in the flow chart shown in Fig. 3.1. All these elements are described in the sections below

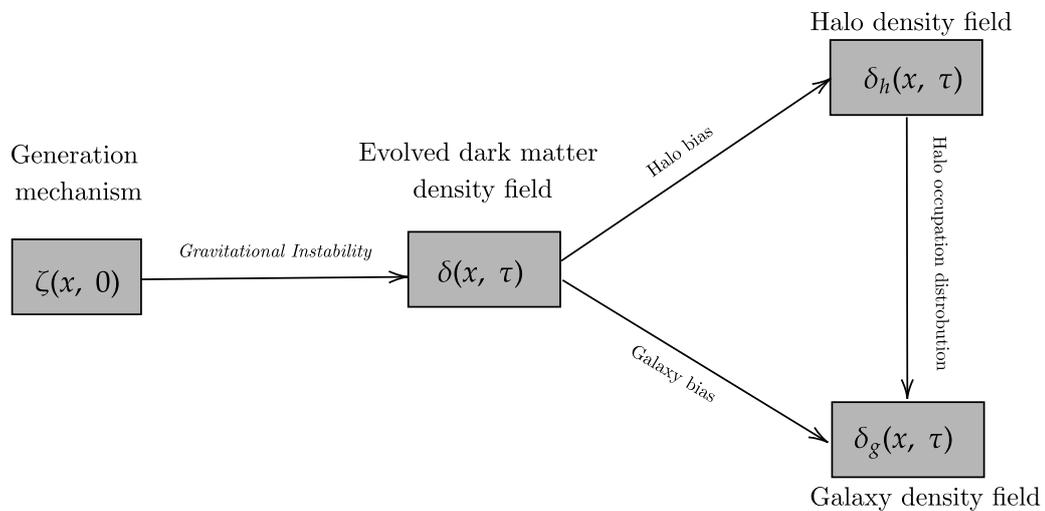


Figure 3.1 Description of how non-linearities in the gravitation evolution, in the biasing, overlap to the initial condition set by inflation.

I briefly review the status of their modelling in sec. 3.3. The final step in the investigation of these processes is the anisotropic contribution from redshift space distortion (RSD), which I do not consider in this Thesis, as I limit the status of modelling in the so-called real space. i.e. not considering distortion effect due to peculiar velocities and (RSD) and incorrect cosmological parameters used to inferred separation between observed objects, also known

as the Alcock-Paczynski effect (AP). The next sections follow other treatments of the physics of LSS such as (14; 82; 83; 84).

3.1 Primordial perturbations

Our understanding of structures is that they originated from small perturbations which grew along the Universe's evolution. Our best guess for the origin of these perturbations is quantum fluctuations during the inflationary era. Then, let me consider how small departures from the background inflaton field φ can give the source of primordial density fluctuations

$$\varphi(\mathbf{x}, t) = \bar{\varphi}(t) + \delta\varphi(\mathbf{x}, t). \quad (3.1)$$

Applying Eq. 2.57 to Eq. 3.1 in Fourier space, I get

$$\delta\ddot{\varphi}(\mathbf{k}) + 3H\delta\dot{\varphi}(\mathbf{k}) + \frac{k^2}{a^2}\delta\varphi(\mathbf{k}) = 0. \quad (3.2)$$

This equation of motion has the same structure as a harmonic oscillator with a friction term due to Hubble expansion. This structure can be used as an analogy to give a quantum picture of the perturbation of the inflaton field

$$\delta\hat{\varphi}(\mathbf{k}, t) = v(\mathbf{k}, t)\hat{a}(\mathbf{k}) + v^*(-\mathbf{k}, t)\hat{a}^\dagger(-\mathbf{k}), \quad (3.3)$$

where $v(\mathbf{k}, t)$ is a complex function and \hat{a}, \hat{a}^\dagger are the creation and annihilation operators satisfying the usual quantum commutation rules

$$[\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k}')] = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}'). \quad (3.4)$$

This leads to finding a solution of Eq. 3.3, and it is possible to derive an expression for the vacuum fluctuation on a super-horizon scale

$$\langle \delta\hat{\varphi}(\mathbf{k}, t)\delta\hat{\varphi}(\mathbf{k}', t) \rangle = (2\pi)^3 \frac{H^2}{2k^3} \delta_D(\mathbf{k} + \mathbf{k}') \equiv (2\pi)^3 P_{\delta\varphi}(k) \delta_D(\mathbf{k} + \mathbf{k}') \quad (3.5)$$

where I define here the power spectrum of inflaton perturbation

$$P_{\delta\varphi}(k) \equiv \frac{H^2}{2k^3}. \quad (3.6)$$

let me also define the dimensionless power spectrum

$$\Delta_{\delta\varphi}^2 \equiv k^3 \frac{P_{\delta\varphi}(k)}{2\pi^2} \propto H^2 \quad (3.7)$$

that is scale-independent. Departures from scale invariance are usually parametrised by the so-called spectral index n_s , so that

$$n_s - 1 \equiv \frac{d\ln\Delta_{\delta\varphi}^2}{d\ln k} \simeq 2\eta - 6\epsilon \quad (3.8)$$

where we have stressed the relation between deviation from scale invariance and the values of slow-roll parameters. This is a testable prediction, confirmed recently by CMB experiments (85). Finally, the linearity of Eq. 3.3 gives Gaussian perturbations. Once perturbation of the inflaton field has been described, it is possible to wonder how large-scale structures are related. To explain, let me refer to Fig. 3.2. The Hubble radius evolves as

$$r_H = cH^{-1} = \left[\frac{8\pi c^2 G \rho(a)}{3} \right]^{(1/2)} = \begin{cases} a^2 & \text{radiation - dom,} \\ a^{3/2} & \text{matter - dom} \end{cases} \quad (3.9)$$

In radiation and matter eras, the Hubble radius, depicted by the solid green line, increases faster than the scale factor, represented by the red line. Since gravity interacts with any component of the universe, small fluctuations of the inflaton field are related to fluctuations of the space-time metric, giving rise to fluctuations in the gravitational potential. So, the wavelength λ of a perturbation leaves the Hubble radius soon. Reentering the horizon the perturbation of the gravitational potential (properly speaking the curvature, for details, see (80)) gives rise to matter perturbation via the *Poisson equation* $\nabla^2\Phi = 4\pi G\rho$. Summarising:

- Quantum fluctuations of the inflation field are excited during inflation. Being connected to the fluctuations in the metrics, these latter are stretched to cosmological scales,
- Gravity interacting with baryons and photons gives rise to scalar and tensor perturbations.

Although the inflationary model has been explored as a solid paradigm, departures from Gaussianity coming from the single-field or multi-field inflation models have been studied in the last years, and they are an object of a rich debate (86; 87).

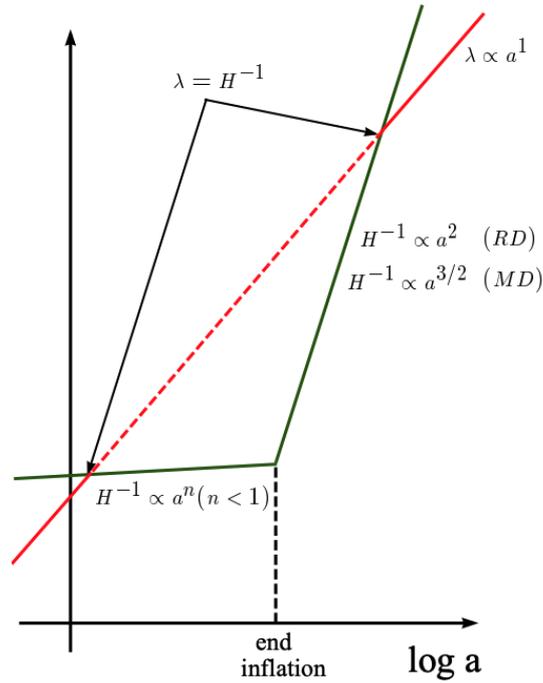


Figure 3.2 Schematisation of how the scale factor re-enters the Hubble radius after the inflationary epoch. The red line refers to the scale factor and the green line to the Hubble radius in different epochs. Picture from (13).

3.2 Gravitational instability

3.2.1 Fluid equations

Considering the equation of motion for a particle at physical position \mathbf{r} in the Newtonian regime, i.e. for small distances $x \ll H^{-1}$ and $v \ll 1$, I have

$$\ddot{\mathbf{r}} = -\nabla_{\mathbf{r}}\Phi, \quad (3.10)$$

where Φ is the gravitational potential, in the Newtonian description, is the potential induced by the local mass density $\rho(\mathbf{r})$

$$\Phi(\mathbf{r}) = G \int d^3\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \quad (3.11)$$

with G the Newtonian constant and the comoving coordinates are defined as $\mathbf{r} = a\mathbf{x}$, and so on $\nabla_{\mathbf{x}} = a\nabla_{\mathbf{r}}$. Taking the derivative of the physical coordinate with respect to physical time,

I get

$$\dot{\mathbf{r}} = \mathcal{H}\mathbf{x} + \mathbf{x}' \quad (3.12)$$

and for the second derivative

$$\ddot{\mathbf{r}} = \frac{1}{a}(\mathcal{H}'\mathbf{x} + \mathcal{H}\mathbf{x}' + \mathbf{x}'') = \frac{1}{a}\nabla\Phi \quad (3.13)$$

where the term proportional to the position is peculiar and arises from the comoving coordinates. Hence, I can define the peculiar potential ϕ

$$\Phi = \frac{1}{2}\mathcal{H}'\mathbf{x}^2 + \phi. \quad (3.14)$$

The peculiar potential ϕ is sourced only by density fluctuations that, under the assumption that dark energy is homogeneous, are dominated by matter fluctuations. The Poisson equation becomes

$$\nabla_{\mathbf{x}}^2\phi = \frac{3}{2}\Omega_{m,0}H_0^2\frac{\delta}{a}. \quad (3.15)$$

It is possible to split the equation of motion in terms of physical and conformal time - defined in 2.37 -

$$\ddot{\mathbf{x}} + 2H\dot{\mathbf{x}} = -\frac{\nabla\phi}{a^2}, \quad \mathbf{x}'' + \mathcal{H}\mathbf{x}' = -\nabla\phi. \quad (3.16)$$

Defining the canonical momentum

$$\mathbf{p} = am\mathbf{x}', \quad (3.17)$$

I get the equation of motion

$$\mathbf{p}' = -am\nabla_{\mathbf{x}}\phi. \quad (3.18)$$

Using the definition in Eq. 3.17, I can express the conservation of phase-space density, yielding the collisionless Boltzmann equation, known as the *Vlasov equation*

$$\begin{aligned} \frac{df}{d\tau} &= \frac{\partial f}{\partial \tau} + \frac{d\mathbf{x}}{d\tau} \cdot \frac{\partial f}{\partial \mathbf{x}} + \frac{d\mathbf{p}}{d\tau} \cdot \frac{\partial f}{\partial \mathbf{p}} \\ &= \frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \frac{\partial f}{\partial \mathbf{x}} + -am\nabla\phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \end{aligned} \quad (3.19)$$

Taking the zeroth and the first moment of the Vlasov equation, I get the continuity and Euler equations

$$\delta' + \nabla \cdot [\mathbf{v}(1 + \delta)] = 0, \quad (3.20)$$

$$v'_i + \mathcal{H}v_i + \mathbf{v} \cdot \nabla v_i = -\nabla_i \phi - \frac{1}{\rho} \nabla_i (\rho \sigma_{ij}) \quad (3.21)$$

where I have defined the density, the mean streaming velocity and the velocity dispersion as

$$\rho(\mathbf{x}, \tau) = \frac{m}{a^3} \int d^3p f(\mathbf{x}, \mathbf{p}, \tau), \quad (3.22)$$

$$v_i(\mathbf{x}, \tau) = \int d^3p \frac{p_i}{am} f(\mathbf{x}, \mathbf{p}, \tau) / \int d^3p f(\mathbf{x}, \mathbf{p}, \tau), \quad (3.23)$$

$$\sigma_{ij}(\mathbf{x}, \tau) = \int d^3p \frac{p_i}{am} \frac{p_j}{am} f(\mathbf{x}, \mathbf{p}, \tau) / \int d^3p f(\mathbf{x}, \mathbf{p}, \tau) - v_i(\mathbf{x})v_j(\mathbf{x}). \quad (3.24)$$

so that $\delta(\mathbf{x}, \tau) = 1 - \rho(\mathbf{x}, \tau)/\bar{\rho}$ where $\bar{\rho}$ represents the spatial average of the field $\rho(\mathbf{x}, \tau)$. The velocity dispersion is also referred to as anisotropic stress and describes the deviation from a single coherent flow.

3.2.2 Linear growth

Defining $\omega = \nabla \times \mathbf{v}$ as the vorticity and $\theta = \nabla \cdot \mathbf{v}$ as the divergence of the velocity field, if I neglect the quadratic terms, the continuity and Euler equations become

$$\theta' + \mathcal{H}\theta = -\nabla^2 \phi, \quad (3.25)$$

$$\omega' + \mathcal{H}\omega = 0. \quad (3.26)$$

The solution of the vorticity equation gives $\omega \simeq a^{-1}$, meaning any initial vorticity decays at the linear stage. Combining the two equations, it is possible to get a single equation for the density field

$$\delta''(\mathbf{x}, \tau) + \mathcal{H}(\tau)\delta'(\mathbf{x}, \tau) - \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(\mathbf{x}, \tau) = 0 \quad (3.27)$$

This equation reveals that different perturbations at different scales evolve independently, at least in the linear regime, because, Fourier transforming it, all modes would grow at the same rate. This means it can split into a spatial and a time-dependent part, and because the time derivatives appear up to the second order, I expect two independent solutions:

$$\delta(\mathbf{x}, \tau) = D_+(\tau)\Delta_+(\mathbf{x}) + D_-(\tau)\Delta_-(\mathbf{x}). \quad (3.28)$$

For the Einstein-de-Sitter (EdS) matter-only and radiation-only Universe case, analytical solutions exist. In the first case giving the following linear growth factor

$$D_+(\tau) = a(\tau), \quad D_-(\tau) = a(\tau)^{-3/2}. \quad (3.29)$$

The second solution decays very quickly in time, while the first one is the growing one. By applying the linearised continuity equation, it is possible to get a corresponding expression for θ

$$-\frac{\theta(\mathbf{x}, \tau)}{\mathcal{H}} = f_+ D_+(\tau) \Delta_+(\mathbf{x}) + f_- D_-(\tau) \Delta_-(\mathbf{x}), \quad (3.30)$$

where the logarithmic growth rate

$$f_{\pm} = \frac{\ln D_{\pm}}{\ln a} \quad (3.31)$$

has been introduced.

3.2.3 Standard perturbation theory

After treating the fluid equations in the linear regime, I can now return to the full equation without neglecting the quadratic terms. In order to develop an easier approach to modelling the non-linear regime, it is usual to work in Fourier space, where the Euler and continuity equations become

$$\delta(\mathbf{k}') + \theta(\mathbf{k}) = - \int \frac{d^3 q}{2\pi^3} \frac{d^3 q'}{2\pi^3} (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{q} - \mathbf{q}') \alpha(\mathbf{q}, \mathbf{q}') \theta(\mathbf{q}) \delta(\mathbf{q}'), \quad (3.32)$$

$$\theta'(\mathbf{k}') + \mathcal{H}\theta(\mathbf{k}) + \frac{3}{2}\Omega_m(a)\mathcal{H}^2\delta(\mathbf{k}) = - \int \frac{d^3 q}{2\pi^3} \frac{d^3 q'}{2\pi^3} (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{q} - \mathbf{q}') \beta(\mathbf{q}, \mathbf{q}') \delta(\mathbf{q}) \delta(\mathbf{q}'). \quad (3.33)$$

where the coupling kernels are defined as

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) \equiv \frac{\mathbf{k}_1 \cdot (\mathbf{k}_1 + \mathbf{k}_2)}{k_1^2} \quad (3.34)$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) \equiv \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_2}{k_1} + \frac{k_1}{k_2} \right) + \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}. \quad (3.35)$$

The fluid Eqs 3.32 and 3.33 are non-linear coupled differential equations for the density and velocity divergence. In general, an exact solution does not exist, but it is possible to try to solve with a perturbative approach, where $\delta \ll 1$ and $\theta \ll 1$. Standard Perturbation Theory (see (15) for an extensive review) aims to solve the fluid equations using the following power law ansatz

$$\delta(\mathbf{k}, \tau) = \sum_i^{+\infty} a^i(\tau) \delta^i(\mathbf{k}), \quad \theta(\mathbf{k}, \tau) = -\mathcal{H}(\tau) \sum_i^{+\infty} a^i(\tau) \tilde{\theta}^i(\mathbf{k}). \quad (3.36)$$

The expansion is in powers of the linear density field, and it is possible to write the n -th order solutions in the form

$$\delta^{(n)}(\mathbf{k}) = \prod_{m=1}^n \left\{ \int \frac{d^3 q_m}{(2\pi)^3} \delta^{(1)}(\mathbf{q}_m) \right\} F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) (2\pi)^3 \delta^{(D)}(\mathbf{k} - \mathbf{q}_1^n), \quad (3.37)$$

$$\tilde{\theta}^{(n)}(\mathbf{k}) = \prod_{m=1}^n \left\{ \int \frac{d^3 q_m}{(2\pi)^3} \delta^{(1)}(\mathbf{q}_m) \right\} G_n(\mathbf{q}_1, \dots, \mathbf{q}_n) (2\pi)^3 \delta^{(D)}(\mathbf{k} - \mathbf{q}_1^n), \quad (3.38)$$

where the kernels can be obtained by means of the recursive relations

$$F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} \left[(2n+1)\alpha(\mathbf{q}_1^m, \mathbf{q}_{m+1}^n) F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \right. \\ \left. + \beta(\mathbf{q}_1^m, \mathbf{q}_{m+1}^n) G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \right], \quad (3.39)$$

$$G_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} \left[3\alpha(\mathbf{q}_1^m, \mathbf{q}_{m+1}^n) F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \right. \\ \left. + 2n\beta(\mathbf{q}_1^m, \mathbf{q}_{m+1}^n) G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) \right]. \quad (3.40)$$

If we explicit the second-order density kernels, we found

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{17}{21} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_2}{k_1} + \frac{k_1}{k_2} \right) + \frac{2}{7} \left[\frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} - \frac{1}{3} \right], \quad (3.41)$$

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{3}{21} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_2}{k_1} + \frac{k_1}{k_2} \right) + \frac{4}{7} \left[\frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} \right]. \quad (3.42)$$

Although second-order kernels are symmetric, generalised Eq. 3.39 and Eq. 3.40 are not symmetrised over the arguments yet.

3.2.4 Lagrangian perturbation theory

An alternative way to deal with nonlinear modelling has been proposed in (88), known as Lagrangian Perturbation Theory (LPT). The idea is to trace the motion of individual fluid elements $\mathbf{x}(\mathbf{q}, \tau)$ with initial spatial condition \mathbf{q} throughout the so-called displacement field $\Psi(\mathbf{q}, \tau)$

$$\mathbf{x}(\mathbf{q}, \tau) = \mathbf{q} + \Psi(\mathbf{q}, \tau). \quad (3.43)$$

By using the conservation of mass, it is possible to write a relation between the Jacobian of the transformation and the density perturbation

$$J(\mathbf{q}, \tau) \equiv \det[\delta_{ij} + \nabla_{q,j}\Psi_i] = \frac{1}{1 + \delta(\mathbf{x}, \tau)}. \quad (3.44)$$

Taking the divergence of Eq. 3.43, it is possible to obtain the following equation for the displacement field Ψ

$$J(\mathbf{q}, \tau) \nabla \cdot \left[\frac{d^2 \Psi}{d\tau^2} + \mathcal{H}(\tau) \frac{d\Psi}{d\tau} \right] = \frac{3}{2} \Omega_m \mathcal{H}^2 (J - 1). \quad (3.45)$$

By Fourier transformation, it turns out that

$$\delta(\mathbf{k}, \tau) = \int d^3 q e^{-i\mathbf{k} \cdot \mathbf{q}} \sum_{m=1}^{\infty} \frac{[-i\mathbf{k} \cdot \Psi]^m}{m!}, \quad (3.46)$$

that shows how even a linear displacement $\Psi^{(1)}$ contributes to all orders. Particularly at the linear order

$$\nabla_q \cdot \Psi^{(1)} = -D(\tau) \delta_0(\mathbf{q}), \quad (3.47)$$

which is known as the Zel'dovich approximation. Going to higher-order corrections, it reads as

$$\Psi(\mathbf{q}, \tau) = \Psi^{(1)}(\mathbf{q}, \tau) + \Psi^{(2)}(\mathbf{q}, \tau) + \Psi^{(3)}(\mathbf{q}, \tau) + \Psi^{(4)}(\mathbf{q}, \tau) + \dots, \quad (3.48)$$

where, assuming no velocity curl modes, they determined as follows (89; 90)

$$\Psi^{(n)}(\mathbf{q}, \tau) = D_n(\tau) \nabla_q \phi_n(\mathbf{q}) \quad (3.49)$$

where D_n are the n -th growth factors. Up to the second order, the growth factors read as (15; 91)

$$D_1(\tau) = D(\tau), \quad (3.50)$$

$$D_2(\tau) = -\frac{3}{14} D(\tau) \Omega_m^{-\frac{1}{43}}. \quad (3.51)$$

The solution for the potential read as

$$\nabla_q^2 \phi_1(\mathbf{q}) = -\delta_0(\mathbf{q}), \quad (3.52)$$

$$\nabla_q^2 \phi_2(\mathbf{q}) = -\mathcal{G}_2[\phi_1(\mathbf{q}), \phi_1(\mathbf{q})] \quad (3.53)$$

where I have defined the second order *Galileian* operator.

$$\mathcal{G}_2(A, B) = \nabla_{ij} A \nabla_{ij} B - \nabla^2 A \nabla^2 B. \quad (3.54)$$

Up to the second order, Lagrangian evolution rules as

$$\mathbf{x}(\mathbf{q}, \tau) = \mathbf{q} + D_1(\tau) \nabla_q \phi_1(\mathbf{q}) + D_2(\tau) \nabla_q \phi_2(\mathbf{q}) \quad (3.55)$$

which is the initial condition for the position, particularly important for an N-body simulation environment.

3.3 From matter to galaxies

For the purpose of finding a clear picture of the physical dynamics of structure formation in the non-linear regime, I mainly distinguish two approaches:

1. Typically, small-scales are described with N-body simulations, which prove to be a powerful means of solving the dynamics of particles under the influence of gravitational evolution. However, these simulations are subject to limitations in terms of both spatial and temporal resolution, as well as the number of points used to trace the matter distribution. Additionally, while simulations that do not take collisions into account offer an accurate representation of dark matter evolution, the modelling of baryonic matter is often achieved through the use of hydrodynamic simulations, which incorporate semi-analytic models and phenomenology for star formation and energy feedback,
2. In general, it is possible to define an approach that describes the mapping between the distribution of dark matter and that of galaxies. Perturbation theory represents an effective means of achieving this, requiring introducing a range of bias parameters to describe the mapping in the small-scale regime. The halo model (92) and the halo abundance matching (93) are some of the notable alternatives within this category.

3.3.1 Halo model

Spherical collapse

The spherical collapse model (94) represents a well-established method for understanding the non-linear evolution of an isolated, spherical overdensity within the context of large-scale structure cosmology. The model represents an example of a non-linear gravitational system

that can be solved exactly, and it is often used to parametrize the growth of a single, spherical overdensity Δ detached from the background density at a given time t

$$\Delta(t) = 1 + \frac{9}{2} \frac{(\theta - \sin(\theta))^2}{(1 - \cos(\theta))^3} \quad (3.56)$$

with θ spanning in the range $[0, 2\pi]$ while shell crossing is not considered. From Eq. 3.56, I can distinguish three main phases:

- Expanding phase: The overdensity first begins to expand following the background dynamics, but it stops growing at $\theta = \pi$ due to the decelerating gravitational field. This moment is also known as the turn-around time t_a ,
- Collapsing phase: The overdensity evolves independently with respect to the background and starts collapsing toward a singularity of vanishing size. The spherical collapse model remains predictive during this phase, as long as the shells do not cross each other,
- Virialization: After reaching the turn-around time t_a , the density perturbation has reached values above unity, and the system enters a phase of violent relaxation, also known as the Lynden-Bell mechanism (95). This is where the interaction of a single particle with the rapidly changing gravitational potential of the system leads to an efficient equipartition of energy among the particles, causing them to deviate from purely radial trajectories and leading to complete virialization of the system. This typically occurs at $\theta = 3\pi$.

In the linear regime, we know that all perturbations are growing proportionally to the growth factor of the background, which considering the Einstein de-Sitter case scales as $D(t) = a(t) \propto \left(\frac{t}{t_a}\right)^{\frac{2}{3}}$. For $\theta \ll 1$, the linear over-density must match with, which fixes the proportionality constant, so that

$$\Delta_0(t) = \frac{3}{20} \frac{6\pi}{(t/t_a)^{\frac{2}{3}}} (6\theta - 6\sin\theta)^{\frac{2}{3}}. \quad (3.57)$$

When the virial theorem is applied, the process is considered complete when the average potential energy is equal to twice the average kinetic energy. This happens when $R_{vir} = R_{ta}/2$, and assuming that this occurs at $\theta = 2\pi$, it is possible to see that $\Delta_{vir} \approx 178$. In comparison, the linearly extrapolated density perturbation in Eq. 3.57 is $\delta_{cr} \equiv \Delta_0(t_{vir}) \approx 1.686$, which is defined as the critical collapse density.

Halo mass function

From the spherical collapse model, I outlined the idea that a halo forms whenever the overdensity exceeds a critical value. To predict the abundance of dark matter halos during the cosmic expansion, we can start identifying a halo mass M and filtering it on a scale R that is related to the halo mass through the background density $(4\pi/3)\bar{\rho}R^3$. Integrating the probability density functions over densities larger than the critical one, we get the fractional volume relative to halos corresponding to size R or bigger

$$F_V(\delta_{0,R} > \delta_{\text{cr}}) = \frac{1}{\sqrt{2\pi\sigma^2(R)}} \int_{\delta_{\text{cr}}}^{\infty} d\delta e^{-\frac{\delta^2}{2\sigma^2(R)}}, \quad (3.58)$$

where $\sigma^2(R)$ completely characterise the overdensity $\delta_{0,R}^2$ because of Gaussianity assumed from the inflationary paradigm. From the assumption of Press and Schechter (96), the associated mass fraction $F_M(> M)$ is the same as fractional volume. Defining the comoving number density of halos per logarithmic mass bin as the halo mass function $\bar{n}_h(M)$ so that

$$F_M(> M) = \frac{1}{\bar{\rho}} \int_M^{\infty} d\ln M' M' \bar{n}_h(M') = F_V(\delta_{0,R} > \delta_{\text{cr}}). \quad (3.59)$$

Hence, by differentiating, I can obtain the so-called Press-Schechter mass function (96)

$$\bar{n}_h(M) = f(\sigma(M)) \frac{\bar{\rho}}{M} \frac{d\ln\sigma(M)^{-1}}{dM} \quad (3.60)$$

where $f(\sigma(M))$ is the amount of mass encapsulated in perturbation of typical scale R , that can be expressed as

$$f(\sigma(M)) = \sqrt{\frac{2}{\pi}} \frac{\delta_{\text{cr}}}{\sigma(M)} e^{-\frac{\delta_{\text{cr}}^2}{2\sigma^2(M)}}. \quad (3.61)$$

This result is known as the so-called Press-Schechter theory (PS) (96), and other works extended the theory taking into account the cloud-in-cloud problem, that is the possibility that over-densities smaller than the critical one can be part of halo with size bigger than R (97). For instance, the well-known Sheth and Tormen (98) model proposed an extension of the mass function based on elliptical collapse, modifying Eq.3.61, and provided a better fit compared to N-body simulations (98; 99).

However, while the halo mass function provides insight into the average number density of halos, one may also investigate its variations with respect to changes in the background matter density. In the case of a spherical perturbation, denoted by Δ , which modifies the local matter density relative to the background density as $\rho = (1 + \Delta)\bar{\rho}$, the response to this

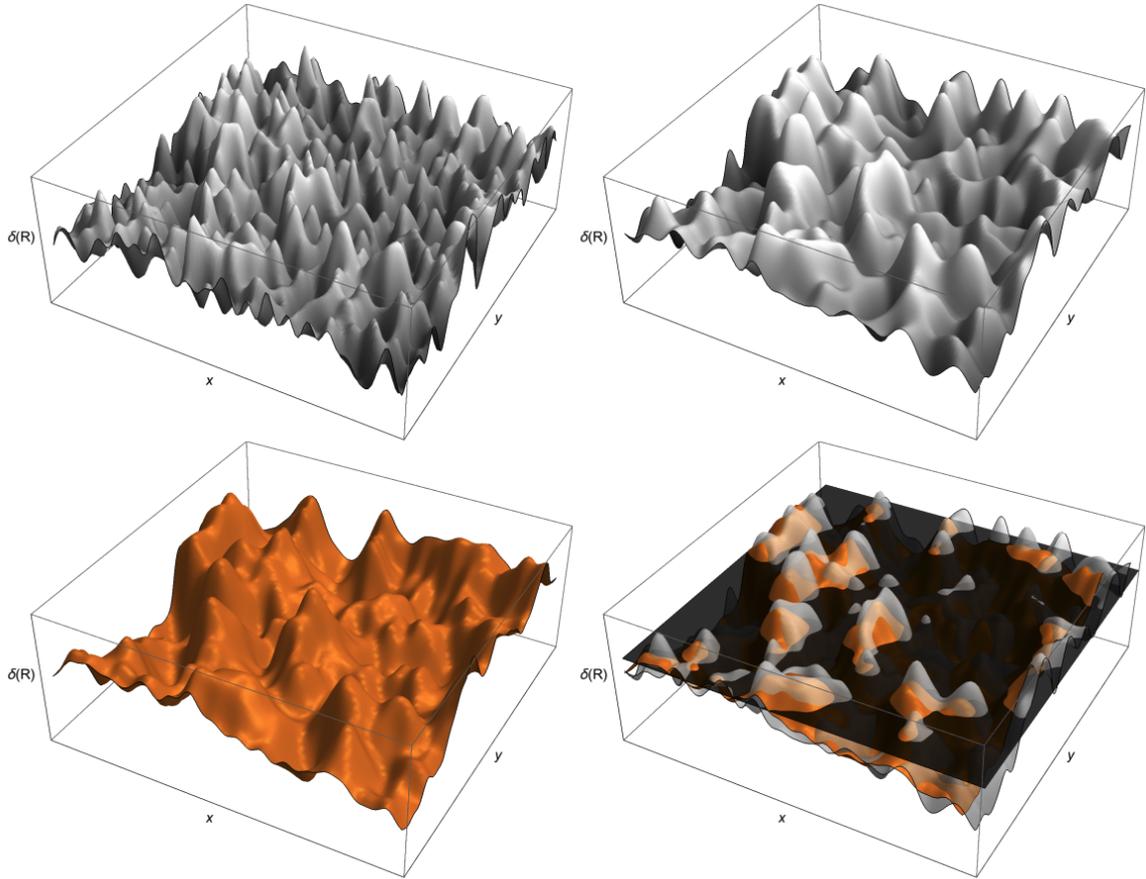


Figure 3.3 Illustration of a two-dimensional Gaussian density field $\delta(R)$ smoothed on different scales $R : [0.4, 1, 2]h^{-1}\text{Mpc}$, respectively *top-left*, *top-right*, *bottom-left*. The *bottom right* panel refers to $R = 2h^{-1}\text{Mpc}$ along with the spherical collapse threshold. Picture from (14).

perturbation can be quantified through the peak-background bias parameter (100)

$$b \equiv \frac{1}{\bar{n}_h|_{\Delta=0}} \frac{\partial \bar{n}_h|_{\Delta}}{\partial \Delta} \Big|_{\Delta=0} \quad (3.62)$$

Considering Δ as a spatial perturbation of the background matter density, I can identify the local abundance of halos $n_h(\mathbf{x}) = \bar{n}_h|_{\Delta}$ through which it is possible to predict the relation between matter and halos perturbations

$$\delta_h(\mathbf{x}) = \frac{n_h(\mathbf{x})}{\bar{n}_h} - 1 = b\Delta, \quad (3.63)$$

that explain how halos are biased tracers of the underlying dark matter perturbations. When applying Eq. 3.61 to Eq. 3.62, I find,

$$b(M) = \frac{\sigma(M)^2}{\delta_{\text{cr}}} \frac{d \ln f(\sigma(M))}{d\sigma(M)}, \quad (3.64)$$

which has been tested with many mass functions against N-body simulations (101).

Halo occupation distribution

Once we have found that dark matter halos are tracers of the underlying dark matter fluctuations, we can expect this can be applied to galaxies. Actually, the spherical collapse model and its assumptions do not hold in galaxy formation physics. However, an attempt to give a picture of how galaxies trace dark matter perturbation in a similar scenario has been given in the so-called halo occupation distribution (HOD) framework (102). In most implementations of the HOD approach, I distinguish between “*central*” and “*satellite*” galaxies. Every halo hosts one or zero central galaxies close to the centre of mass of the halo, and all other galaxies are satellite ones. Let me call N_c and N_s the number of central and satellite galaxies in a given halo so that $N_c \in [0, 1]$ and $N_s \in [0, 1, 2, \dots]$. In the model of (103) the number of central and satellite galaxies is given by

$$\langle N_c(M) \rangle = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln M - \ln M_{\min}}{\sigma_{\ln M}} \right) \right], \quad (3.65)$$

$$\langle N_s(M) \rangle = \langle N_c(M) \rangle \times \begin{cases} \left(\frac{M - M_0}{M_1} \right)^\alpha, & M > M_0 \\ 0, & M \leq M_0 \end{cases} \quad (3.66)$$

where M_0 is the minimum mass for a halo to host a satellite galaxy and $M_1 + M_0$ is the typical mass of a satellite galaxy, while $\sigma_{\ln M}$ is the logarithmic scatter between galaxy luminosity and halo mass and M_{\min} represents a cut-off for the probability for hosting a galaxy. In general, since the existence of satellite galaxies is conditioned by the presence of a central galaxy, it is a usual practice to define the parameter \mathcal{N}_s so that $N_s = N_c \mathcal{N}_s$. Assuming this condition, it is possible to express the average number of galaxies as

$$\bar{n}_g = \int d \ln M \bar{n}_h(M) \langle N_c(M) \rangle [1 + \langle \mathcal{N}(M)_s \rangle], \quad (3.67)$$

where $\langle N_c | M \rangle$, $\langle \mathcal{N}_s \rangle$ are respectively the values of central and satellite galaxy numbers in a halo of mass M and the parameter M_{\min} can be adjusted to match the number density of

some observed population of galaxies. Therefore, I can express the linear bias parameter as

$$b_g = \frac{1}{\bar{n}_g} \int d \ln M \bar{n}_h(M) \langle N_c(M) [1 + \langle \mathcal{N}(M)_s \rangle] b(M). \quad (3.68)$$

This equation represents, according to the peak-background argument, the linear response of the mean galaxy density to a change in the background density. The galaxy bias depends on the same properties of the galaxy sample, such as their luminosities.

3.3.2 Perturbative bias

So far, the HOD model has provided us with a picture of galaxy fluctuations in the nonlinear regime, making the assumption that occupation numbers only depend on the mass of the hosting halo. On the other hand, perturbative approaches to galaxy bias address the problem, at least on a large scale, by adopting an alternative point of view. Coming back to the peak background split argument, the linear relation between galaxies and dark matter perturbations $\delta_g(\mathbf{x}) = b\delta(\mathbf{x})$ holds only at large scales and where the evolution is far from being in the nonlinear regime due to the assumption that $\delta(\mathbf{x}) \ll 1$. An alternative to the HOD model and a general extension to nonlinear small scale has been provided in (104). The main idea is expanding the galaxy density contrast in a perturbative series, a power-law function of matter fluctuations

$$\delta_g(\mathbf{x}) = \sum_{n=1} \frac{b_n}{n!} \delta(\mathbf{x})^n \quad (3.69)$$

where I recognize the linear bias $b = b_1$. The perturbative approach encapsulates the spherical collapse and also takes into account other additional effects emerging at higher order. One of them is the impact of the tidal field (105; 106). At the second order, the perturbative galaxy bias expansion reads

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + \frac{b_2}{2} \delta(\mathbf{x})^2 + \gamma_2 \mathcal{G}_2(\Phi|\mathbf{x}), \quad (3.70)$$

where, following (105), $\mathcal{G}_2(\Phi|\mathbf{x})$ is the so-called second-order *Galileon* operator, a Galileian invariant term given by

$$\mathcal{G}_2(\Phi|\mathbf{x}) = (\nabla_{ij} \Phi_v)^2 - (\nabla^2 \Phi)_v^2, \quad (3.71)$$

and generalised in Eq. 3.54, where Φ_v is the velocity potential, defined as $\nabla^2 \Phi_v(\mathbf{x}, \tau) \equiv \theta \equiv \nabla \cdot v$. Concerning the expansion in Eq. 3.70:

- The linear bias term, b_1 , quantifies the relationship between the galaxy density and the dark matter density on large scales. It is defined as the ratio of the galaxy density fluctuation to the dark matter density fluctuation. In the simplest models, it is assumed

that b_1 is a constant, independent from scale and time, and that the galaxy density is proportional to the dark matter density. However, studies have shown that the value of b_1 can vary with scale, redshift, and galaxy properties. The b_1 parameter has been extensively studied in the literature, with a focus on understanding the underlying physical mechanisms that determine the galaxy-dark matter connection. Several works have investigated this relationship using different observational and theoretical approaches, such as halo models, perturbation theory, and large-scale structure simulations (107; 108; 109),

- The b_2 parameter quantifies the non-linearity in the relationship between the distribution of galaxies and the distribution of dark matter. This non-linearity arises from the complex physics of galaxy formation and bias, which deviates from a simple linear relation between the galaxy and dark matter density fields. The b_2 parameter has been studied in a number of theoretical and observational works, see (14) for a review. These works have examined the relationship between the distribution of galaxies and the distribution of dark matter on various scales and in different cosmological models, providing insights into the underlying physical mechanisms and the dependence on cosmological parameters,
- The term γ_2 is a parameter used in the study of the large-scale structure of the universe that describes the relationship between the density of galaxies and the gravitational potential field. Without considering any source of primordial non-gaussianity, this relationship is commonly assumed to be linear, with the density of galaxies being proportional to the gravitational potential field. However, the γ_2 parameter accounts for any nonlinear dependencies of the galaxy distribution on the gravitational potential field. This is particularly relevant in the presence of dark matter, which is known to significantly impact the gravitational potential field. The γ_2 term has been studied in several works; see (14) for an extensive review.

The relationships between these three bias parameters have been extensively analyzed in the literature.

- The relationship $b_2(b_1, \gamma_2)$ reported in (100), expressed as:

$$b_2 = 0.412 - 2.143b_1 + 0.929b_1^2 + 0.008b_1^3 + \frac{4}{3}\gamma_2 \quad (3.72)$$

is a fitting formula that has been derived from separate universe simulations. It is important to mention that the $\frac{4}{3}\gamma_2$ term is not given in (100) due to the difference in the bias expansion used in that study,

- Several studies have focused on the relationship $\gamma_2(b_1)$. Specifically, as will be presented later, the local Lagrangian relationships (see Eq 4.68) have been studied, and some works (100; 110) have slightly invalidated the $\gamma_{2,LL}(b_1)$ where *LL* refers to local Lagrangian. An alternative estimation of the tidal bias parameter in the context of the excursion set approach was discussed in (111). They make a prediction for γ_2 , which can be represented by the following quadratic fit:

$$\gamma_2(b_1) = 0.524 - 0.547b_1 + 0.046b_1^2. \quad (3.73)$$

This fit provides a slightly better description of the measurements compared to the assumption of local Lagrangian.

Nonlinear quantities like δ^n receive contributions from all scales, even large. A common practice is usually to impose a cutoff smoothing δ in order to remove contributions below a certain scale Λ , an arbitrary quantity we fix to parametrise our ignorance of small-scale physics. An approach to remove these sources of issues has been developed in the so-called *renormalized halo bias* (83), for which large-scale contributions can be systematically removed by adding local counterterms. Each of the bias parameters that appear in the bias perturbative expansion is free, and it should be marginalised in a cosmological analysis of a survey. Their value parametrises the small-scale physics of galaxy formation imprinted into the large-scale signal. The perturbative bias approach, which I will follow in this Thesis, substantially differs from the HOD framework for which there is a model to treat small-scale processes. Contrary, the perturbative bias approach is an agnostic strategy to model small-scale physics, and for the general approach to the problem, it can be used and extended to different tracers of the matter fluctuations, such as, for example, the 21 cm emission line.

4

Matter and galaxy clustering statistics

So far, we have briefly presented the theory of gravitational instability to describe the generation and subsequent amplification of small departures from homogeneity and their evolution into observable extragalactic objects. Their spatial distribution encodes precious information on the geometry of the Universe and on the properties of the main components (dark matter, dark energy, baryons, photons, neutrinos) that contribute to the mass-energy budget of the system. Efficient extraction of this precious information can be obtained by measuring the statistical properties of this distribution and their comparison with theoretical predictions that, as we shall see, can be obtained using either a standard or effective perturbative theory approach. In their early stage of evolution, cosmological fields are well described by Gaussian statistics which, in turn, only require considering two-point statistics for a complete statistical characterization. Deviations from Gaussianity not present in the initial conditions are induced by non-linear dynamical processes as well as from the matter-to-galaxy mapping known as galaxy bias. In this case, higher-order statistics are required to collect all available information. The modelling and the measurement of the first and most rewarding of these statistics, the 3-point correlation function, constitutes the focus of this and the following chapters. This has provided valuable cosmological information during the years (112; 113; 114).

4.1 Correlation functions of density fluctuations

Correlation functions are central to Galaxy Clustering, which involves studying the distribution and clustering of galaxies. As discussed in the previous chapter, correlation functions can be represented in either configuration space or Fourier space. These functions are closely related to the statistical properties of inhomogeneities and how they evolve over time. Understanding correlation functions can provide insight into the underlying physical processes

driving the distribution and clustering of galaxies, as well as the overall structure of the universe.

4.1.1 Statistical description

Let us consider one perturbation $g(\mathbf{x})$ associated with a *random field*. This term denotes a set of functions $g_n(\mathbf{x})$, each coming with a probability \mathcal{P}_n . The set is referred to as the *ensemble*, and the individual function is the so-called *realization*. Let us define the two-point correlator

$$\langle g(\mathbf{x})g(\mathbf{x}') \rangle \equiv \sum_n \mathcal{P}_n g_n(\mathbf{x})g_n(\mathbf{x}'). \quad (4.1)$$

The random field is typically assumed to be statistically homogeneous and isotropic. This means that the probabilities assigned to its possible realizations are unchanged under translations, rotations, and a transformation that reverses the orientation of the coordinate system. Homogeneity, or translation invariance, means that the probability assigned to a realization $g_n(\mathbf{x})$ is the same as that assigned to $g_n(\mathbf{x} + \mathbf{X})$ for any fixed \mathbf{X} . One realization can be used to generate the entire random field by letting \mathbf{X} take on all possible values. This is known as the *ergodic* property of the field. According to the ergodic theorem, the ensemble average $\langle g(\mathbf{x})g(\mathbf{x}') \rangle$ can be considered as a spatial average at a fixed $\mathbf{x}' - \mathbf{x}$ for a single realization of the ensemble, as in the realistic case in which we have a single cosmological realisation - our universe -, and the same holds for higher-order correlators.

It is possible to Fourier expand the perturbation; working in a box of comoving size L , we have

$$g(\mathbf{x}) = \frac{1}{L^3} \sum_n g_n e^{-i\mathbf{k}_n \cdot \mathbf{x}}, \quad (4.2)$$

$$g_n = \int d^3 x g(\mathbf{x}) e^{-i\mathbf{k}_n \cdot \mathbf{x}}. \quad (4.3)$$

where $k_n = 2n\pi/L$. In the limit $L \rightarrow +\infty$ we have

$$g(\mathbf{x}) = \frac{1}{L^3} \int d^3 k g_n e^{-i\mathbf{k}_n \cdot \mathbf{x}}, \quad (4.4)$$

$$g_n = \int d^3 x g(\mathbf{x}) e^{-i\mathbf{k}_n \cdot \mathbf{x}}. \quad (4.5)$$

The simplest type of random field is the Gaussian random field, defined as one whose Fourier coefficients have no correlation except for the reality condition

$$g(-\mathbf{k}) = g^*(\mathbf{k}) \quad (4.6)$$

Going into configuration space, let me remind that the quantity $g(\mathbf{x})$ is a superposition of Fourier modes. Because of the central limit theorem, the sum of uncorrelated quantities has a Gaussian probability distribution. We can conclude that the probability distribution $g(\mathbf{x})$ at a given point is Gaussian with mean-square being

$$\sigma_g^2(\mathbf{x}) \equiv \langle g^2(\mathbf{x}) \rangle = \frac{1}{(2\pi)^3} \int_0^{+\infty} P_g(k) d^3k = \int_0^{+\infty} \mathcal{P}_g(k) \frac{dk}{k}. \quad (4.7)$$

We can note that $\mathcal{P}_g(k)$ represents the contribution to σ^2 per unit logarithmic interval of k .

4.1.2 Two-point statistics

Let us apply the correlation formalism to the cosmological density field traced by either mass particles or discrete objects. Considering the perturbed density field, which departure from homogeneity is characterised by the quantity $\delta(\mathbf{x})$ that is so spatial-dependent, it is possible to define the so-called two-point correlation function as follows

$$\xi(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle = \int d\delta(\mathbf{x}_1) d\delta(\mathbf{x}_2) \mathcal{P}[\delta(\mathbf{x}_1) \delta(\mathbf{x}_2)] \delta(\mathbf{x}_1) \delta(\mathbf{x}_2). \quad (4.8)$$

where the quantity $\mathcal{P}[\delta(\mathbf{x}_1) \delta(\mathbf{x}_2)]$ represents the probability density function (PDF) associated to the pair $\delta(\mathbf{x}_1) \delta(\mathbf{x}_2)$ and expresses the probability

$$\mathcal{P}[\delta(\mathbf{x}_1) \delta(\mathbf{x}_2)] \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \quad (4.9)$$

for a field δ to have values between $\delta(\mathbf{x}_i)$ and $\delta(\mathbf{x}_i) + d\delta(\mathbf{x}_i)$ with $i = 1, 2$. Because of statistical homogeneity and isotropy, Eq. 4.8 is a function of the modulus of the separation $r = |\mathbf{x}_1 - \mathbf{x}_2|$. To give a physical picture, let us consider a discrete sampling of the continuous field $\delta(\mathbf{x})$. This is known to be a Poisson process, where the probability of finding an object within an element of volume V is proportional to the density, following a Poissonian distribution with intensity $\rho(\mathbf{x}) \delta V / m$ where m is the mass of the discrete tracer. Taking the ensemble average over the Poisson distribution, denoted as $\langle \rangle_{\times}$, we get

$$\langle n(\mathbf{x}) \rangle_{\times} = \bar{n} [1 + \delta(\mathbf{x})] = \bar{n} + \delta n(\mathbf{x}), \quad (4.10)$$

where \bar{n} refers to the ensemble-averaged number density. Taking the average number of objects given two coordinates $\mathbf{x}_1, \mathbf{x}_2$ we obtain

$$\langle n(\mathbf{x}_1) n(\mathbf{x}_2) \rangle_{\times} = \bar{n}^2 [1 + \delta(\mathbf{x}_1)] [1 + \delta(\mathbf{x}_2)] + \bar{n} [1 + \delta(\mathbf{x}_1)] \delta_D(\mathbf{x}_1 - \mathbf{x}_2). \quad (4.11)$$

Taking the average over the ensemble, we get

$$\langle \langle n(\mathbf{x}_1)n(\mathbf{x}_2) \rangle_{\times} \rangle = \bar{n}^2[1 + \xi(r)] + \bar{n} \delta_D(\mathbf{x}_1 - \mathbf{x}_2). \quad (4.12)$$

The physical meaning is that if $\xi > 0$, the objects indicate an excess probability over the Poisson expectation, and $\xi < 0$ indicates the opposite case.

If the density perturbations are Gaussian, it means the probability density functions involved in the computation of higher-order statistics is given through the following PDF

$$\mathcal{P}[\delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_n)] = \frac{1}{\sqrt{(2\pi)^n}} \exp \left[-\frac{1}{2} \sum_{i,j=1}^n \delta(\mathbf{x}_i) \Theta_{ij}^{-1} \delta(\mathbf{x}_j) \right], \quad (4.13)$$

where the matrix Θ encodes the correlation between fields at various positions, i.e. $\Theta_{ij} = \xi(|\mathbf{x}_i - \mathbf{x}_j|)$. This stresses the importance of the two-point correlation function, as it completely characterises an N-point PDF under the Gaussian assumption.

So far, we have presented the two-point correlation function in the so-called configuration space. Actually, even more used, the Fourier space is an optimal space to deal with clustering statistics because the Fourier coefficients $\delta(\mathbf{k})$ evolve independently in the linear regime, and it facilitates the description in the nonlinear regime. For this purpose, let us define the two-point statistics in Fourier space as the inverse Fourier transform of the correlator shown in Eq. 4.8 as

$$\begin{aligned} \langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \rangle &= \int d^3x_1 d^3x_2 e^{-i(\mathbf{k}_1 \cdot \mathbf{x}_1 + \mathbf{k}_2 \cdot \mathbf{x}_2)} \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle \\ &= (2\pi)^3 \delta_D(\mathbf{k}_1 - \mathbf{k}_2) \int d^3r e^{-i\mathbf{k}_1 \cdot \mathbf{r}} \xi(\mathbf{r}) \\ &= (2\pi)^3 P(k_1) \delta_D(\mathbf{k}_1 - \mathbf{k}_2), \end{aligned} \quad (4.14)$$

where δ_D is the Dirac delta, $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ and it turns out the definition of the power spectrum $P(k)$ as the Fourier transform of the two-point correlation function, results known as the Wiener-Kinchin theorem

$$P(k) = (2\pi)^3 \int d^3r e^{-i\mathbf{k} \cdot \mathbf{r}} \xi(\mathbf{r}), \quad (4.15)$$

$$\xi(r) = (2\pi)^3 \int d^3k e^{i\mathbf{k} \cdot \mathbf{r}} P(\mathbf{k}). \quad (4.16)$$

Adopting the same procedure applied before, we can treat the case in which the power spectrum is computed for discrete tracers. It turns out that the discrete power spectrum is

modulated by a constant term, the shot noise, that contributes on all scales

$$\langle\langle\delta n(\mathbf{k}_1)\delta n(\mathbf{k}_2)\rangle\rangle_{\times} = (2\pi)^3 \left[P(k) + \frac{1}{\bar{n}} \right], \quad (4.17)$$

where we have focused on the perturbations to the number density $\delta n(\mathbf{k}_1)$.

4.1.3 Higher-order statistics

The power spectrum is a well-defined quantity for almost all homogenous fields, and it is particularly useful when we deal with Gaussian fields. This means that any odd expectation value of a Gaussian field vanishes, while even expectations can be expressed as a sum of all possible two-point contractions. This is known as the *Wick theorem*, for which

$$\langle\delta(\mathbf{k}_1)\dots\delta(\mathbf{k}_{2p+1})\rangle = 0 \quad (4.18)$$

$$\langle\delta(\mathbf{k}_1)\dots\delta(\mathbf{k}_{2p})\rangle = \sum_{\text{all pair associated}} \prod_{p \text{ pairs } (i, k)} \langle\delta(\mathbf{k}_i)\delta(\mathbf{k}_j)\rangle. \quad (4.19)$$

The physical meaning is that all the statistical properties of the random field $\delta(\mathbf{k})$ are entirely determined by the shape and normalisation of $P(k)$.

In general, it is possible to define the so-called higher-order correlation functions as the connected part of the joint ensemble average of the density in an arbitrary number of locations. They can be expressed as

$$\begin{aligned} \xi_N(\mathbf{x}_1, \dots, \mathbf{x}_N) &= \langle\delta(\mathbf{x}_1, \dots, \mathbf{x}_N)\rangle_c \\ &= \langle\delta(\mathbf{x}_1, \dots, \mathbf{x}_N)\rangle - \sum_{\mathcal{S} \in \mathcal{P}(\{\mathbf{x}_1, \dots, \mathbf{x}_N\})} \prod_{s_i \in \mathcal{S}} \delta(\mathbf{x}_{s_1}, \dots, \mathbf{x}_{s_n}) \end{aligned} \quad (4.20)$$

where $N \geq 3$ and the sum and the products are made over the proper partition $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, the subscript $\langle\rangle_c$ indicates the connected part and s is a subset of the partition \mathcal{S} . The decomposition between the connected part and the non-connected one can be explained throughout Fig. 4.1, for which any ensemble average can be decomposed into a product of connected parts. In the case of Gaussianity, all the connected correlation functions are zero except for the two-point correlation function as a consequence of the Wick theorem.

For non-Gaussian overdensity fields, in general, the two-point correlation function is not enough to give a complete description, and we need to rely on higher-order correlation statistics. A visual impression of how non-Gaussianity affects the clustering properties can be obtained from Fig 4.2. The two plots, taken from (16) represent a mock, though realistic, Gaussian distribution of SDSS galaxies in a slice 50Mpc/ h thick (left) and a non-Gaussian

$$\langle \delta_1 \delta_2 \delta_3 \rangle = \text{[Diagram 1]} + \text{[Diagram 2]} + \text{[Diagram 3]} + \text{[Diagram 4]} + \text{[Diagram 5]}$$

Figure 4.1 Figurative description of the three-point moment in terms of the unconnected (first four diagrams, from *left to right*) and connected (last diagram) parts. Picture from (15)

realisation of a Rayleigh-Lèvy flight (i.e. a collection of random walks). Both samples have the same two-point correlation functions but remarkably different visual appearances and higher-order correlation properties. In general, even if non-Gaussianity is not imprinted into the primordial density fluctuations as in the case of the so-called primordial non-Gaussianity, it can arise from the nonlinear gravitational evolution of perturbations or the nonlinear nature of biasing between dark matter fluctuations and their tracers.

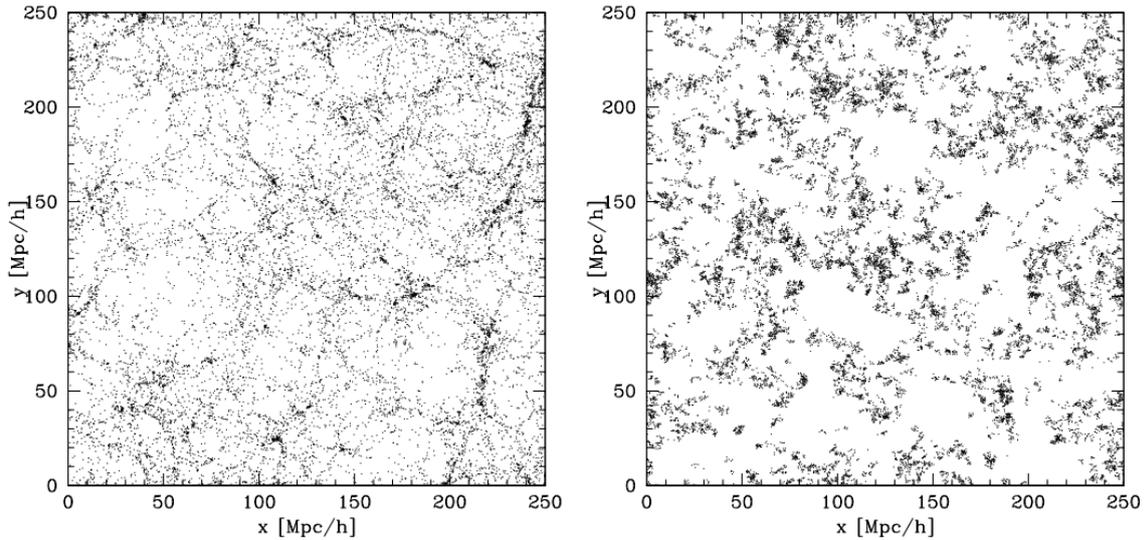


Figure 4.2 Slices of thickness $50h^{-1}\text{Mpc}$ of a mock galaxy distribution for SDSS (left) and a realisation of a Rayleigh-Lèvy flight (right). They have the same two-point correlation function, and they exhibit visible differences. Figure from (16)

Particularly interesting is the three-point correlation function, or its Fourier space counterpart, the bispectrum, because it is the higher-order statistics with the highest signal-to-noise ratio in the nonlinear regime. It turns out that it is crucial to disentangle cosmological information hidden in nonlinear evolution. It defines the probability of finding a triplet of objects in excess over a random distribution of points. As for the two-point case, we can

define the three-point correlation function for a set of discrete tracers. We get

$$\begin{aligned} \langle n(\mathbf{x}_1)n(\mathbf{x}_2)n(\mathbf{x}_3) \rangle_{\times} &= \bar{n}^3 [1 + \xi(r_{12}) + \xi(r_{23}) + \xi(r_{31}) + \overbrace{\zeta(r_{12}, r_{23}, r_{31})}^{3\text{PCF}}] \\ &+ \bar{n}^2 [1 + \xi(r_{12})] \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_2) + \text{cyc} \\ &+ \bar{n} \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_2) \delta_{\text{D}}(\mathbf{x}_2 - \mathbf{x}_3) \end{aligned} \quad (4.21)$$

where we used $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$. Let us note that by the assumption of homogeneity and isotropy, the three-point correlator is here a function of three sides only. As for the two-point case, the discrete case brings shot noise contributions. Following the same path as in the two-point case, we can apply the Wiener-Kinchin theorem. It turns out that the bispectrum $B(k_1, k_2, k_3)$, defined as

$$\langle \delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \rangle \equiv B(k_1, k_2, k_3) \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \quad (4.22)$$

is linked to the three-point correlation function $\zeta(r_{12}, r_{13}, r_{23})$ by a Fourier transform

$$B(k_1, k_2, k_3) = (2\pi)^9 \int d^3r_1 d^3r_2 d^3r_3 \zeta(r_{12}, r_{13}, r_{23}) e^{-i(\mathbf{k}_1 \cdot \mathbf{r}_1 + \mathbf{k}_2 \cdot \mathbf{r}_2 + \mathbf{k}_3 \cdot \mathbf{r}_3)}, \quad (4.23)$$

$$\begin{aligned} \zeta(r_{12}, r_{23}, r_{31}) &= \int \frac{d^3k_1 d^3k_2 d^3k_3}{(2\pi)^9} B(k_1, k_2, k_3) e^{i(\mathbf{k}_1 \cdot \mathbf{r}_{12} + \mathbf{k}_2 \cdot \mathbf{r}_{23} + \mathbf{k}_3 \cdot \mathbf{r}_{31})} \\ &\times \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \end{aligned} \quad (4.24)$$

Going to the discrete tracer of the bispectrum, we obtain

$$\begin{aligned} \langle \langle \delta n(\mathbf{k}_1) \delta n(\mathbf{k}_2) \delta n(\mathbf{k}_3) \rangle_{\times} \rangle &= (2\pi)^3 \left[B(k_1, k_2, k_3) + \frac{1}{\bar{n}} [P(k_1) + P(k_2) + P(k_3)] + \frac{1}{\bar{n}^2} \right] \\ &\times \delta_{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \end{aligned} \quad (4.25)$$

As in the two-point case, the discrete tracer of the bispectrum brings a constant term that contributes on all scales.

4.2 Modeling clustering statistics from perturbation theory

Correlation functions have been modelled using both perturbative and non-perturbative methods (115; 116; 117). The focus of this thesis is on perturbative modelling, which has been a useful tool for studying the clustering properties of celestial objects. The most common approach is known as Standard Perturbation Theory (SPT) (see (15) for a review).

However, more advanced approaches have been developed to address the limitations of SPT, such as Effective Field Theory (EFT). There is typically a distinction between modelling perturbations in dark matter and those in halos and galaxies. Research in this area is addressed by examining the relationship between the fluctuations of luminous tracers (such as galaxies) and dark matter perturbations. Furthermore, it is worth stressing that Perturbation Theory provides a framework for understanding how non-Gaussianity arises from the gravitational evolution of fluctuations. Although initial fluctuations may be Gaussian, their correlations with nonlinear evolution can lead to higher-order statistics that are not zero.

4.2.1 Modelling two- and three- point statistics in Fourier space: the matter case

Perturbative model for the matter power spectrum

The simplest and trivial application of Perturbation Theory is given by the linear power spectrum, for which we do not need to consider perturbative expansions except the leading one, i.e. when the overdensities linearly evolve (see 3.2.2). In this case, the linear power spectrum $P_L(k)$ is usually computed throughout relativistic Boltzmann codes (118; 119), despite analytical approaches having been developed (120). Generally, it can be decomposed as follows:

$$P_L(k) = k^{n_p} T^2(k) \quad (4.26)$$

where n_p is the primordial spectral index. A $n_p = 1$ value characterises the so-called Harrison-Zel'dovich scale-invariant spectrum, and $T(k)$ is the transfer function describing the evolution of the density field perturbations through decoupling.

To model the power spectrum in the nonlinear regime, it is usually assumed as common practice to deal with perturbative expansion at higher orders than the leading one. In this context, it is useful to introduce the notion of loops defined as the contribution to the N -point correlation function coming from considering higher perturbative expansions in the overdensity field in Eq. 3.38 with respect to the leading one (i.e. $\langle \delta_1^m, \dots, \delta_N^n \rangle$, with $m + \dots + n = N$). To give an idea, one-loop contributions consider $\langle \delta_1^m, \dots, \delta_N^n \rangle$, with $m + \dots + n = 2N$, two-loops contributions with $m + \dots + n = 3N$ and so on. For example, the second order in the overdensity field is the leading order in the power spectrum. The next-to-leading order accounts for four orders in the overdensity fields by the Wick theorem. To build four order overdensity fields term, we do it by correlating the linear one with a third one, twice given two different wavevectors \mathbf{k} and \mathbf{k}' , or by correlating two second-order overdensity fields.

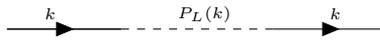
The correlation functions read as

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \langle \delta^{(1)}(\mathbf{k})\delta^{(1)}(\mathbf{k}') \rangle + 2\langle \delta^{(1)}(\mathbf{k})\delta^{(3)}(\mathbf{k}') \rangle + \langle \delta^{(2)}(\mathbf{k})\delta^{(2)}(\mathbf{k}') \rangle + \mathcal{O}(\delta^{(6)}) \quad (4.27)$$

Actually, a popular and useful prescription to compute these terms can be found by adopting the so-called *Feynman-like rules*. For the calculation of the i -th order contribution to the n -spectrum:

1. Draw all the connected diagrams with n external lines up to the desired order i in $\delta^{(1)}$:
 - (a) For each vertex V with ingoing momenta q_i and outgoing momentum k , write a Dirac delta of the form $(2\pi)^3 \delta_{\mathbb{D}}(\mathbf{k} - \sum_i \mathbf{q}_i)$ and a coupling kernel $V(\mathbf{q}_1, \dots, \mathbf{q}_m)$. In real space modelling, they are $F_m(\mathbf{q}_1, \dots, \mathbf{q}_m)$,
 - (b) Assign a term of the form $(2\pi)^3 \delta_{\mathbb{D}} P_L(q)$ - called the propagator of the diagram - to each outgoing momenta \mathbf{q} and \mathbf{q}' ,
 - (c) Integrate over loop momenta $\int d^3 \frac{q_i}{(2\pi)^3}$,
 - (d) Multiply the symmetry factor of each diagram,
 - (e) Sum over possible permutations of external lines.
2. Iterate the steps for all the possible diagrams at the i -th order contribution to the n -spectrum.

An example that clarifies how to use the Feynman rules is shown in Fig. 4.3 and Fig. 4.4. Particularly at the so-called tree-level in Fig. 4.3, applying the Feynman rules becomes trivial.



$$P_{11}$$

Figure 4.3 Tree-level power spectrum diagram. P_{11} represents the diagrammatic term corresponding to P_L

In fact, the tree-level contribution to the power spectrum corresponds to the linear propagator itself, i.e. the linear power spectrum

$$P_{\text{tree-level}}(k) \equiv P_{11}(k) = P_L(k). \quad (4.28)$$

Concerning the one-loop expansion of the power spectrum, the two diagrams involved in the expansion are represented in Fig. 4.4. Here, the two diagrams at one loop represent

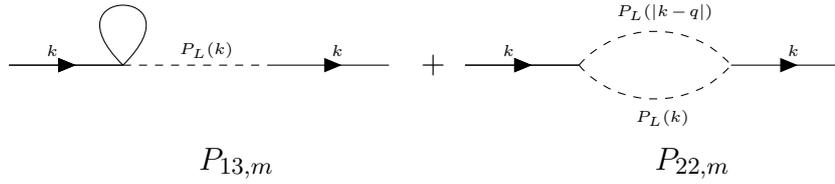


Figure 4.4 One-loop power spectrum diagrams. Left diagram refers to $P_{13,m}$, right to $P_{22,m}$

the two possible perturbative contributions to the two-point correlator in Eq. 4.27. The loop in $P_{13,m}$ must be permuted over the two possible vertexes. Hence, a factor of two will be accounted for when it is considered in the full evaluation at one-loop. By applying the mentioned rules, we can write the contributions in terms of loop integrals

$$P^{\text{one-loop}}(k) = P_L(k) + 2P_{13,m}(k) + P_{22,m}(k) \quad (4.29)$$

where - indicating through the index m the matter contribution -,

$$P_{13,m}(k) = 3P_L(k) \int \frac{d^3q}{(2\pi)^3} P_L(q) F_3^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q}), \quad (4.30)$$

$$P_{22,m}(k) = \int \frac{d^3q}{(2\pi)^3} P_L(q) P_L(|\mathbf{k} - \mathbf{q}|) |F_2^{(s)}(\mathbf{q}, \mathbf{k} - \mathbf{q})|^2, \quad (4.31)$$

where the kernels $F_2^{(s)}(\mathbf{q}, \mathbf{k} - \mathbf{q})$ and $F_3^{(s)}(\mathbf{k}, \mathbf{q}, -\mathbf{q})$ are defined in Eqs. 3.42, 3.39 and symmetrised over the arguments. These integrals must be evaluated numerically due to the numerical nature of P_L . However, the term P_{13} contains propagators that depend only on the radial part of the loop variable \mathbf{q} . This leads to further simplification and, in general, reduction to a one-dimensional integration by integrating over the angular dimension. Contrary, the $P_{22,m}$ term contains propagators that also depend on the one angle of the loop variable \mathbf{q} , meaning it can be reduced maximum to a two-dimensional integration.

The divergence properties of loop integrals depend on the power-law behaviour of the linear power spectrum, which is summarised in Tab 4.1. Despite $P_{13,m}$, $P_{22,m}$ being

	UV-divergent	IR-divergent
$P_{13,m}$	$n_p > 1$	$n_p < -1$
$P_{22,m}$	$n_p > 1/2$	$n_p < -1$

Table 4.1 Divergence conditions for one-loop terms $P_{13,m}$ and $P_{22,m}$

convergent in a power-law cosmology respectively in the case $n = 1$, $-1 \leq n \leq 1/2$, their

sum, the total contribution $P^{1\text{-loop}}$, is well defined in the range $-3 \leq n_p \leq -1$. The contribution of the one-loop terms, as represented in Fig. 4.5, is sizeable in the large scale range. In general, SPT (see (15) for a review) of clustering statistics has proved, so far, an effective way of accessing nonlinear scales in Fourier space up to wave number as large as $k_{\text{max}} \sim 0.15 h\text{Mpc}^{-1}$ (60). Hence, the widespread use of perturbative expansion techniques to investigate the clustering properties of the matter in the universe (55; 56; 57; 58; 59; 60).

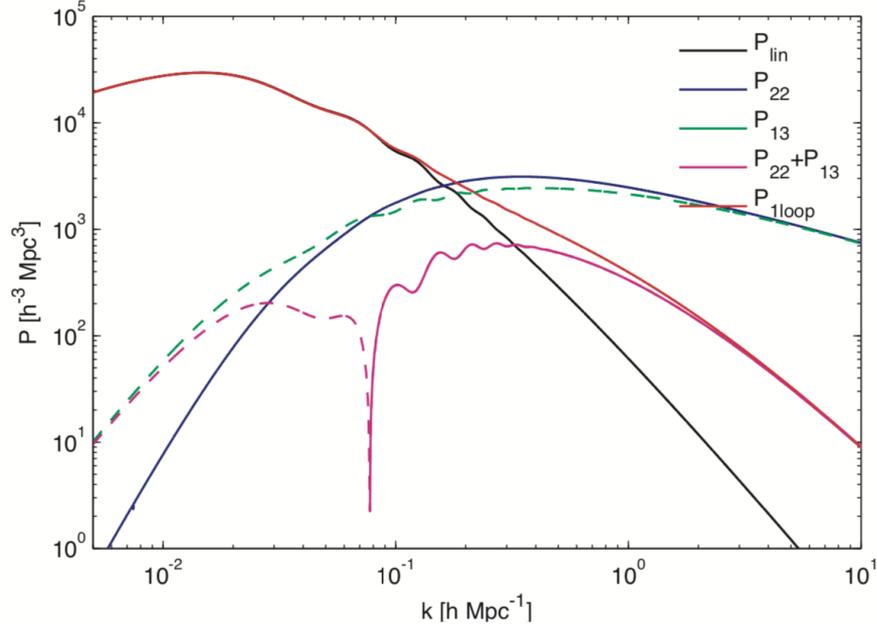


Figure 4.5 Power spectrum at one-loop and single contributions, assuming ΛCDM cosmology from Planck (3).

Furthermore, to improve the accuracy of the model at the scale of the baryonic acoustic oscillations (BAO) feature, it is a common practice to perform the so-called infrared (IR) resummation (68; 121; 122). This is a crucial step to correctly account for higher perturbative order contributions, as best shown by the case of the two-point correlation function in configuration space. In that case, neglecting resummation would generate a spurious double peak around the BAO scale, making it impossible to use it as a standard ruler for precise observational tests. We also include the IR resummation in our models to avoid generating similar spurious features in the 3PCF. To include the IR resummation, we first decompose the linear power spectrum into a smooth and wiggly component so that (123)

$$P_L(k) = P_{\text{nw}}(k) + P_{\text{w}}(k). \quad (4.32)$$

The effect is splitting the oscillating contribution from BAO imprinting into the power spectrum from the smoothed part from the remaining non-baryonic physics ruling the shape

of the power spectrum. Then, we damp the wiggling part to recover the infrared tree-level power spectrum (122)

$$P_L^{\text{IR}}(k) = P_{\text{nw}}(k) + e^{-k^2 \Sigma^2} P_w(k), \quad (4.33)$$

where the damping factor is given by the relative displacement two-point function in the Zel'dovich approximation at the BAO scale (121)

$$\Sigma^2 = \int_0^{k_S} \frac{d^3 q}{(2\pi)^3} \frac{P_{\text{nw}}(q)}{3q^2} \left[1 - j_0\left(\frac{q_0}{k_{\text{BAO}}}\right) + 2j_2\left(\frac{q_0}{k_{\text{BAO}}}\right) \right], \quad (4.34)$$

where $j_n(x)$ are spherical Bessel functions and $k_{\text{BAO}} = \pi/\ell_{\text{BAO}}$ with $\ell_{\text{BAO}} = 110 \text{Mpc}/h$. The k_S cutoff is commonly fixed to $k_S = 0.2h/\text{Mpc}$ (121; 124). We can finally recall the infrared resummed power spectrum as follows

$$\begin{aligned} P_{\text{LO},m}(k) &= P_L^{\text{IR}}(k), \\ P_{\text{NLO},m}(k) &= P_{\text{LO},m}(k) + P_{13,m}^{\text{IR}}(k) + P_{22,m}^{\text{IR}}(k) \end{aligned}$$

where we have defined the leading and the next-to-leading order for the matter power spectrum.

Perturbative model for the matter bispectrum

Moving to bispectrum statistics, it is possible to draw the associated diagram in Fig. 4.6 at the tree-level order. By applying the Feynman rules, the corresponding model reads as

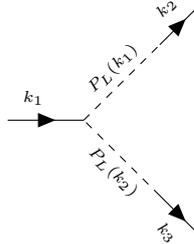


Figure 4.6 Tree-level bispectrum diagram. $B_{211,m}$ represents the diagrammatic term corresponding to $B_{\text{tree-level},m}$

$$B_{\text{tree-level},m}(k_1, k_2, k_3) \equiv B_{211,m}(k_1, k_2, k_3) = 2F_2(\mathbf{k}_1, \mathbf{k}_2) P_L(k_1) P_L(k_2) + \text{cyc.} \quad (4.35)$$

where “cyc” refers to cyclic permutations over the external momenta k_1, k_2, k_3 , and the factor of two derives from the symmetry factor of the diagram. The resulting bispectrum scales as P_L^2 . In Fig. 4.7, I report the four diagrams associated with one-loop perturbative expansions. As in the previous case, it is possible to compute the associated one-loop bispectrum as

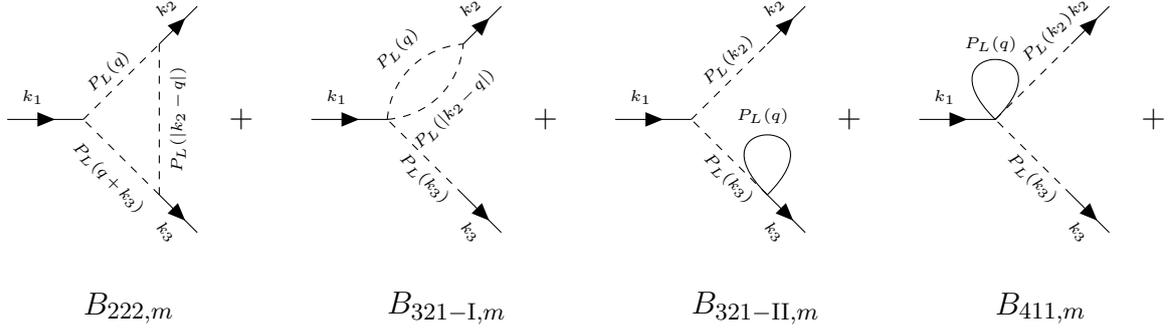


Figure 4.7 One-loop bispectrum diagrams. From right to left: $B_{222,m}$, $B_{321-I,m}$, $B_{321-II,m}$, $B_{411,m}$

follows

$$B_m^{\text{one-loop}} = B_{\text{tree-level},m} + B_{222,m} + B_{321-I,m} + B_{321-II,m} + B_{411,m}, \quad (4.36)$$

where

$$B_{222,m} = 8 \int d^3q F_2(-\mathbf{q}, \mathbf{k}_3 + \mathbf{q}) F_2(\mathbf{k}_3 + \mathbf{q}, \mathbf{k}_2 - \mathbf{q}) F_2(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) \\ \times P_L(q) P_L(|\mathbf{k}_2 - \mathbf{q}|) P_L(|\mathbf{k}_3 + \mathbf{q}|), \quad (4.37)$$

$$B_{321-I,m} = 6 P_L(k_3) \int d^3q F_3(-\mathbf{q}, -\mathbf{k}_2 + \mathbf{q}, -\mathbf{k}_3) F_2(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) \\ \times P_L(|\mathbf{k}_2 - \mathbf{q}|) P_L(q) + \text{all perms}, \quad (4.38)$$

$$B_{321-II,m} = 6 P_L(k_2) P_L(k_3) F_2(\mathbf{k}_2, \mathbf{k}_3) \int d^3q F_3(\mathbf{k}_3, \mathbf{q}, -\mathbf{q}) \\ \times P_L(q) + \text{all perms}, \quad (4.39)$$

$$B_{411,m} = 12 P_L(k_2) P_L(k_3) \int d^3q F_4(\mathbf{q}, -\mathbf{q}, -\mathbf{k}_2, -\mathbf{k}_3) \\ \times P_L(q) + \text{cyc}. \quad (4.40)$$

where I denoted with “all perms” and “cyc”, respectively, permutation over all possible exchanges and the already mentioned cyclic permutations. Similar to the power spectrum

case, the propagators involved in the different diagrams are functions of the radial and, in some cases, the angular part of the loop variables. In $B_{411,m}$ and $B_{321-\text{II},m}$ diagram, they depend only on the radial so that these integrals can be reduced to one dimension. For the $B_{321-\text{I},m}$, the propagator depends only on one angle so that it can be reduced to a two-dimensional integral. Finally, $B_{222,m}$ must be evaluated in its three dimensions due to the dependence of the propagators on all the degrees of freedom of the loop variable. At the one-loop level, the bispectrum comprises two distinct components: the tree-level contributions, which involve the combination of two linear power spectra (represented by P_L), and the one-loop contributions, which are comprised of three linear power spectra (represented by P_L). This serves as a means of breaking the degeneracies between the parameters of bias and the amplitude of the power spectra during the fitting process. As for the power spectrum case, to reach sub-per cent accuracy on the BAO scale, an IR-resummation step is also added to the bispectrum modelling. It simply consists of using the IR-resummed power spectrum as input power spectrum in both the tree-level and one-loop contributions for the bispectrum. This allows using the leading and next-to-leading orders, connected with tree-level and one-loop contributions by the infrared resummation scheme

$$B_{\text{LO}}(k_1, k_2, k_3) = B_{\text{tree}}^{\text{IR}}(k_1, k_2, k_3), \quad (4.41)$$

$$B_{\text{NLO}}(k_1, k_2, k_3) = B_{\text{tree}}^{\text{IR}}(k_1, k_2, k_3) + B_{\text{one-loop}}^{\text{IR}}(k_1, k_2, k_3). \quad (4.42)$$

4.2.2 The effective field theory approach for modelling for the matter power spectrum and bispectrum: the matter case

Despite SPT providing us with a powerful description of nonlinear scales, it exhibits some shortcomings, such as the fact that on small scales, the density is large and the power series does not converge, and shell crossing leads to multi-streaming and departures from the assumption of pressureless fluid. In addition to the mentioned fact that the behaviour of the linear power spectrum determines that for certain power laws potential divergencies. Those shortcomings represent the main basis over which the attempt to build an effective field theory (EFT) description of small scales has been provided. Indeed, a series of works (65; 66; 67; 68) addressed an alternative effective description of long-wavelength modes where SPT usually fails. The EFT procedure uses an effective stress tensor expressed in terms of all operators of long-wave-length density and velocity fields. Taylor expanding the effective stress tensor gives us an infinite series of unconstrained parameters associated with each perturbative expansion. These parameters can be treated as coupling constants in the Wilson approach

to renormalisation (125). EFT coupling coefficients can be seen as counterterms that can be used to cancel the dependence of SPT on the UV scale and to model a non-ideal stress tensor. The remaining cutoff-independent part of the counterterms quantifies the impact of physics that cannot be described perturbatively by introducing effective interaction on long-wavelength modes. For the power spectrum, the EFT procedure adds a counter-term at the one-loop level, depending on the parameter c_0 , that should be fitted in a cosmological analysis

$$P_{\text{EFT}}^{1\text{-loop}}(k) = P_L(k) + 2P_{13,m}(k) + P_{22,m}(k) - 2c_0k^2P_L(k), \quad (4.43)$$

being $P_{\text{ctr}}(k) = -2c_0k^2P_L(k)$ the counter-term.

Concerning modelling bispectrum in the EFT framework at the next-to-leading order, it is possible to express this contribution as the sum of four counterterms (60; 126; 127):

$$B_{\text{EFT}}(k_1, k_2, k_3) = B_{\text{NLO}}(k_1, k_2, k_3) + B_{c_0} + B_{c_1} + B_{c_2} + B_{c_3}, \quad (4.44)$$

where

$$B_{c_0} = c_0(z) [D(z)]^4 \left[2P_L(k_1)P_L(k_2)\tilde{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}) + 2 \text{ perms} \right. \\ \left. - 2k_1^2P_L(k_1)P_L(k_2)F_2(\mathbf{k}_1, \mathbf{k}) + 5 \text{ perms} \right], \quad (4.45)$$

$$B_{c_1} = -2c_1(z)[D(z)]^4k_1^2P_L(k_2)P_L(k_3) + 2 \text{ perms}, \quad (4.46)$$

$$B_{c_2} = -2c_2(z)[D(z)]^4k_1^2\frac{(\mathbf{k}_2 \cdot \mathbf{k}_3)^2}{k_2^2k_3^2}P_L(k_2)P_L(k_3) + 2 \text{ perms}, \quad (4.47)$$

$$B_{c_3} = -2c_3(z)[D(z)]^4(\mathbf{k}_2 \cdot \mathbf{k}_3)\left[\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_2^2} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_3}{k_3^2}\right]P_L(k_2)P_L(k_3) + 2 \text{ perms}, \quad (4.48)$$

where $\tilde{F}_2^{(s)}$ is defined as

$$\tilde{F}_2^{(s)}(\mathbf{k}_1, \mathbf{k}_2) = \frac{1}{(1+\zeta)(7+2\zeta)} \left[\left(5 + \frac{113\zeta}{14} + \frac{17\zeta^2}{7}\right)(k_1^2 + k_2^2) \right. \\ + \left(7 + \frac{148\zeta}{7} + \frac{48\zeta^2}{7}\right)(\mathbf{k}_1 \cdot \mathbf{k}_2) \\ + \left(2 + \frac{59\zeta}{7} + \frac{18\zeta^2}{7}\right)\left(\frac{1}{k_1^2} + \frac{1}{k_2^2}\right)(\mathbf{k}_1 \cdot \mathbf{k}_2)^2 \\ + \left(\frac{7}{2} + \frac{9\zeta}{2} + \zeta^2\right)\left(\frac{k_1^2}{k_2^2} + \frac{k_2^2}{k_1^2}\right)(\mathbf{k}_1 \cdot \mathbf{k}_2) \\ \left. + \left(\frac{20\zeta}{7} + \frac{8\zeta^2}{7}\right)\frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^3}{k_1^2k_2^2} \right]. \quad (4.49)$$

Following (67) $\zeta = 3.1$, but in (68) there is no evidence of different predictions from adopting $\zeta = 2$. In this thesis, I use the fiducial value of $\zeta = 3.1$.

4.2.3 Modelling two- and three- point statistics in Fourier space: the galaxy case

Perturbative model for the galaxy power spectrum and bispectrum

The perturbative modelling of two and three -point statistics for the galaxy distribution requires a more sophisticated treatment compared to the matter case examined in the previous section. Here, I explore the case within the framework of SPT, subsequently applied in this Thesis, not focusing on the so-called stress-tensor contributions, which are analogous to the EFT modelling in the galaxy case. For details, see (126). Concerning the modelling of luminous tracers as galaxies are, model predictions of the correlation functions usually require a relation between galaxy perturbation and matter fluctuations. Usually, it is assumed to be described as a Taylor expansion, as presented in Eq. 3.69, that here is reported

$$\delta_g(\mathbf{x}) = \sum_{n=1} \frac{b_n}{n!} \delta(\mathbf{x})^n = \sum_{n=1} \frac{1}{n!} \left(\frac{\partial^n \delta_g}{\partial \delta^n} \Big|_0 \right) \delta(\mathbf{x})^n = \bar{b}_0 + \bar{b}_1 \delta + \frac{\bar{b}_2}{2} \delta^2 + \frac{\bar{b}_3}{3!} \delta^3 + \text{h.o.} \quad (4.50)$$

where I identified the bias parameters as the derivative of the overdensity computed when it is equal to zero. This expansion, however, suffers from the fact that the bias parameters are *not observable*. As an example, using Eq. 4.14 along with the expansion in Eq. 4.50 limiting to the third order and reminding that Fourier space transforming for galaxy overdensities works as

$$\delta_g(\mathbf{x}) = \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} \delta_g(\mathbf{k}), \quad (4.51)$$

where $\int_{\mathbf{k}} = \int d^3 k$, so that the power spectrum and the bispectrum now read as

$$\begin{aligned} P_g(k) &= \bar{b}_1^2 P_L(k) + (\bar{b}_1^2 \bar{b}_3^2 \sigma^2) P_L(k) + \frac{\bar{b}_2^2}{2} \int_{\mathbf{q}} P_L(|\mathbf{k} - \mathbf{q}|) P_L(q) \\ &\quad + \text{higher order terms} \quad (4.52) \\ B_g(k_1, k_2, k_3) &= \bar{b}_1^2 \bar{b}_2 P_L(k_1) P_L(k_2) + \left[\bar{b}_1^2 \bar{b}_2^2 \bar{b}_3^2 \sigma^2 + \frac{1}{2} \bar{b}_1^2 \bar{b}_4 \sigma^2 \right] P_L(k_1) P_L(k_2) \\ &\quad + \frac{1}{2} \bar{b}_1^2 \bar{b}_2^2 \bar{b}_3^2 P_L(k_1) \int_{\mathbf{q}} P_L(\mathbf{k}_2 - \mathbf{q}) P_L(q) \\ &\quad + \bar{b}_2^3 \int_{\mathbf{q}} P_L(\mathbf{k}_1 - \mathbf{q}) P_L(\mathbf{k}_2 + \mathbf{q}) P_L(q) \\ &\quad + \text{cyc} \end{aligned}$$

$$+ \text{higher order terms} \quad (4.53)$$

where $\sigma^2 \equiv \langle \delta^2(\mathbf{x}) \rangle$, a quantity highly sensitive to nonlinear regime, meaning that power spectrum and bispectrum models are affected by these terms in the large scale regime. In these expressions, higher-order bias parameters enter the leading order terms at different higher perturbative orders. This is because, as defined in Eq. 4.50, the linear bias is *not observable*, while the *observable* linear bias should be defined as the coefficient in front of P_L . Redefining the linear and the quadratic biases coefficients as $b_1 = \bar{b}_1 + \bar{b}_3\sigma^2/2$, $b_2 = \bar{b}_2 + \bar{b}_4\sigma^2/2$, we get

$$\begin{aligned} P_g(k) &= b_1^2 P_L(k) + \frac{\bar{b}_2^2}{2} \int_{\mathbf{q}} P_L(|\mathbf{k} - \mathbf{q}|) P_L(q) \\ &\quad + \text{higher order terms} \\ B_g(k_1, k_2, k_3) &= b_1^2 b_2 P_L(k_1) P_L(k_2) + \\ &\quad + \frac{1}{2} \bar{b}_1^2 \bar{b}_2^2 \bar{b}_3^2 P_L(k_1) \int_{\mathbf{q}} P_L(\mathbf{k}_2 - \mathbf{q}) P_L(q) \\ &\quad + \bar{b}_2^3 \int_{\mathbf{q}} P_L(\mathbf{k}_1 - \mathbf{q}) P_L(\mathbf{k}_2 + \mathbf{q}) P_L(q) \\ &\quad + \text{cyc} \\ &\quad + \text{higher order terms} \end{aligned} \quad (4.54)$$

The problem with this redefinition is that it depends on the perturbative order we consider. In fact, as more bias loops are included, the expression for b_1 keeps changing but, in principle, remains the *observed* linear bias, defined as the coefficient in front of P_L . In general, redefinitions like this lead to possible inconsistencies between the same bias parameters involved in the power spectrum, bispectrum and higher-order statistics. The problem has been considered in (126; 128), which proposed a more general basis to include higher-order local and non-local contributions. On this basis, the galaxy overdensity reads as

$$\begin{aligned} \delta_g(\mathbf{x}) &= \bar{b}_1 \delta(\mathbf{x}) \\ &\quad + \frac{\bar{b}_2}{2} \delta^2(\mathbf{x}) + \bar{\gamma}_2 \mathcal{G}_2(\Phi_v | \mathbf{x}) + \\ &\quad + \frac{\bar{b}_3}{6} \delta^3(\mathbf{x}) + \bar{\gamma}_2^\times \mathcal{G}_2(\Phi_v | \mathbf{x}) \delta(\mathbf{x}) + \bar{\gamma}_3 \mathcal{G}_3(\Phi_v | \mathbf{x}) + \bar{\gamma}_{21} \mathcal{G}_2(\phi_2, \phi_1 | \mathbf{x}) \\ &\quad + \bar{\gamma}_{21}^\times \mathcal{G}_2(\phi_2, \phi_1) \delta(\mathbf{x}) + \bar{\gamma}_{211} \mathcal{G}_3(\phi_2, \phi_1, \phi_1) + \bar{\gamma}_{22} \mathcal{G}_2(\phi_2, \phi_2) + \bar{\gamma}_{31} \mathcal{G}_3(\phi_3, \phi_1). \end{aligned} \quad (4.55)$$

Every term in this expansion combines a different operator with different associated bias parameters. Also, some terms involve the so-called *Galileian* operators, already encountered

in Sec.3.3.2 and defined in Eq. 3.54 and Eq. 3.71. The potentials ϕ_n are the LPT potential introduced in Sec. 3.2.4. In this representation of bias expansions, there are two main groups of operators: local and nonlocal. The first ones are built up throughout delta operators, and the two Galileian operators $\mathcal{G}_2(\Phi_v|\mathbf{x})$ and $\mathcal{G}_3(\Phi_v|\mathbf{x})$. On the other hand, all the other operators involved, such as $\mathcal{G}_3(\phi_2, \phi_1, \phi_1)$ and $\mathcal{G}_2(\phi_2, \phi_1)$, belong to the nonlocal group. They reflect the fact that gravity acts over long distances, and it is evident in the LPT potentials in all terms beyond the Zel'dovich approximation.

Still, in the basis presented in Eq. 4.55, the bias parameters depend on the expansion order. In order to guarantee that the expansion coefficients are observable quantities, it is possible to expand the galaxy density contrast in Fourier space in terms of Wiener-Hermite functionals \mathcal{H}_n (126; 128)

$$\delta_g(\mathbf{k}) = [\Gamma_g^{(1)} \otimes \mathcal{H}_1](\mathbf{k}) + \frac{1}{2!} [\Gamma_g^{(2)} \otimes \mathcal{H}_2](\mathbf{k}) + \frac{1}{3!} [\Gamma_g^{(3)} \otimes \mathcal{H}_3](\mathbf{k}) + \dots \quad (4.56)$$

where

$$[\Gamma_g^{(n)} \otimes \mathcal{H}_n](\mathbf{k}) \equiv (2\pi)^3 \int d^3k_1, \dots, d^3k_n \delta_D(\mathbf{k} - \mathbf{k}_{1\dots n}) \Gamma_g^{(n)}(\mathbf{k}_1, \dots, \mathbf{k}_n) \mathcal{H}_n(\mathbf{k}_1, \dots, \mathbf{k}_n) \quad (4.57)$$

with δ_D being the Dirac delta, $\mathbf{k}_{1,\dots,n} \equiv \mathbf{k}_1 + \dots + \mathbf{k}_n$, and

$$\mathcal{H}_1(\mathbf{k}) = \delta^*(\mathbf{k}), \quad (4.58)$$

$$\mathcal{H}_1(\mathbf{k}_1, \mathbf{k}_2) = \delta^*(\mathbf{k}_1)\delta^*(\mathbf{k}_2) - \langle \delta^*(\mathbf{k}_1)\delta^*(\mathbf{k}_2) \rangle, \quad (4.59)$$

$$\mathcal{H}_1(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \delta^*(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\delta^*(\mathbf{k}_3) - [\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle \delta_L(\mathbf{k}_3) + \text{cyc}] \quad (4.60)$$

In this expansion, usually known as a *renormalisation basis*, the functions $\Gamma_g^{(n)}$ take the role of scale-dependent biases, and they are called galaxy *multipoint propagator* in renormalized perturbation theories. In this context, the bias parameters can be seen as the ensemble-averaged derivatives of δ_g with respect to the linear perturbation δ_L

$$b_n \equiv \left\langle \frac{\partial^n \delta_g(\mathbf{k})}{\partial \delta_L(\mathbf{k}_1) \dots \partial \delta_L(\mathbf{k}_n)} \right\rangle \equiv (2\pi)^3 \Gamma_g^{(n)}(\mathbf{k}_1, \dots, \mathbf{k}_n) \delta_D(\mathbf{k} - \mathbf{k}_{1,\dots,n}) \quad (4.61)$$

The multipoint propagators are directly related to cross-correlations between galaxy and dark matter fluctuations, providing they are observable. For instance, on the large-scale limit, $\Gamma_g(k)$ plays the role of the linear bias as measured in real or simulated datasets. The scale dependence of multipoint propagators is parameterised by the same operators introduced in the basis represented in Eq. 4.50. It has been demonstrated in (126) that loop corrections to

multipoint propagators are functions of only nonlinear operators. Up to the first loop order, the first three Lagrangian multipoint propagators read as

$$\Gamma_{g,\mathcal{L}}^{(1)}(\mathbf{k}) = \overbrace{b_{1,\mathcal{L}}}^{\text{tree-level}} + \overbrace{2\gamma_{21,\mathcal{L}} \int d^3q K(\mathbf{k}-\mathbf{q}, \mathbf{q}) K(\mathbf{k}, \mathbf{q}) P_L(q)}^{\text{one-loop}}, \quad (4.62)$$

$$\begin{aligned} \Gamma_{g,\mathcal{L}}^{(2)}(\mathbf{k}_1, \mathbf{k}_2) &= \overbrace{b_{2,\mathcal{L}} + 2\gamma_{2,\mathcal{L}} K(\mathbf{k}_1, \mathbf{k}_2)}^{\text{tree-level}} \\ &\quad + \overbrace{12 \int d^3 [\gamma_{21,\mathcal{L}}^\times \mathcal{K}_{\delta\mathcal{G}_2(\phi_2, \phi_1)}^{(4,F)} + \gamma_{211,\mathcal{L}}^\times \mathcal{K}_{\mathcal{G}_2(\phi_2, \phi_1, \phi_1)}^{(4)} \\ &\quad + \gamma_{22,\mathcal{L}}^\times \mathcal{K}_{\mathcal{G}_2(\phi_2, \phi_2)}^{(4)} + \gamma_{31,\mathcal{L}}^\times \mathcal{K}_{\mathcal{G}_3(\phi_3, \phi_1)}^{(4)}] P_L(q)}^{\text{one-loop}}, \end{aligned} \quad (4.63)$$

$$\begin{aligned} \Gamma_{g,\mathcal{L}}^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= b_{3,\mathcal{L}} + 2\gamma_{2,\mathcal{L}}^\times [K(\mathbf{k}_1, \mathbf{k}_2) + \text{cyc}] \\ &\quad + \overbrace{2\gamma_{21,\mathcal{L}} [K(\mathbf{k}_1, \mathbf{k}_2) K(\mathbf{k}_{12}, \mathbf{k}_3) + \text{cyc}] 6\gamma_{3,\mathcal{L}} L(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}^{\text{tree-level}} \end{aligned} \quad (4.64)$$

where we have highlighted the tree-level and one-loop structures and the kernels involved are defined in Appendix A.2. The transition from the Lagrangian to the Eulerian picture, i.e. to the late-time galaxy fluctuations, induces corrections due to the time evolution. Then, new contributions enter as an additional loop correction, and the constraint on nonlocal operators involved in loop corrections breaks down at this moment. It has been demonstrated in (126) that additional contributions take the form

$$\Delta\Gamma_g^{(1)}|_{\text{one-loop}} = 3b_1 \int d^3q F_3(\mathbf{k}, \mathbf{q}, -\mathbf{q}) P_L(q) + 4\gamma_2 \int d^3q K(\mathbf{k}-\mathbf{q}) G_2(\mathbf{k}, -\mathbf{q}) P_L(q), \quad (4.65)$$

$$\Delta\Gamma_g^{(2)}|_{\text{tree-level}} = 2b_1 F_2(\mathbf{k}_1, \mathbf{k}_2), \quad (4.66)$$

$$\Delta\Gamma_g^{(3)}|_{\text{tree-level}} = 6b_1 F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + 2 \left[b_2 F_2(\mathbf{k}_1, \mathbf{k}_2 + 2\gamma_2 K(\mathbf{k}_1 + \mathbf{k}_2, \mathbf{k}_3) G_2(\mathbf{k}_1, \mathbf{k}_2) + \text{cyc} \right]. \quad (4.67)$$

While the first-order propagator is affected by nonlinear evolution at the one-loop correction order, the second and third-order propagators are already affected at the tree level. While multipoint propagators are presented as a function of bias parameters in the Lagrangian basis, subscripted through the suffix \mathcal{L} , the parameters themselves can be grouped in the Eulerian basis, giving place to a set of coevolution relations between bias parameters (126)

$$\gamma_2 = -\frac{2}{7}(b_1 - 1) + \gamma_{2,\mathcal{L}}, \quad (4.68)$$

$$\gamma_2^\times = -\frac{2}{7}b_2 + \gamma_{2,\mathcal{L}}, \quad (4.69)$$

$$\gamma_3 = \frac{1}{9}(b_1 - 1) - \gamma_2 + \gamma_{3,\mathcal{L}}, \quad (4.70)$$

$$\gamma_{21} = \frac{2}{21}(b_1 - 1) + \frac{6}{7}\gamma_2 + \gamma_{21,\mathcal{L}}, \quad (4.71)$$

$$\gamma_{21}^\times = \frac{2}{21}b_2 + \frac{6}{7}\gamma_2^\times + \gamma_{21,\mathcal{L}}, \quad (4.72)$$

$$\gamma_{211} = \frac{5}{77}(b_1 - 1) + \frac{15}{14}\gamma_2 - \frac{9}{7}\gamma_3 + \gamma_{21} + \gamma_{211,\mathcal{L}}, \quad (4.73)$$

$$\gamma_{22} = \frac{6}{539}(b_1 - 1) - \frac{9}{49}\gamma_2 + \gamma_{21,\mathcal{L}}, \quad (4.74)$$

$$\gamma_{31} = -\frac{4}{11}(b_1 - 1) - 6\gamma_2 + \gamma_{31,\mathcal{L}}, \quad (4.75)$$

where $b_1 = 1 + b_{1,\mathcal{L}}$ and $b_2 = b_{2,\mathcal{L}}$. Particularly, bias operators that are absent in the Lagrangian picture arise by nonlinear evolution contributions. A consequence of the peak-background split approach, briefly described in Sec.3.3.2, led to the bias relation $b_2(b_1)$ presented in Eq. 3.72 by fitting simulated dataset. Moreover, provided us with another relation between the bias parameters b_3 and the linear bias b_1 , reading as

$$b_3(b_1) = -1.028 + 7.646b_1 - 6.227b_1^2 + 0.912b_1^3. \quad (4.76)$$

This relation can be proven useful when used to reduce the parameter space dimensionality at higher perturbative orders.

Power spectrum and bispectrum

The multipoint propagator approach provides us with a very effective method when modelling correlation functions in the nonlinear regime and, in general, when investigating the bias relation between dark matter fluctuations and their luminous tracers. In particular, addressing the renormalisation of various operators and bias parameters involved in the expansion basis, this theoretical framework provides us with a diagrammatic approach to compute renormalised correlation functions. To better explain how standard Feynmann diagrams extensively used in SPT are used in this context, I reported the diagram expansion in terms of the general multipoint propagator. In Fig. 4.8, the multipoint propagator for n -points is represented diagrammatically, showing it encodes various loop orders at the corresponding vertex. In general, any loop correction can be portrayed as a diagram with a fixed number of incoming momenta and various numbers of internal lines. In this case, the number of internal lines is associated with the number of loops at the vertex. Particularly, I stress the bias expansion is here written in terms of the sum over all reducible diagrams. The

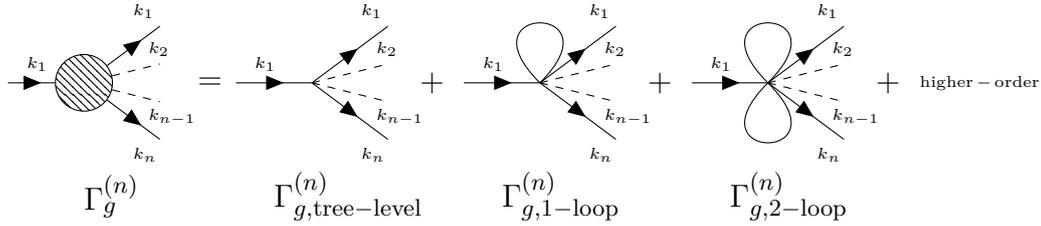


Figure 4.8 Diagrammatic representation for the n -th multipoint propagator. The vertexes correspond to the n incoming momenta. The blob vertex corresponds to the diagrammatic representation of a vertex in a renormalised perturbation theory. The represented diagrams show the multipoint propagator involved in a generic correlation function up to two-loop plus higher order corrections.

multipoint propagator diagram representation is particularly powerful because of its ability to describe easily and uniquely any complicated loop diagram. More specifically, the reducible subdiagrams are absorbed into the multipoint propagators, while the subset of irreducible diagrams is recovered by connecting multipoint propagators between themselves. The multipoint propagator approach provides us with a powerful tool to recover loop expansions of correlation functions in a renormalised perturbation theory. In order to explain how multipoint propagators are involved in the computation of correlation functions, I report the case for the power spectrum at one-loop in Fig. 4.9. The diagram for the power spectrum

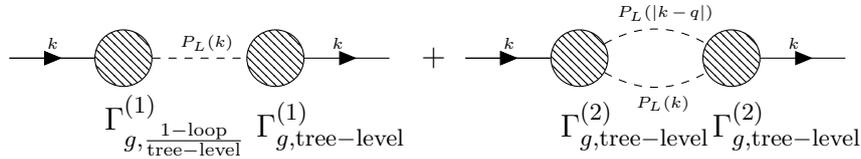


Figure 4.9 Power spectrum at one-loop in the bias loops multipoint propagator approach. The possible arising terms at one-loop are represented by the combination of 1-loop and tree-level *one*-point propagators (left diagram) connected by one internal line, and two tree-level *two*-point propagators connected by two internal lines (right diagram). The tree-level power spectrum is recovered by two tree-level *one*-point propagators (left diagram) connected by one internal line.

at one-loop order is recovered by glueing together two of the multipoint propagator shown in Fig. 4.8, fixing n , i.e. the number of external lines, to $n = 1$, where each combination of the incoming lines produces an internal line, i.e. the linear power spectrum. The shaded area represents the vertexes which are associated with the multipoint propagator, including loop corrections at the desired order. With this prescription, it is possible to recover the expressions for the one-loop power spectrum

$$P_{gg}(k) = \left[\Gamma_{g, \text{one-loop}}^{(1)}(k) \right]^2 P_L(k) + \frac{1}{2} \int d^3q \left[\Gamma_{g, \text{tree-level}}^{(2)}(\mathbf{k} - \mathbf{q}, \mathbf{q}) \right]^2 P_L(|\mathbf{k} - \mathbf{q}|) P_L(q). \quad (4.77)$$

where $\Gamma_g^{(1)}$ and $\Gamma_g^{(2)}$ are evaluated, respectively, at one-loop and tree-level order.

With the same approach, it is possible to get the one-loop bias expansion for the bispectrum. In Fig. 4.10, it is reported the diagrams involved in the mentioned expansion. As for

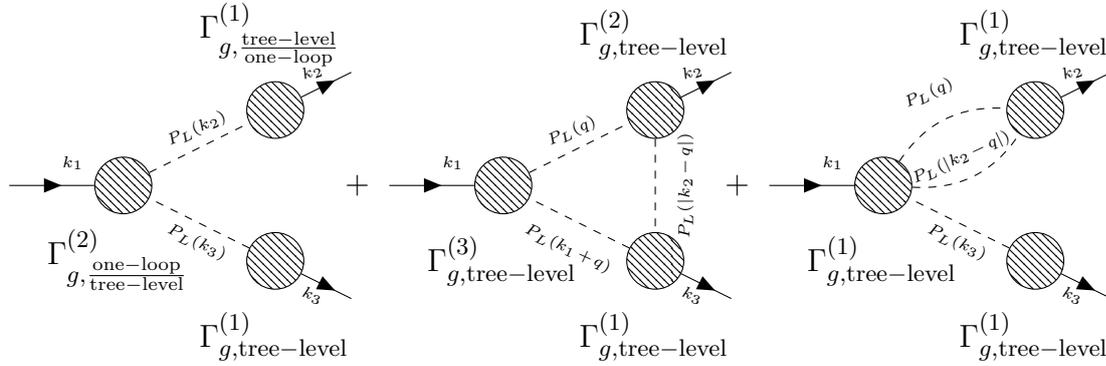


Figure 4.10 Bispectrum at one-loop in the bias loops multipoint propagator approach. The possible arising terms at one-loop are represented by the combination of three *one*-point propagators (left diagram, two at three-level, one at one-loop) connected by two internal lines, three tree-level *two*-point propagators connected by three internal lines (middle diagram) and the same but connecting tree-level one *one*-point, *two*-point, *three*-point propagator (right diagram). The three-level bispectrum is recovered by three *one*-point propagators at three-level connected by two internal lines (left diagram).

the power spectrum case, the one-loop bispectrum in the multipoint propagator approach is obtained by combining different multipoint propagators at different orders connecting them with two or three internal lines. The resulting expression for the bispectrum read as

$$\begin{aligned}
B_{ggg}(k_1, k_2, k_3) = & \Gamma_g^{(2)}(\mathbf{k}_1, \mathbf{k}_2) \Gamma_g^{(1)}(k_1) \Gamma_g^{(1)}(k_2) P_L(k_1) P_L(k_2) + \text{cyc} \\
& + \left[\int d^3 q \Gamma_g^{(2)}(\mathbf{k}_1 - \mathbf{q}, \mathbf{q}) \Gamma_g^{(2)}(\mathbf{k}_1 + \mathbf{q}, -\mathbf{q}) \Gamma_g^{(2)}(\mathbf{k}_1 + \mathbf{q}, \mathbf{k}_2 + \mathbf{q}) \right. \\
& \quad \times P_L(|\mathbf{k}_1 - \mathbf{q}|) P_L(|\mathbf{k}_1 + \mathbf{q}|) P_L(|\mathbf{k}_2 + \mathbf{q}|) \\
& + \frac{1}{2} \int d^3 q \Gamma_g^{(3)}(\mathbf{k}_3, \mathbf{k}_2 - \mathbf{q}, \mathbf{q}) \Gamma_g^{(2)}(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) \Gamma_g^{(1)}(k_3) \\
& \quad \left. \times P_L(|\mathbf{k}_2 - \mathbf{q}|) P_L(q) P_L(k_3) + \text{cyc} \right] \quad (4.78)
\end{aligned}$$

Let me stress, as explained in (126), that the bias loops corrections do not only affect small scales, but they have an impact on large scales. This is because a subset of the bias loops does not vanish in the large-scale limit. From the power spectrum, the non-vanishing contribution

comes from the following term

$$\lim_{k \rightarrow 0} P_g(k) = \frac{1}{2} \int d^3q \left[\Gamma_g^{(2)}(-\mathbf{q}, \mathbf{q}) \right]^2 P_L^2(q) = \frac{b_2^2}{2} \int d^3q P_L(q)^2. \quad (4.79)$$

Concerning the bispectrum, the non-vanishing contributions from the large-scale limit come from configurations in which all three triangles sides tend to zero, such as

$$\lim_{k_1, k_2 \rightarrow 0} B_g(k_1, k_2, k_3) = \int d^3q \left[\Gamma_g^{(2)} \right]^3 P_L(q)^3 = b_2^3 \int d^3q P_L(q)^3 \quad (4.80)$$

for the terms containing tree-level two-point propagators and

$$\lim_{k_1, k_2 \rightarrow 0} B_g(k_1, k_2, k_3) = b_1 b_2 \left[\frac{115}{42} b_2 + b_3 - \frac{8}{3} \gamma_2^\times \right] (P_L(k_1) + P_L(k_2) + P_L(k_3)) \int d^3q P_L(q)^2 \quad (4.81)$$

These effects can be interpreted as the impact of small-scale perturbations on the formation of galaxies that perturbative approaches cannot model. To take account of these terms, usually, they are subtracted and substituted by *a posteriori* shot-noise terms that have the purpose of fitting these effects on large scales by introducing the N_0 , ϵ_0 and η_0 terms

$$P_{\text{shot,LO}}(k) = N_0, \quad (4.82)$$

$$B_{\text{shot,LO}}(k_1, k_2, k_3) = \epsilon_0 + \eta_0 [P_L(k_1) + P_L(k_2) + P_L(k_3)]. \quad (4.83)$$

Actually, by considering those terms as first-order terms of a general Taylor expansion of the stochastic terms, it is possible to write the next-to-leading corrections to the stochastic contribution that read as

$$P_{\text{shot,NLO}}(k) = N_2 k^2, \quad (4.84)$$

$$B_{\text{shot,NLO}}(k_1, k_2, k_3) = \epsilon_2 (k_1^2 + k_2^2 + k_3^2) + [\eta_{2,1} k_1^2 + \eta_{2,2} (k_2^2 + k_3^2)] P(k_1) + \text{cyc.} \quad (4.85)$$

To summarize, the comprehensive modelling in the bias loops framework involves combining the contributions of shot noise described in Eqs. 4.85 with the modelling of galactic bias outlined in Eqs. 4.77, 4.78. The bias loops approach enables the accurate modelling of small scales by incorporating a thorough treatment of galaxy bias, while also capturing the significant sources of non-Gaussianity.

5

Modelling and measuring two- and three-point statistics in configuration space

In this Chapter, we build upon the material presented in Sec. 4.2.3, i.e. the nonlinear models for the power spectrum and the bispectrum and present a strategy to obtain a three-point correlation function model at the next-to-leading order. Indeed, the goal of this Thesis is to fill the gap between configuration space and Fourier space in modelling three-point statistics in Perturbation Theory. I will start by introducing the basic idea for the case of two-point clustering and then move to the three-point correlation model.

Since the implementation of the model requires performing multi-dimensional numerical integrals that I estimate using the 2D-FFTLLog tool, I will discuss this technique and assess the adequacy of its implementation in the specific case of the three-point correlation function. To validate the three-point correlation function models, I will compare them to measurements of this function performed in simulated catalogues extracted from N-body simulation experiments. The validation tests and their results will be described in Chapters 6, 7, 8.

Here I also describe the estimator used to measure the three-point correlation function and the techniques that we have used to estimate the uncertainties and their covariance. Finally, I will summarise the material presented in this Chapter by presenting the pipeline that I have implemented to model the three-point correlation function and to compare these predictions to the actual measurements.

5.1 Configuration space vs. Fourier space

The next few years will be critical for studying the large-scale structure clustering properties of the universe. Large amounts of data from upcoming surveys will significantly expand our understanding of the structure and evolution of the universe from its earliest moments to the present day. Two main probes that will be used to investigate these properties are weak lensing and galaxy clustering. As the latter is the focus of this Thesis, there are two main

spaces in which clustering properties can be analysed: Fourier space and configuration space. These two spaces are equivalent, but they have some important differences. Fourier space is the most commonly used space for clustering analyses, and it has a long history of success in cosmological analyses (55; 56; 57; 58; 59; 60). One reason for its popularity is the ease with which it allows for the modelling of perturbation evolution and clustering properties in general. In fact, Fourier space allows for a relatively simple treatment of the equations governing the motion of cosmological fluids and the use of perturbative solutions in regimes where the data are expected to be affected by nonlinear evolution. Efficient estimators for the main correlation functions have been developed in Fourier space, making it a reliable space for comparing predictions to observations. However, Fourier space does have some limitations. Survey geometries are finite and affect the observations, and it can be challenging to model this in Fourier space. This issue can be addressed by deconvolving the correlation signal measurements or the theory using a window function (75; 76), but this is not necessary for configuration space, where complex survey geometries can be easily considered in the estimator. However, although configuration space has some advantages over Fourier space, estimators in configuration space, particularly for the three-point correlation function (3PCF), have historically been slower and computationally intensive to be used in a cosmological analysis. However, recent developments, such as the estimators proposed in (17; 77) for the 3PCF, have improved the efficiency and effectiveness of estimators in configuration space. Despite these advances, the status of models in configuration space is still somewhat behind those in Fourier space. Specifically, in the nonlinear regime, there is currently no Lagrangian modelling in configuration space, so the corresponding models must be obtained by using their Fourier counterparts by inverse Fourier transforming. However, it is still important to develop models in configuration space, as it allows for the recovery of scales in the nonlinear regime, where important cosmological information may be hidden. Additionally, modelling the baryon acoustic oscillation (BAO) peak in configuration space can provide important constraints on the universe's expansion history. In configuration space, the BAO peak is a single, distinct feature in the clustering pattern that can be isolated, modelled, and studied in its evolution history to investigate its properties in constraining universe's properties (129; 130; 131; 132; 133; 134; 135). For the two-point correlation function, treatments have been developed to the same perturbative order in the Standard Perturbation Theory (SPT). In contrast, for the 3PCF, only a tree-level model currently exists, while several one-loop models have been developed and tested for the bispectrum in both dark matter and galaxy/halo perturbations. The goal of this PhD Thesis is to bridge this gap by developing a three-point one-loop model that is equivalent to those developed for the bispectrum.

5.2 Perturbative model for the two-point correlation function

Regarding a perturbative treatment for the two-point correlation functions, due to the lack of a perturbative approach to fluid equations in configuration space, models are usually obtained by inverse Fourier transforming the corresponding power spectrum. In general, by assuming isotropy, Eq. 4.16 can be simplified in

$$\xi_m(r) = \int \frac{d^3k}{(2\pi)^3} P_m(k) e^{i\mathbf{k}\cdot\mathbf{r}} = \int_0^\infty \frac{dk}{2\pi^2} k^2 j_0(kr) P_m(k). \quad (5.1)$$

The Fourier transform can be evaluated either by direct integration or more efficiently by using the FFTLog algorithm. The real-space, one-loop PT prediction for the matter correlation function can be written as

$$\xi_m(r) = \xi_L(r) + \xi_m^{1\text{-loop}}(r) + \xi_m^{\text{ctr}}(r), \quad (5.2)$$

where, ignoring for the time being IR-resummation, we have

$$\xi_m^{1\text{-loop}}(k) = \xi_{22,m}^{1\text{-loop}}(k) + \xi_{13,m}^{1\text{-loop}}(k) \quad (5.3)$$

where ξ_L , $\xi_{13,m}$, $\xi_{22,m}$, ξ_m^{ctr} are the Fourier-transform pairs of P_L , $P_{13,m}$, $P_{22,m}$, $P_{\text{ctr},m}$. To simplify the analysis, it is possible to decompose the linear power spectrum using the FFTLog algorithm (136). This allows us to express the linear power spectrum as a sum of power laws:

$$P_L(k_n) = \sum_{m=-N_{\text{max}}/2}^{N_{\text{max}}/2} c_m k_n^{-2\nu_m}, \quad (5.4)$$

where $\nu_m = \nu + i\eta_m$, ν is a fixed number called bias, and $\eta_m = 2\pi/\log(k_{\text{max}}/k_{\text{min}})$. Using this decomposition, it is possible to express the one-loop power spectra as a matrix multiplication:

$$P_X(k_n) = k_n^3 \sum_{m_1, m_2} c_{m_1} k_n^{-2\nu_{m_1}} M_X(\nu_{m_1}, \nu_{m_2}) k_n^{-2\nu_{m_2}} c_{m_2}, \quad (5.5)$$

where X denotes $13, m$ or $22, m$. By performing an analytical integration, the correlation function can be expressed as a simple matrix multiplication based on the decompositions of

the linear and one-loop power spectra:

$$\int_0^\infty dx j_\mu(x) x^{2-2\nu} = \sqrt{\pi} 2^{1-2\nu} \frac{\Gamma((3+\mu-2\nu)/2)}{\Gamma((\mu+2\nu)/2)}. \quad (5.6)$$

Plugging all the expressions, we have:

$$\xi(r) = \sum_m c_m r^{2\nu_m-3} \frac{2^{-2\nu_m} \Gamma(\frac{3}{2} - \nu_m)}{\pi^{3/2} \Gamma(\nu_m)} \quad (5.7)$$

It is important to carefully choose the boundaries k_{\min} , k_{\max} , and the number of sampling points N_{\max} to avoid spurious oscillations in the resulting power spectrum or correlation function. In the analysis presented in Chaps. 7, 8, I choose $k_{\min} = 10^{-6}$, $k_{\max} = 10^2$.

5.3 Perturbative model for the three-point correlation function

Despite its importance in various fields of research, the use of the three-point correlation function in research has been limited in the past due to the high computational cost involved in estimating it. However, the development of a new estimator proposed in (17; 77) has revitalized the entire field of research involving the three-point correlation function (see Sec. 5.4). In this approach, decomposing the three-point correlation function into multipoles is deemed crucial. The multipole expansion is a mathematical technique employed to investigate the angular dependence of the function rather than its overall magnitude. This approach enables a more in-depth and specific examination of the relationships between variables, thereby providing valuable insights into the statistical characteristics of the system. In our case of interest, the aforementioned decomposition reads as

$$\zeta(r_{12}, r_{13}, r_{23}) = \sum_\ell \zeta_\ell(r_{12}, r_{13}) \mathcal{L}_\ell(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}), \quad (5.8)$$

where the Legendre transformation gives the coefficients of the expansion

$$\zeta_\ell(r_{12}, r_{13}) = \frac{2\ell+1}{2} \int_{-1}^{+1} d\mu \zeta(r_{12}, r_{13}, r_{23}) \mathcal{L}_\ell(\mu), \quad (5.9)$$

where $\mathcal{L}_\ell(\mu)$ are the Legendre polynomials and $\mu = \frac{\mathbf{r}_{12} \cdot \mathbf{r}_{13}}{r_{12} r_{13}}$. The decomposition of the three-point correlation function into multipoles proved not only to be a method of speeding up the

estimation process but also an essential element in achieving the modelling of the function. In terms of modelling, both the multipole expansion and the original form of the three-point correlation function are equally important, albeit they adopt different perspectives of the scientific concepts under investigation. The multipole expansion affords a more concise and manageable representation of the function, thereby facilitating the modelling process. Nonetheless, both methods are equivalent and hold the potential to further our comprehension of the relationships between variables.

Compared to its Fourier space equivalent, the bispectrum, the three-point correlation function remains a less developed area of modelling. This is due to the absence of a direct approach for constructing a perturbative three-point model in configuration space. Hence, the sole procedure for modelling the three-point correlation function at present is based on Eq. 4.24, which is presented here again for clarity

$$\zeta(r_{12}, r_{13}, r_{23}) = \int \frac{d^3k_1 d^3k_2 d^3k_3}{(2\pi)^9} B(k_1, k_2, k_3) e^{i(\mathbf{k}_1 \cdot \mathbf{r}_{12} + \mathbf{k}_2 \cdot \mathbf{r}_{13} + \mathbf{k}_3 \cdot \mathbf{r}_{23})} \times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \quad (5.10)$$

However, this equation, which is a 6-dimensional integral, is not particularly useful due to its complexity. To overcome this issue, the inverse Fourier transform is often evaluated using the multipole expansion in Eq. 5.9, which allows it to be expressed as a Hankel transform. This, in fact, allows for a more manageable and tractable representation of the function, which makes it easier to analyze and study. Indeed, the complexity of the problem can be reduced by integrating the angular variables and by using the expansion of a plane wave in spherical harmonics (see Appendix in (137) for a usage)

$$e^{i\mathbf{k} \cdot \mathbf{r}} = 4\pi \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{\ell} i^\ell j_\ell(kr) Y_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}^*(\hat{\mathbf{k}}), \quad (5.11)$$

so that Eq. 4.24 reduces to

$$\zeta_\ell(r_{12}, r_{13}) = (-1)^\ell \int \frac{k_{12}^2 k_{13}^2 dk_{12} dk_{13}}{(2\pi^2)^2} B_\ell(k_{12}, k_{13}) j_\ell(k_{12} r_{12}) j_\ell(k_{13} r_{13}). \quad (5.12)$$

In (53), a simplification was proposed. The treatment was restricted to the tree level, simplifying the expression by performing an analytical integration, leading to a 1D Hankel transform that can be considered numerically using the FFTLog algorithm. This treatment is only valid at the tree level - or leading order if infrared resummation is implemented -. The modelling process is facilitated by using the 1D-Hankel transform, and its computational

evaluation through the FFTLog algorithm (136), which allows for a more efficient and manageable representation of the 3PCF. However, this approach is limited in considering tree-level contributions and does not account for higher-order corrections. The model for galaxies, reduced to its real space part considered in this Thesis, reads as

$$\zeta(r_{12}, r_{13}, r_{23}) = \sum_{\ell=0}^{\ell=2} \zeta_{\text{pc},\ell}(r_{12}, r_{13}) \mathcal{L}_{\ell}(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}), \quad (5.13)$$

where

$$\zeta_{\text{pc},0}(r_{12}, r_{13}) = \frac{34}{21} (b_1^3 + 2b_2 b_1) [\xi_1^{[0]} \xi_2^{[0]}], \quad (5.14)$$

$$\zeta_{\text{pc},1}(r_{12}, r_{13}) = -b_1^3 [\xi_1^{[1+]} \xi_2^{[1-]} + \xi_2^{[1+]} \xi_1^{[1-]}], \quad (5.15)$$

$$\zeta_{\text{pc},2}(r_{12}, r_{13}) = \left(\frac{8}{21} b_1^3 + \frac{16}{15} \gamma_2 \right) [\xi_1^{[2]} \xi_2^{[2]}]. \quad (5.16)$$

The *pre-cyc* terms, referred to as pc, are intermediate stages in decomposing the three-point correlation function into multipoles. To obtain the full multipoles, it is generally necessary to sum the pre-cyc terms, permute them cyclically, and then perform the decomposition. This PhD Thesis focuses on the resummed case rather than the multipole representation, which will be clarified in the next chapters. In the tree-level treatment of the three-point correlation function, as shown, second-order bias parameters are introduced to consider the biases of luminous tracers. These parameters include the linear bias parameter b_1 , the local bias term b_2 , and a term proportional to the tidal field, which has been shown to be important for the analysis of the three-point correlation function as it captures the influence of tidal forces on the distribution of galaxies. This term corresponds to the bias parameter γ_2 in the loop bias treatment, which considers the effects of higher-order correlations. For a complete discussion of these three bias parameters, see Sec. 3.3.2. In the effort to compare the 3PCF against dark matter perturbations, it is possible to recover the matter case by setting $b_1 = 1$ and $b_2 = \gamma_2 = 0$.

5.4 Measuring the three-point correlation function

A direct estimate of the 3PCF would scale as $N(nV_{R_{\text{max}}})^2$, where N is the number of objects in the survey, n is the survey number density, and R_{max} is the maximum scale.

Recently, a new estimator has been introduced that significantly reduces the computational time required for estimating the three-point correlation function. This has revitalized research in this area, allowing for a more efficient study of the relationships between variables. The

development of this estimator has opened up new possibilities for research and has the potential to greatly provides us with cosmological information coming from higher order statistics. To explain how the estimator presented in (17) let me introduce the expansion of the three-point correlation function in multipoles that is typically used to analyze the angular dependence of the function rather than its overall magnitude. In this context, the expansion is being introduced without considering the average over the translation. This means that the position of the primary galaxy, denoted by \mathbf{s} , is being used as a reference point from which the distances between the other variables are calculated. This allows the angular dependence of the function to be studied more specifically rather than considering the function as a whole. This allows expanding the 3PCF into a series of multipoles, as done in (17) as follows

$$\zeta(\mathbf{r}_{12}, \mathbf{r}_{13}, \hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}; \mathbf{s}) = \sum_{\ell_1, \ell_2, m_1, m_2} \zeta_{\ell_1, \ell_2}^{m_1, m_2}(r_{12}, r_{13}, \mathbf{s}) Y_{\ell_1, m_1}(\hat{\mathbf{r}}_{12}) Y_{\ell_2, m_2}^*(\hat{\mathbf{r}}_{13}), \quad (5.17)$$

where the multipoles $\zeta_{\ell_1, \ell_2}^{m_1, m_2}$ are defined as

$$\zeta_{\ell_1, \ell_2}^{m_1, m_2}(r_{12}, r_{13}; \mathbf{s}) = \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{(4\pi)^2} \int d^2\hat{\mathbf{r}}_{12} \int d^2\hat{\mathbf{r}}_{13} \zeta(\mathbf{r}_{12}, \mathbf{r}_{13}, \hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}; \mathbf{s}) \times Y_{\ell_1, m_1}(\hat{\mathbf{r}}_{12}) Y_{\ell_2, m_2}^*(\hat{\mathbf{r}}_{13}) \quad (5.18)$$

where r_i represents the norm $|\mathbf{r}_i|$. Let me define the quantity

$$\zeta_\ell(\mathbf{r}_{12}, \mathbf{r}_{13}, \hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}; \mathbf{s}) = 2\pi \sum_m \zeta_{\ell, \ell}^{m, m}(r_{12}, r_{13}; \mathbf{s}) \quad (5.19)$$

and its average over translations

$$\zeta_\ell(\mathbf{r}_{12}, \mathbf{r}_{13}) = \frac{1}{V} \int d^3 \zeta_\ell(\mathbf{r}_{12}, \mathbf{r}_{13}, \hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}; \mathbf{s}) \quad (5.20)$$

where V is the survey volume. Now, considering a primary galaxy located at the generic position \mathbf{s} and the local contribution of an overdensity averaged on a bin-shell, this latter reads as

$$\bar{\delta}(\mathbf{r}_i, \mathbf{s}) = \int dr r^2 \phi(r, \mathbf{r}_i) \delta(\mathbf{r} + \mathbf{s}) \quad (5.21)$$

where \mathbf{r}_i represents the i -th radial bin and $\phi(|\mathbf{r}|)$ is the binning function demanding that we are in the bin by its second argument. Throughout the process of radial binning, it is possible to rewrite the multipole decomposition as follows

$$\hat{\zeta}_\ell(r_{12}, r_{13}, \mathbf{s}) = \frac{2\ell + 1}{(4\pi)^2} \delta(\mathbf{s}) \int d\Omega_1 d\Omega_2 \bar{\delta}(\mathbf{r}_{12}; \mathbf{s}) \bar{\delta}(\mathbf{r}_{13}; \mathbf{s}) \mathcal{L}_\ell(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13}). \quad (5.22)$$

where $\mathcal{L}_\ell(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13})$ are the Legendre polynomials and the hat denotes an estimate. Defining the coefficients

$$a_{\ell m}(r_i, \mathbf{s}) = \int d\Omega \bar{\delta}(\mathbf{r}_i, \mathbf{s}) Y_{\ell, m}^* = \int d\Omega Y_{\ell, m}^* \int dr' r'^2 \phi(r', r_i) \delta(\mathbf{r}' + \mathbf{s}), \quad (5.23)$$

and combining the equations, Eq. 5.22 becomes

$$\hat{\zeta}_\ell(r_{12}, r_{13}, \mathbf{s}) = \frac{1}{4\pi} \delta(\mathbf{s}) \sum_{m=-\ell}^{\ell} a_{\ell, m}(r_{12}, \mathbf{s}) a_{\ell, m}(r_{13}, \mathbf{s}). \quad (5.24)$$

From Eq. 5.24, it is clear why radial binning is crucial for speeding up the algorithm. By splitting the integral in Eq. 5.22, it is now possible to directly compute the pre-coefficient $a_{\ell m}$ for every bin and then combine them to compute the associated multipole. The computational time-scaling of evaluating the coefficient $a_{\ell m}$ is an order of $\mathcal{O}(nV_{max})$, where n is the average number of galaxies in a given volume V_{max} , and represents the main computational cost of the algorithm. Furthermore, one might think that the spherical harmonic transform involved in the computation of $a_{\ell m}$ coefficients represent the main computational cost, scaling as $N_g^{3/2}$, being N_g the number of spatial cells of a spherical grid. Actually, because of the low number of multipoles usually required, these coefficients are commonly evaluated in a direct way; i.e. spherical harmonics evaluated at angles given by a galaxy's location are computed. The final step of the algorithm is to get the final estimate of the multipoles by averaging over the possible direction of the displacement vector \mathbf{s} applying the averaging over translation in Eq. 5.24. To understand the algorithm described in the text, it is helpful to refer to Fig. 5.1.

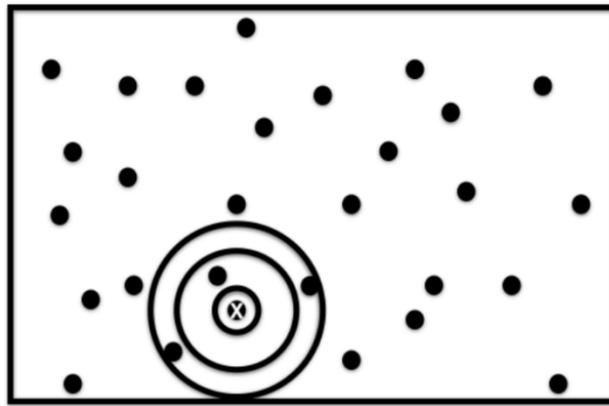


Figure 5.1 Figurative description of the algorithmic approach presented in (17). The primary galaxy is positioned at X , and around it, spherical shells are used to divide the secondary galaxies into radial bins. The coefficients in Eq. 5.23 are then calculated through these radial bins.

Summarising, the algorithm begins by fixing a primary galaxy as the center of the coordinate

system and creating spherical shells around it. The second galaxies in the group are then divided into radial bins, or groups, based on their distance from the primary galaxy. Next, the coefficients defined in Eq. 5.23 are computed and averaged over the entire survey volume being studied. The primary galaxy's location is then changed, and the process is repeated. This step allows the algorithm to analyze the average properties of the group of galaxies from different perspectives rather than just from the viewpoint of the primary galaxy. Finally, the multipoles calculated in each iteration are averaged together to provide the final estimation of the average properties of the group of galaxies. This final result can then be used to provide a fast estimation of the three-point correlation function.

5.5 Modelling the three-point correlation function in the small-scale regime: the 2D-FFTLog algorithm

In order to bridge the gap between Fourier space and configuration space three-point statistics models, as outlined as the objective of this Thesis, it is now legitimate and necessary to ask: what strategy should we use to bridge this gap between Fourier space and configuration space three-point statistics models? It will be necessary to consider the relationship between the bispectrum and the 3PCF. As previously mentioned, the two quantities are theoretically linked through a 6D integral, as expressed in Eq. 4.24. However, in practice, it is common to decompose this integral into multipoles, as shown in Eq. 5.8 and Eq. 5.9. The complexity of the problem of developing a three-point model for the 3PCF in configuration space can be reduced by integrating the angular variables and expanding plane waves in spherical harmonics.

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{+\infty} \sum_{m=-l}^l i^l j_l(kr) Y_{lm}(\hat{\mathbf{r}}) Y_{lm}^*(\hat{\mathbf{k}}). \quad (5.25)$$

This approach has been demonstrated to be effective in previous work (53). By simplifying the problem in this way, it may be possible to develop an accurate and efficient link between models

$$\zeta_\ell(r_{12}, r_{13}) = (-1)^\ell \int \frac{k_{12}^2 k_{13}^2 dk_{12} dk_{13}}{(2\pi^2)^2} B_\ell(k_{12}, k_{13}) j_\ell(k_{12}r_{12}) j_\ell(k_{13}r_{13}). \quad (5.26)$$

Generally speaking, this mathematical relationship provides us with a means of bridging the gap with Fourier space. Nonetheless, the integral in Eq. 5.26, while being theoretically significant for comprehending the relationship between the bispectrum and the 3PCF, still poses computational burdens that have made its direct computation challenging. Actually,

evaluating the integral from a $N_r \times N_r$ grid by employing the standard quadrature method to compute Eq. (5.26) needs an order of $N_r^2 N_k^2$ steps, where N_r and N_k represent the number of sampled points in k and r . Additionally, the rapid and oscillatory nature of Bessel functions necessitates performing a great many integration steps, making the precise evaluation of the integrals computationally challenging. To counteract these difficulties, the FFTLog algorithms - also known as Fast Fourier Transform with a Logarithmic transform - have arisen as an effective tool for efficiently and accurately evaluating Hankel transforms. By using FFTLog algorithms, it is feasible to overcome the computational challenges associated with computing the integral in 5.26 and develop an efficient implementation of the model for the 3PCF.

5.5.1 2D-FFTLog technique

A common problem in physics is the numerical evaluation of integrals involving the product of two or more Bessel functions as in Eq. 5.26. Evaluating these integrals represents the main computational burden of our 3PCF modelling procedure. Evaluating the integral from a grid $N_r \times N_r$ using the standard quadrature method to compute Eq (5.26) requires an order of $N_r^2 N_k^2$ steps where N_r and N_k are the number of k and r sampled points. Furthermore, the rapid and oscillatory behaviour of Bessel functions requires performing many integration steps, making the accurate evaluation of the integrals computationally challenging. The 1D-FFTLog method, originally conceived to address atomic physics problems (138) and then applied to cosmology by (136), has been used over the years to efficiently evaluate Fourier transform with logarithmic variables involving single Bessel integration of the form $\int_0^\infty dk f(k) j_\ell(kr)$ where $f(k)$ is a generic smooth function. The main idea of the 1D-FFTLog algorithm is using the expansion $f(k) = \sum_m c_m k^{z_m}$ with z_m , in general, a complex number to obtain a term that can be integrated analytically to speed up the computation by evaluating a sum over c_m Fourier coefficients instead of a 1D-integral. The FFTLog approach has proven to be useful, as can be seen by three notable and quite different applications of the method. The first one is the evaluation of the Bessel integrals in the angular power spectra and bispectra expressions (139; 140; 141). The second one by (142) consists of using 1D-FFTLog to evaluate Bessel integrals coming from the deconvolution of multi-dimensional integrations. Finally, FFTLog has been used to model higher-order statistics in Fourier space. In this latter case, there is no Bessel integral involved. Instead, the idea is to obtain integrable expressions to which the FFTLog tool can be applied. Examples include one-loop (143; 144) and two-loop (77; 145; 146) perturbation models. Furthermore, an application of the 2D-FFTLog algorithm to the tree-level anisotropic 3PCF can be found in (147).

For the 3PCF model, in general, one has to consider the 2D-dimensional integral in Eq. 5.26, which involves Bessel functions. In this case, the usual 1D-FFTLLog approach is useless since one cannot separate terms involving k_{12} and k_{13} . (148) has recently proposed an extension of 1D-FFTLLog that can be adapted to perform fast and accurate evaluations of integrals involving the product of two Bessel functions. This work was originally addressed to the computation of real-space 3PCF covariances, similar to Eq. 5.26. The 2D extension of the FFTLog algorithm is sensitive, as the 1D version, to all sources of aliasing and ringing (136). The grid we use is 2D, parametrised by two k_1, k_2 after integrating over the angle between the two, θ . In general, grid spacing Δk , regulates the integration accuracy. The choice of k_{\min} and k_{\max} is crucial to avoid aliasing and ringing effects. Zero padding is also advised to reduce the possibility of generating spurious wiggles.

In this Thesis, we aim to use and adapt the 2D-FFTLLog to inverse Fourier transform a general bispectrum model, focusing on the one-loop perturbative expansion. In order to present how the algorithm works, let me start by considering the main input object, the dimensionless bispectrum multipoles defined as:

$$\Delta_\ell(\mathbf{k}_{12}, \mathbf{k}_{13}) = \frac{k_{12}^3 k_{13}^3}{(2\pi^2)^2} B_\ell(\mathbf{k}_{12}, \mathbf{k}_{13}). \quad (5.27)$$

The dimensionless bispectrum multipoles can be decomposed so into a series of products of two-power laws, assuming that k_1 and k_2 are the same and logarithmically sampled

$$\Delta_\ell(\mathbf{k}_p, \mathbf{k}_q) = \frac{1}{N^2} \sum_{m=-\frac{N}{2}}^{N/2} \sum_{n=-\frac{N}{2}}^{N/2} \tilde{c}_{\ell, mn} k_0^{-i\eta_m} k_0^{-i\eta_m} k_p^{\nu_1+i\eta_m} k_q^{\nu_2+i\eta_m} \quad (5.28)$$

where $\eta_m = 2\pi \frac{m}{N\Delta_{\ln k}}$, N is the size of k -array, ν_1, ν_2 are the so-called bias parameters, i.e. the real part of the power laws, $\Delta_{\ln k}$ is the logarithmic spacing in k , coefficients $c_{\ell, mn}$ are given by the discrete Fourier transform.

$$\tilde{c}_{\ell, mn} = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \frac{\Delta_\ell(\mathbf{k}_p, \mathbf{k}_q)}{k_p^{\nu_1} k_q^{\nu_1}} e^{-2\pi i(m p + n q)/N}. \quad (5.29)$$

To remove sharp edges at the boundary of $c_{\ell, mn}$ a filtering process of the form

$$c_{\ell, mn} = W_m W_n \tilde{c}_{\ell, mn} \quad (5.30)$$

is usually applied, where \mathbf{W} is a one-dimensional window function (for details, see (149)). Using the decomposition in 5.26, with the following mathematical steps, becomes

$$\zeta_\ell(r_{12}, r_{13}) = (-1)^\ell \sum_{m,n=-N/2}^{N/2} c_{\ell,mn} \frac{1}{k_{1,0}^{i\nu_m}} \frac{1}{k_{2,0}^{i\nu_n}} \quad (5.31)$$

$$\times \int \frac{dk}{k_1} k_1^{\nu_1+i\nu_m} j_\ell(k_1 r_1) \int \frac{dk}{k_2} k_2^{\nu_2+i\nu_m} j_\ell(k_2 r_2) \quad (5.32)$$

$$= (-1)^\ell \frac{\pi}{16 r_{12}^{\nu_1} r_{13}^{\nu_2}} \sum_{m,n=-N/2}^{N/2} c_{\ell,mn} k_0^{-i(\eta_m+\eta_n)} r_{12}^{-i\eta_m} r_{13}^{-i\eta_n} \quad (5.33)$$

$$\times g_\ell(\nu_1 + i\eta_m) g_\ell(\nu_2 + i\eta_n) \quad (5.34)$$

where $g_\ell(\omega) = 2^\omega \frac{\Gamma(\frac{\ell+\omega}{2})}{\Gamma(\frac{3+\ell-\omega}{2})}$. The range of validity of ν_1 and ν_2 are $-\ell < \nu_1, \nu_2 < 2$. The 2D-FFTLLog method uses 2D-FFT twice, and thus, it is an $\mathcal{O}(N^2 \log N)$

Actually, when dealing with data, they are usually bin averaged, i.e. multipoles are obtained by considering the following integral for every pair (r_{12}, r_{13})

$$\zeta_\ell(\bar{r}_i, \bar{r}_j) = \frac{\int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_1 \int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_2 r_1^2 r_2^2 \zeta_\ell(r_1, r_2)}{\int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_1 r_1^2 \int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_2 r_2^2}. \quad (5.35)$$

Using (5.35) into (5.31) and recalling the denominator as A it is possible to get the bin-averaged expression for the 2D-FFTLLog algorithm in Eq. (5.36)

$$\zeta_\ell(\bar{r}_i, \bar{r}_j) = \frac{\pi \bar{r}_{i,\min}^{2-\nu_1} \bar{r}_{j,\min}^{2-\nu_2}}{16 A N^2} \sum_{m,n=-N/2}^{N/2} c_{\ell,mn} k_0^{-i(\eta_m+\eta_n)} r_{12}^{-i\eta_m} r_{13}^{-i\eta_n} \times g_\ell(\nu_1 + i\eta_m) g_\ell(\nu_2 + i\eta_n) s(D - \nu_1 - i\eta_m, \lambda) s(D - \nu_2 - i\eta_n, n) \quad (5.36)$$

where

$$s(D, n) = \frac{n-1}{D}, \quad (5.37)$$

where $\frac{\bar{r}_{i,\max}}{\bar{r}_{i,\min}} = n$ is the linear bin width, and

$$A = \int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_1 r_1^2 \int_{\bar{r}_{i,\min}}^{\bar{r}_{i,\max}} dr_2 r_2^2 = \bar{r}_{i,\min}^2 \bar{r}_{j,\min}^2 [s(2, n)]^2. \quad (5.38)$$

The 2D-FFTLLog machinery described above provides an important calculation tool for evaluating the mathematical relationship between Fourier space and configuration space at the level of three-point statistics. However, it relies on certain input parameters, which are

crucial to its accuracy and must be carefully considered. These parameters must be carefully studied to ensure the accuracy of the 2D-FFTLog algorithm.

5.5.2 2D-FFTLog calibration

In (150), we studied the algorithm as a function of input parameters. We considered the parameters k_{\min} , k_{\max} , the minimum and maximum values over we perform Eq 5.26, a damping factor k_0 , i.e.

$$P_{k_0}(k) = P_L(k)e^{-k^2/k_0^2}, \quad (5.39)$$

we used to study convergence properties and ringing sensibility of 1D and 2D-FFTLog algorithms, N_k and N_μ numbers of point in k -direction (1D and 2D) in θ -direction (only 2D), with $P_L(k)$ generated throughout the Boltzmann code CAMB (118). In Fig. 5.2, we tested,

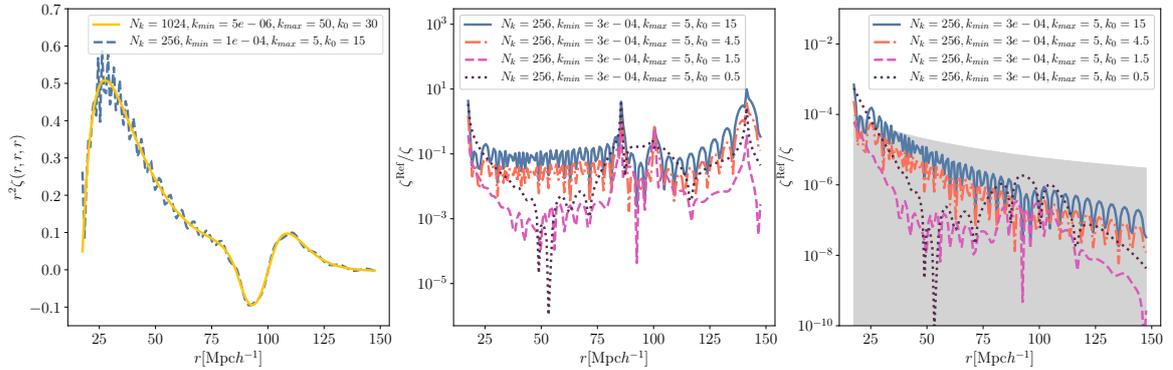


Figure 5.2 Different computation of ζ for equilateral computation with different grid parameters using 1D-FFTLog. Left: plot of different evaluations. Middle: ratio of previous computations. Right: residuals of previous evaluations.

for an equilateral configuration at $z = 0.5$, different usage of parameters reproducing the LO-1D model at the centre of the bin and considering an extended k -range from $k_{\min} = 5 \times 10^{-6}$ and $k_{\max} = 5 \times 10^1$ as a reference model. An incorrect using of damping k_0 over a shorter grid implies a source of ringing effect. Using $k_0 = 1.5h\text{Mpc}^{-1}$ is the best solution to mimic a usage of a larger grid and minimise ringing effects. In Fig. 5.3, we did the same test but using the 2D version of the FFTLog algorithm, considering the leading-order model computed through the 2D-FFTLog algorithm, from now dubbed LO-2D, including several sample points in θ -direction but, differently from the previous case, considering the bin-averaged model. In this case, we fix the damping factor according to the previous case $k_0 = 1.5h\text{Mpc}^{-1}$ where

$$B_{k_0}(k_1, k_2, k_3) = B(k_1, k_2, k_3)e^{-(k_1^2+k_2^2+k_3^2)/k_0^2}, \quad (5.40)$$

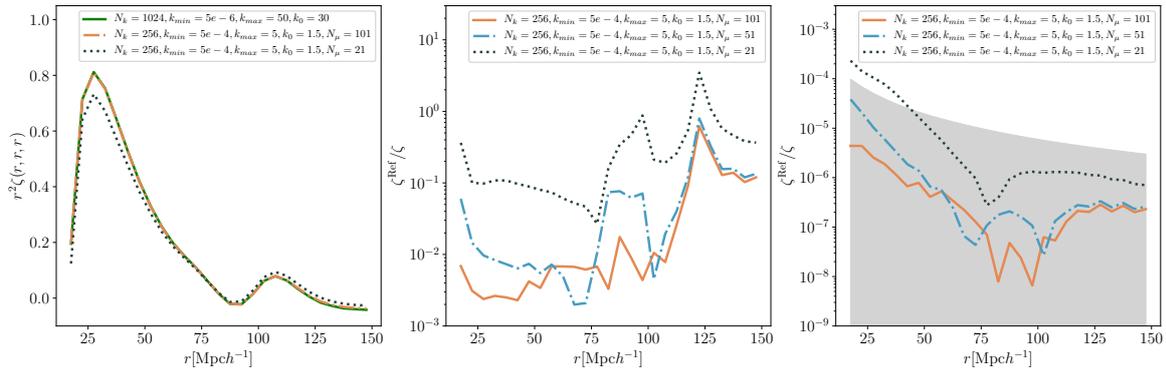


Figure 5.3 Different computation of ζ for equilateral computation with different grid parameters using 2D-FFTLog. Left: plot of different evaluations. Middle: residuals of previous computations. Right: difference between previous evaluations compared to theoretical error assuming theoretical covariance.

We found N_θ needs to be at least equal to 51 to minimise the difference with the reference model (1D-FFTlog evaluation). We use $N_\theta = 51$, $N_k = 256$, $k_{\min} = 5 \times 10^{-6}$ and $k_{\max} = 5 \times 10^1$ in our computations.

5.6 Model vs. simulated data: evaluating statistical errors

Accurately estimating uncertainty is essential in any cosmological analysis, including those that use the three-point correlation function to study the clustering of galaxies. There are two main approaches for estimating the covariance matrix, which represents the uncertainty in the measurement: theoretical estimates and numerical estimates based on mock analyses of simulated data sets. Theoretical estimates are typically based on measures of higher-order statistics if non-Gaussian signals are considered, while numerical estimates are obtained by analysing mock catalogues. Both of these approaches have their own benefits and limitations, and it is important to carefully consider which method is most appropriate for a given analysis.

5.6.1 Numerical covariance

One common way to use mock data sets is to estimate the covariance matrix, which is a measure of how much data are expected to vary from one sample to another. To estimate the covariance matrix from a mock data set, one typically follows a specific set of steps. These steps may involve analysing the mock data in a manner similar to how the real data will be analysed and using statistical techniques to calculate the covariance matrix from the results

of this analysis. The exact steps involved in this process will depend on the specific statistical methods used and the specific characteristics of the data being analysed.

Once the covariance matrix has been estimated from the mock data set, it can be used to quantify the uncertainties in real data analysis. For example, the covariance matrix can be used to calculate the standard errors of the estimates or to construct confidence intervals around the estimates. This information can be used to assess the reliability of the results of the analysis and to make more informed decisions based on the data. Covariances from mocks are usually estimated as follows.

$$C_{ij} = \frac{1}{N_m - 1} \sum_{k=0}^{N_m} \left(\hat{\zeta}_i^k - \bar{\zeta}_i \right) \left(\hat{\zeta}_j^k - \bar{\zeta}_j \right) \quad (5.41)$$

where N_m refers to the number of mock, i and j identify any two triangles in the samples, $\hat{\zeta}$ and $\bar{\zeta}$ refer, respectively, to the measured 3PCF in the n -th mock and the average among the mocks. computing the covariance matrix from a set of mocks (simulated data sets used to represent the statistical properties of the real data) can be computationally expensive, especially if a large number of mocks is needed to achieve statistical significance. (151) investigated the impact of using different estimates of covariance based on different sets of mocks and also compared these estimates to a theoretical estimate of the covariance matrix, which is presented in the following Sec.

Due to the high computational cost deriving from the computation of several mocks, some alternative approaches to estimating covariances matrices have been provided. It is worth highlighting the jackknife and bootstrap techniques which are internal error estimators that involve subsampling the original catalogue and computing the correlation function in all but one subsample (jackknife) or in a random selection of subsamples (bootstrap). This approach has the advantage of being fast and accounting for possible systematic effects and biases, as well as contributions from higher-order correlations. However, it may not provide an accurate estimate of the variance due to its reliance on the survey volume, and the off-diagonal terms may be noisy, which can affect the analysis of the correlation function in configuration space.

5.6.2 Theoretical covariance

An alternative approach in estimating covariance matrices has been proven in the theoretical estimates. These latter bring computational benefits, avoiding the computation of a large number of mocks from an N-body simulation. Several works addressed two and three-point cases (76; 152; 153; 154).

Regarding the two-point correlation function, it has been shown in (155) that, under the Gaussian hypothesis, the computation of the covariance matrix reduces to

$$\text{Cov}(\hat{\xi}^a, \hat{\xi}^b) = \frac{1}{2\pi^2} \int_0^\infty dk k^2 \sigma(k) \bar{j}_0(kr_a) \bar{j}_0(kr_b) \quad (5.42)$$

where

$$\bar{j}_0(kr_a) = \frac{4\pi}{V_{r_a}} \int_{r_a - \Delta s/2}^{r_a - \Delta r/2} r^2 j_0(kr) dr, \quad (5.43)$$

$$\sigma(k) = \frac{2}{V_s} \left(P(k)^2 + \frac{1}{\hat{n}} \right)^2, \quad (5.44)$$

where it has been defined the bin-averaged Bessel function over the volume encapsulated between $r_{a,max}$ and $r_{a,min}$.

Similarly, theoretical estimates have been proved useful for providing uncertainties for higher-order statistics (76; 152; 154). Even in this case, under the Gaussian hypothesis, it is possible to derive simplified expressions for the three-point correlation function multipoles (17) as follows

$$\begin{aligned} \text{Cov}(\hat{\zeta}_{\ell_1}^{ab}, \hat{\zeta}_{\ell_2}^{cd}) &= \frac{4\pi}{V} (2\ell_1 + 1) (2\ell_2 + 1) (-1)^{\ell_1 + \ell_2} \\ &\times \int r^2 dr \mathcal{M}_{\ell_1 \ell_2} \times \left\{ (-1)^2 \xi_0(r) \left[f_{\ell' \ell_1 \ell_2}(r; r_a, r'_a) f_{\ell' \ell_1 \ell_2}(r; r_b, r'_b) \right. \right. \\ &+ (-1)^{\binom{\ell_1 + \ell_2 + \ell'}{2}} \left[f_{\ell_1 \ell_1}(r; r_a) f_{\ell_2 \ell_2}(r; r'_a) f_{\ell' \ell_1 \ell_2}(r; r_b, r'_b) \right. \\ &+ f_{\ell_1 \ell_1}(r; r_a) f_{\ell_2 \ell_2}(r; r'_b) f_{\ell' \ell_1 \ell_2}(r; r_b, r'_a) \\ &+ f_{\ell_1 \ell_1}(r; r_b) f_{\ell_2 \ell_2}(r; r'_a) f_{\ell' \ell_1 \ell_2}(r; r_a, r'_b) \\ &\left. \left. + f_{\ell_1 \ell_1}(r; r_b) f_{\ell_2 \ell_2}(r; r'_b) f_{\ell' \ell_1 \ell_2}(r; r_a, r'_a) \right] \right\}, \quad (5.45) \end{aligned}$$

where

$$\mathcal{M}_{\ell_1 \ell_2} = \sum_{\ell'} (2\ell' + 1) \binom{\ell_1 \quad \ell_2 \quad \ell'}{0 \quad 0 \quad 0}^2 \quad (5.46)$$

and

$$f_{\ell \ell}(r; r_a) = \int \frac{k^2 dk}{2\pi^2} P(k) j_\ell(kr_a) j_\ell(kr), \quad (5.47)$$

$$f_{\ell' \ell_1 \ell_2}(r; r_a, r'_a) = \int \frac{k^2 dk}{2\pi^2} P(k) j_{\ell_1}(kr_a) j_{\ell_2}(kr'_a) j_{\ell'}(kr) \quad (5.48)$$

It is worth stressing that this expression was derived assuming Gaussian assumptions. One advantage is that it makes it relatively easy to compute the theoretical covariance of the three-point correlation function. However, this assumption also means that the three-point correlation function covariance may bias the true underlying relationship between variables in the data, as it does not consider any non-Gaussian contributions. Non-Gaussian contributions refer to deviations from the Gaussian distribution that usually significantly affect the data's analysis, and ignoring them can lead to inaccurate estimates. Therefore, when using the three-point correlation function to analyse mock or real data, it is important to consider the potential underestimation caused by excluding non-Gaussian contributions (151).

5.7 The three-point correlation function pipeline

The goal of this Thesis is to bridge the gap between configuration and Fourier spaces in modelling three-point statistics. The machinery over which next-to-leading order 3PCF models are evaluated is based on the 2D-FFTLog algorithm, which allows generating any 3PCF model from a given bispectrum model. This represents a significant improvement over the previous approach - limited to the first perturbative order - of predicting 3PCF models. However, the computational cost of this procedure is not prohibitive but still significant. This thesis uses this procedure to compare high-order perturbative 3PCF models with simulated clustering measurements, focusing on dark matter and galaxy/haloes perturbations. The strategy used in this Thesis can be illustrated in Fig. 5.4 and resumed in the following steps:

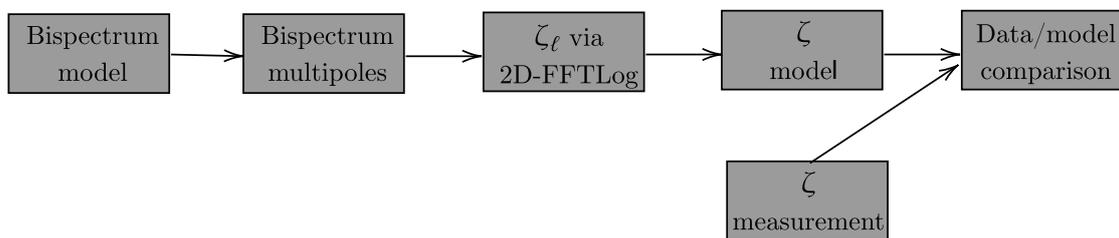


Figure 5.4 Flow chart describing the strategy assumed in this Thesis.

- *Step 1:* Generating an input bispectrum. In this step, the focus is on modelling the bispectrum at the next-to-leading order (i.e. the one-loop level using infrared resummation). The bispectrum at the leading order will also be used, with the understanding that the corresponding 3PCF is equivalent to that presented in (53). In terms of modelling at the next-to-leading order, the focus will initially be on dark matter perturbations and then on galaxy perturbations (126; 128). To obtain the respective bispectrum models, a code written in a combination of C++ and Fortran90 was implemented to efficiently calculate the individual loop terms involved in this modelling. The integrals involved in the next-to-leading order bispectrum model are typically three-dimensional. However, some of these integrals can be simplified into two-dimensional or one-dimensional integrals. The multidimensional integrals will be computed using the integration package *CUBA*¹, while the one-dimensional integrals will be computed using the *GSL-Gnu Scientific Library*². The input grid in which the bispectrum model has computed consists of a $256 \times 256 \times 51$, where the first two dimensions refer to k_1 , k_2 and the last one to the angle θ , the angle between them.
- *Step 2:* Computing the multipoles of the bispectrum model using the Gauss-Legendre procedure throughout the `numpy` specific package. The angular dimension is sampled at 51 equally spaced points within the interval determined by the Gauss-Legendre procedure.
- *Step 3:* Using the 2D-FFTLog algorithm to generate the multipoles of the 3PCF. The algorithm's input is the bispectrum multipoles calculated in the previous step. The algorithm has two parameters, ν_1 and ν_2 , which control its accuracy. The grid dimensions also affect the algorithm's accuracy and were discussed in the previous section.
- *Step 4:* The multipoles produced by the output of the 2D-FFTLog algorithm are summed to obtain the 3PCF. In general, the sum is infinite, but it is, in practice, truncated at $\ell_{\max} = 10$. While this may have some consequences that will be discussed later, it is consistent with the approach used in the measurements.
- *Step 5:* Measurements are performed on a catalogue of input objects. This Thesis will examine three catalogues: one containing only dark matter perturbations and two containing galaxies and halos. An estimator is used to obtain the measurements based on the procedure introduced in (17; 77). Due to the computational cost of the measurements, the maximum number of multipoles is set to $\ell_{\max} = 10$. This impact the

¹feynarts.de/cuba

²gnu/software/gsl

3PCF in most isosceles configurations, for which the higher multipoles are expected to have a significant weight in the Legendre resummation. The uncertainties associated with the measurements will be computed, including both Gaussian uncertainties from the theoretical covariance matrix 5.45 and those from a covariance matrix reconstructed from different mock catalogues.

- *Step 6:* The generated models are compared with measurements extracted at different redshifts, where different perturbative order models are expected to provide slightly different predictions. Initially, the comparison will focus on the data-model comparison of dark matter perturbations only and then move on to the more challenging case of the galaxy/haloes model. In this case, the inferred values of the bias parameters involved in the model will be investigated. The constraining power of a joint 2PCF and 3PCF analysis at both leading and next-to-leading orders will be studied.

6

Testing the next-to-leading order model for the matter

In Chapter 5 we presented the theory behind the next-to-leading order model for the matter or galaxy/halos 3PCF and the pipeline for its numerical implementation and comparison with the data. In this Chapter we make use of that pipeline to generate several matter 3PCF models, compare their performances and, finally, to gauge their accuracy by comparing their predictions with the measurements performed on a set of simulated data. The results presented in this chapter have been taken from the article *Modelling the next-to-leading order matter three-point correlation function using FFTLog* by M. Guidi, A. Veropalumbo, E. Branchini, A. Eggemeier and C. Carbone, currently submitted to JCAP for publication (150).

6.1 3PCF matter models

In my study, I have generated four different models for the matter 3PCF, each of them identified by a different acronym listed in the first column of Tab. 6.1. The first model, dubbed LO-1D, only uses leading order perturbation theory, equivalent to the leading order matter bispectrum (see (15)). Since it relies on the analytic simplification presented in (53) is implemented using the 1D-FFTLog algorithm. The LO-2D also relies on leading order treatment but drops the analytic simplification and, therefore, makes use of 2D-FFTLog for its implementation. Since these two models are expected to provide identical predictions, their comparison will constitute a sanity check for the correct implementation of the 2D-FFTLog procedure. The third model, dubbed SPT-NLO, is based on the next-to-leading order model for the matter bispectrum presented in Sec. 4.2.1 and uses 2D-FFTLog for the 3PCF model implementation. Finally, EFT-NLO relies on the next-to-leading order bispectrum model implemented within the EFT framework presented in Sec. 4.2.2 and on the 2D-FFTLog algorithm. All these 3PCF models assume the same cosmology of the DEMNUni simulation

described in the next section. As a result they do not have free parameters apart from EFT-NLO that has four of them.

A full 3PCF model specifies the amplitude of the correlation for all triangle configurations and its representation requires some convention for the ordering of the triangle sides. In this Thesis we use the following convention: $r_{12} \leq r_{13} \leq r_{23}$.

3PCF model name	Method	Degrees of freedom
LO-1D	1D-FFTLog	0
LO-2D	2D-FFTLog	0
SPT-NLO	2D-FFTLog	0
EFT-NLO	2D-FFTLog	4

Table 6.1 Summary of models used in this paper, methods used to compute them and their degrees of freedom.

6.2 The DEMNUni simulations

The models presented above have been tested against measurements extracted from the “Dark Energy and Massive Neutrinos Universe” (DEMNUi) N-body simulations (156). The DEMNUi simulations have been produced with the aim of investigating large-scale structures in the presence of massive neutrinos and dynamical dark energy, and they were conceived for the nonlinear analysis and modelling of different probes, including dark matter, halo, and galaxy clustering (70; 157; 158; 159; 160; 161; 162; 163), CMB lensing, Sunyaev-Zel’dovich (SZ) and Integrated Sachs–Wolfe (ISW) effects (156; 164; 165), cosmic void statistics (166; 167; 168; 169), and cross-correlations among these probes (170; 171). To this end, they combine a good mass resolution with a large volume to include perturbations both at large and small scales. In fact, these simulations follow the evolution of 2048^3 cold dark matter (CDM) and, when present, 2048^3 neutrino particles in a box of side $L = 2 \text{ Gpc}/h$. The fundamental frequency of the comoving particle snapshot is, therefore, $k_F \approx 3 \times 10^{-3} h/\text{Mpc}$, while the chosen softening length is $20 \text{ kpc}/h$. The simulations are initialized at $z_{\text{ini}} = 99$ with Zel’dovich initial conditions. The initial power spectrum is rescaled to the initial redshift via the rescaling method developed in Ref. (172). Initial conditions are then generated with a modified version of the N-GenIC software, assuming Rayleigh random amplitudes and uniform random phases.

In this Thesis, I have only considered the set of standard Λ CDM model simulation with no neutrinos. From this I have selected two snapshots, at $z = 0.49$ and $z = 1$ out of the 63 available, to match the typical redshifts that will be sampled by next generation spectroscopic surveys like DESI (42), Euclid (41) and Nancy Grace Roman Space telescope (44). Since the number of dark matter particles is far too large for quick and efficient estimates of the matter 3-point correlation function and since the scale at which our 3PCF model is expected to break down is much larger than the inter-particle separation, we decided to random downsample the particle distribution to a dilution factor 2.5%. The corresponding shot-noise error level matches that of the BOSS galaxy sample at $z \sim 0.5$ and the one expected for Euclid at $z \sim 1$. The volume of the box is also comparable to that of the above mentioned catalogs at the corresponding redshifts. As a result, the shot-noise and cosmic variance terms adopted in the Gaussian covariance matrix described in Sec. 5.6.2 are realistic and comparable to those expected for the real galaxy surveys. Finally, the cosmological model adopted in the 3PCF model and to measure the matter 3PCF is the same adopted in the simulation, i.e. the Planck 2013 cosmology with $\Omega_b = 0.05$, $\Omega_M = 0.32$, $h = 0.67$, $n_s = 0.96$ and $\sigma_8 = 0.846$.

6.2.1 Dark matter clustering measurements

For an efficient triplet counting procedure, we use the Szapudi-Szalay estimator complemented with the Spherical Harmonics Decomposition (SHD) technique presented in Sec. 5.4 (17; 74). This technique allows to estimate the Legendre coefficients, $\hat{\zeta}_\ell(r_{12}, r_{13})$, up to $\ell_{max} = 10$ and then obtaining the approximate estimate of the triangle-binned 3PCF:

$$\hat{\zeta}(r_{12}, r_{13}, r_{23}) = \sum_{\ell=0}^{\ell_{max}} \hat{\zeta}_\ell(r_{12}, r_{13}) \tilde{\mathcal{P}}_\ell(r_{12}, r_{13}, r_{23}); \quad (6.1)$$

where $\tilde{\mathcal{P}}_\ell(r_{12}, r_{13}, r_{23})$ is the Legendre polynomial of order ℓ weighted over the triangle

$$\tilde{\mathcal{P}}_\ell(r_{12}, r_{13}, r_{23}) = \frac{\int_{r_{12}-\Delta_r/2}^{r_{12}+\Delta_r/2} dp p^2 \int_{r_{13}-\Delta_r/2}^{r_{13}+\Delta_r/2} dq q^2 \int_{r_{23}-\Delta_r/2}^{r_{23}+\Delta_r/2} ds s^2 \Theta(p, q, s) \mathcal{P}_\ell(p, q, s)}{\int_{r_{12}-\Delta_r/2}^{r_{12}+\Delta_r/2} dp p^2 \int_{r_{13}-\Delta_r/2}^{r_{13}+\Delta_r/2} dq q^2 \int_{r_{23}-\Delta_r/2}^{r_{23}+\Delta_r/2} ds s^2 \Theta(p, q, s)} \quad (6.2)$$

with

$$\Theta(r_{12}, r_{13}, r_{23}) = \begin{cases} 1 & \text{if } \left| \frac{r_{12}^2 + r_{13}^2 - r_{23}^2}{2r_{12}r_{13}} \right| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

where $\tilde{\mathcal{P}}_\ell(r_{12}, r_{13}, r_{23})$ is the Legendre polynomial of order ℓ integrated over the triangle (see Appendix A of (74)), with $\mathcal{P}_\ell(r_{12}, r_{13}, r_{23})$ being the Legendre polynomial computed at a fixed triangle configuration. This estimator is known to be difficult in handling isosceles triangle configurations ($r_{12} = r_{13}$) using a reasonably small number of multipoles ℓ_{max} (74). This problem also extends to nearly isosceles triangles for which $r_{13} \simeq r_{12}$. For this reason, and to separate these cases from the results obtained for all other configurations, we characterise measurements using two parameters the minimum separation, r_{min} , and the minimum value, η_{min} , for the relative difference

$$\eta \equiv \frac{r_{13} - r_{12}}{\Delta r}. \quad (6.4)$$

For a given set Δr - i.e. the bin-width -, r_{min} and r_{max} the case $\eta_{min} = 0$ includes all triangles. The case $\eta_{min} = 1$ includes all triangles with $|r_{13} - r_{12}| > \Delta r$. And so on. Inputs to the 3PCF estimator are the spatial distribution of the objects and the spatial distribution of a *random* set of unclustered objects distributed in the same volume and with the same selection function as the data. Since we are considering snapshots of an N-body simulation our triplet counting algorithm accounts for periodicity in the boundary conditions.

Input to the estimator is a catalog of synthetic objects randomly distributed in the same volume as the real sample and with the same selection criteria, i.e. the so-called random catalog. Comparing the triplets counts in the real and random catalog accounts for sample geometry and selection effects. To reduce the impact of this additional source of shot-noise error we use 20 times more objects in the random than in the real catalog. And to reduce the computational cost we adopt the random splitting technique (173): we repeat the 3PCF measurement of the same sample 20 time with a different random catalogue the same size of the data, and then we average the estimates. This also allows us to explicitly determine the error coming from this procedure.

6.2.2 Gaussian covariance

The final ingredient to compare models with simulated data is the 3PCF covariance matrix \mathbb{C} . Since a numerical estimate of \mathbb{C} is too computationally demanding, we use the Gaussian model of (154) to obtain a theoretical expression for the 3PCF covariance matrix \mathbb{T} . Thanks to the periodic boundary condition, we can further simplify the model expression and ignore mode coupling due to the sample geometry. We also consider the effect of binning in the computation. With these assumptions, we can obtain a simplified expression for the Gaussian 3PCF covariance matrix for each Legendre coefficient, $T_{\ell, \ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})$ and use it to

form the triangle-binned 3PCF covariance matrix:

$$T(r_{12}, r_{13}, r_{23}; r'_{12}, r'_{13}, r'_{23}) = \sum_{\ell, \ell'=0}^{\ell_{max}} T_{\ell, \ell'}(r_{12}, r_{13}; r'_{12}, r'_{13}) \mathcal{P}_{\ell}(\mu) \mathcal{P}_{\ell'}(\mu'), \quad (6.5)$$

where \mathcal{P}_{ℓ} are the Legendre polynomials of degree ℓ and μ is the cosine angle between r_{12} and r_{13} . The expression for $T_{\ell, \ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})$ and the details of each estimate can be found in (17). Its value is contributed by cosmic variance, which in turn depends on the volume of the sample V and on the matter power spectrum $P(k)$, and by the shot-noise of the discrete tracers with mean density \bar{n} . The values of V and \bar{n} are taken from the simulation box whereas we use one-loop SPT to model the power spectrum.

The use of the splitting method to estimate the 3PCF introduces an additional error, σ_r , on top of the shot-noise one. We model its contribution as an additional diagonal term to the theoretical covariance, i.e.

$$\mathbb{C} = \mathbb{T} + \sigma_r^2 \mathbb{I} \quad (6.6)$$

and use \mathbb{C} in the χ^2 analysis.

The accuracy of the Gaussian approximation for the 3PCF covariance matrix has been assessed by (17; 151). Based on their results, we expect that the Gaussian model provides an unbiased estimate of the covariance matrix for separations $r \geq 40 h^{-1} \text{Mpc}$ whereas it underestimates the uncertainties and their correlation on smaller scales. To correctly evaluate the error budget, we renormalise the covariance matrix \mathbb{C} by the factor $\alpha = 1.4$ to get a meaningful value of χ^2 around unity at sufficient large separations for the LO model.

$$\mathbb{C}' = \frac{\mathbb{C}}{\alpha} \quad (6.7)$$

We will use \mathbb{C}' throughout the rest of the paper. The impact of the treatment of uncertainties will be discussed in the framework of our χ^2 analysis, see Sec. 6.4.

6.3 Comparing models to data

To assess the goodness of the model, we perform a χ^2 comparison with the data. We estimate

$$\chi^2 = \sum_{i,j} (M_i - D_i) C_{ij}^{-1} (M_j - D_j) \quad (6.8)$$

where M_i is the model vector, D_i is the data vector and C_{ij} is the Gaussian covariance described in the previous section. The Gaussian hypothesis may fail on scales as small as those considered in our analysis. However, any inaccuracy introduced in the error estimate will affect all χ^2 analyses in a similar fashion and, therefore, will not bias the comparison between models.

We estimate the χ^2 statistics in the eigenspace, in order to diagonalize the covariance matrix, i.e.

$$\chi^2 = \sum_i \frac{(\Lambda_{ij}M_j - \Lambda_{ij}D_j)^2}{\lambda_i} \quad (6.9)$$

where Λ_{ij} are the coefficients of the eigen-matrix and λ_i are the positive eigenvalues. In our analysis, we use the reduced chi-square $\chi_r^2 = \chi^2/\nu$, where ν indicates the number of degrees of freedom.

To quantify the relative performance of models such as SPT-NLO and EFT-NLO we use the cumulative chi-square difference normalised by the degree of freedom of the SPT-NLO (i.e. the number of triangle configurations given the set Δr , r_{\min} and η)

$$\frac{\Delta\chi^2}{\nu_{\text{SPT-NLO}}} = \frac{\chi_{\text{SPT-NLO}}^2 - \chi_{\text{EFT-NLO}}^2}{\nu_{\text{SPT-NLO}}}. \quad (6.10)$$

We will also consider the residuals between models and data normalised to the statistical errors:

$$R_i = \frac{M_i - D_i}{\sigma_i} \quad (6.11)$$

where σ_i are the errors extracted from the diagonal elements of C_{ij} . The level of agreement between models and data is expected to vary with scale. To assess the relative goodness of the LO and NLO models and how it changes with scale we compute the average residuals on all triangle configurations and some selected ones, and compare them in the summary statistics

$$\langle \Delta R \rangle = \frac{1}{N} \sum_i^N |R_i^{\text{NLO}}| - |R_i^{\text{LO}}| \quad (6.12)$$

- where N is the number of configurations examined.

6.4 Results

We now compare the different 3PCF models introduced in Sec. 6.1 with the measurements performed on the DEMNUni simulations. In doing so, we distinguish, somewhat arbitrarily, between large ($> 40h^{-1}\text{Mpc}$) and small ($< 40h^{-1}\text{Mpc}$) separations to indicate the scales on which leading and next-to-leading 3PCF models are expected to provide matching and

different predictions, respectively. In addition, we will focus on triplets in which at least one of the sides matches the BAO scale of $\simeq 110h^{-1}\text{Mpc}$. We call these the BAO configurations.

6.4.1 Comparisons among 3PCF models

We now compare LO-2D and SPT-NLO 3PCF models with the LO-1D one. In Fig. 6.2, we show the difference between the LO-2D and LO-1D 3PCFs, $\Delta\zeta$, as a function of the various triplets configuration, labelled and ordered as described in Sec. 6.1. The top and bottom panels show the results obtained for $z = 1.05$ and $z = 0.49$, respectively. In both cases, the magnitude of the difference is small compared to the expected $1\text{-}\sigma$ Gaussian error represented by the grey band. The largest discrepancies, the spikes in the blue curve, are found for the isosceles configurations $\eta_{\min} = 0$. The angular sampling needed to compute bispectrum multipoles in Eq. 5.26 represents the source of the peaks at quasi-isosceles nonlinear configurations (see Fig. 5.2, and Fig. 5.3). The amplitude of these peaks increases when decreasing the size of the triangles but never exceeds the Gaussian uncertainty.

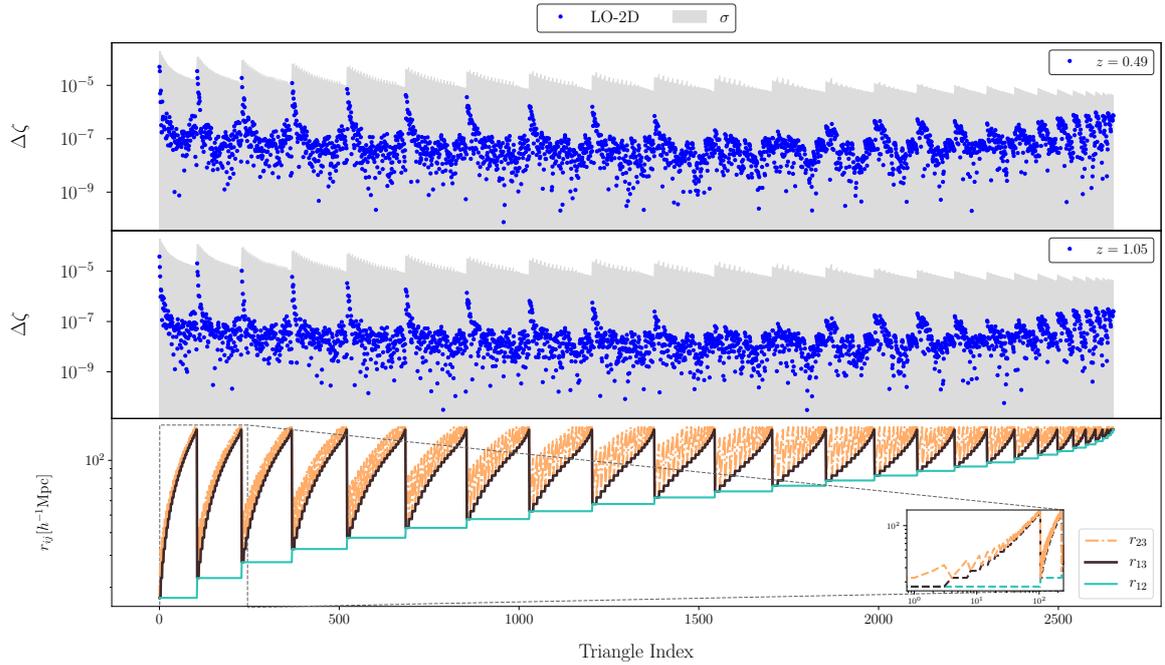


Figure 6.1 Difference between LO-1D, considered as a benchmark, and LO-2D 3PCF models, $\Delta\zeta = \zeta^{\text{LO-2D}} - \zeta^{\text{LO-1D}}$ (blue curve), as a function of the triangle configurations identified by the side lengths. Top and central panels show the model predictions at $z = 1.05$ and at $z = 0.49$. The grey band represents the $1\text{-}\sigma$ Gaussian uncertainty. The bottom panel shows the different sides of the triangles as a function of the Triangle Index, using logarithmic y-axis. The boxed plot represents, using logarithmic x- and y- axes, a zoomed-in area of the bottom panel including only separation distances $r_{12} \leq 17.5h^{-1}\text{Mpc}$, providing a closer look at the region of very small scales.

We conclude that our implementation, or the LO-2D model, which uses the 2D-FFTLog tool, fully agrees with the standard one, LO-1D, used as a benchmark. These results validate the 2D-FFTLog implementation and justify the adoption of this tool to generate a nonlinear 3PCF model.

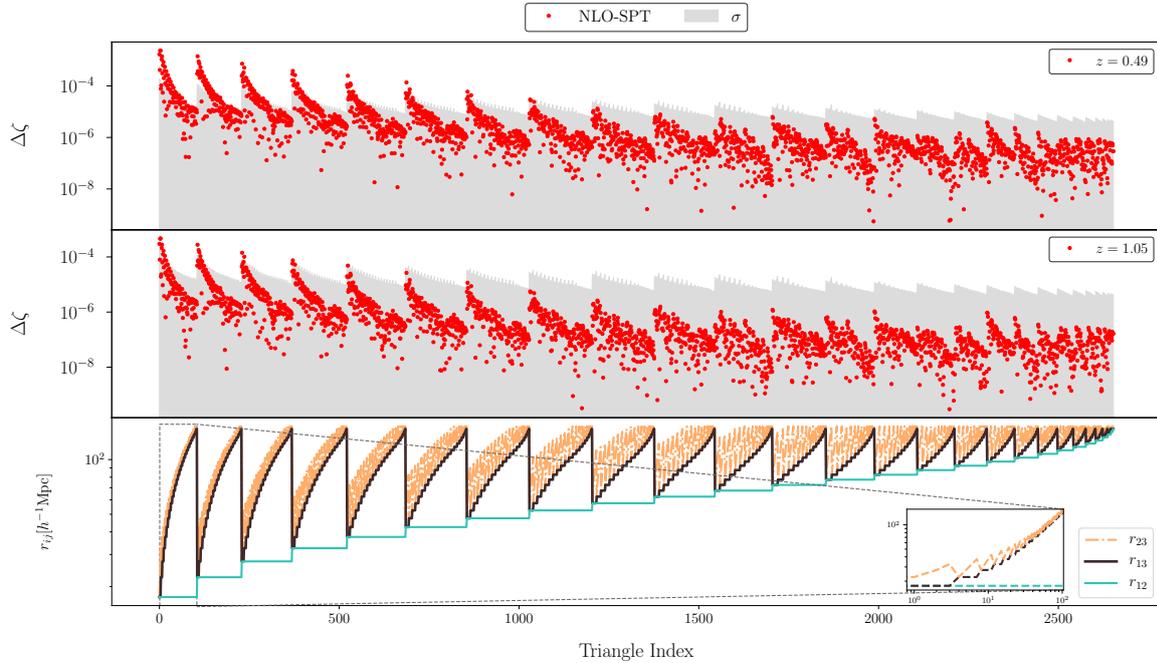


Figure 6.2 Same as figure 6.1 showing the difference between 3PCF models LO-1D and NLO-SPT (red curves).

Fig. 6.2 compares the difference between the SPT-NLO and the LO-1D 3PCF models (red points) to the expected $1\text{-}\sigma$ Gaussian uncertainty (grey band). On large scales (i.e. to the right part of the plots), the differences are about ten times larger than in the LO-1D vs LO-2D case but still well below the expected Gaussian uncertainty, showing that nonlinear effects are small in this regime. Differences increase when moving to the left of the panels, i.e. on small scales, as expected. Differences between the models increase by up to two orders of magnitudes, indicating the importance of nonlinear contributions to the 3PCF signal.

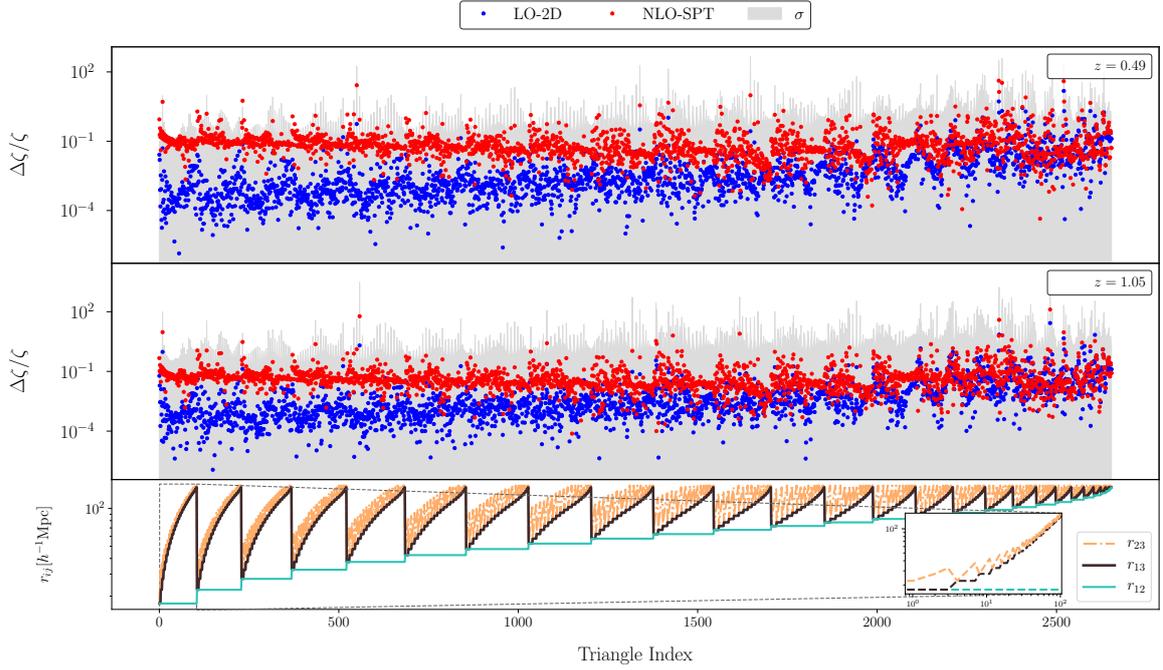


Figure 6.3 Relative difference between LO-1D, considered as a benchmark, and LO-2D 3PCF models, $\Delta\zeta = (\zeta^{\text{LO-2D}} - \zeta^{\text{LO-1D}})/\zeta^{\text{LO-1D}}$ (blue curve), as a function of the triangle configurations identified by the side lengths. Top and central panels show the relative errors between model predictions at $z = 1.05$ and at $z = 0.49$. The grey band represents the $1\text{-}\sigma$ propagated Gaussian uncertainty. The bottom panel shows the different sides of the triangles as a function of the Triangle Index, using logarithmic y-axis. The boxed plot represents, using logarithmic x- and y- axes, a zoomed-in area of the bottom panel including only separation distances $r_{12} \leq 17.5h^{-1}\text{Mpc}$, providing a closer look at the region of very small scales.

Moreover, the magnitude of the discrepancy increases at lower redshift for the same reason. Superimposed on this trend, we still see peaks in correspondence of isosceles triangle configurations, confirming not only that the different techniques that reproduce the same model exhibit most differences on those configurations but also that differences between 3PCF models at different perturbative orders peak on small scales toward $\eta_{\min} = 0$.

We conclude that the SPT-NLO model significantly differs from the LO-1D on small scales. Which, however, does not guarantee that SPT-NLO is a better model. A point that we will address by comparing models with data in the next Sections. These observations are confirmed by visually inspecting Fig. 6.3, which illustrates the relative error between the examined models (NLO-SPT and LO-2D) and the benchmark model (LO-1D). The relative differences in the LO-2D model demonstrate a decreasing trend as the minimum scale decreases. This indicates that the numerical approximation error inherent in the 2D-FFTLog algorithm, applied with the given input grid, impacts the model's predictions more prominently at larger scales, where the signal-to-noise ratio is lower. On the other hand,

the NLO-SPT model exhibits greater per-cent differences on smaller scales, as expected. It is worth noting that the noise introduced on large scales by the 2D-FFTLog algorithm is model-independent.

The fact that the differences between models are larger than statistical errors for small triangle configurations indicates that model nonlinearities would dominate the error budget on small scales. And clearly illustrates the importance of going beyond the linear order 3PCF model.

6.4.2 3PCF measurements

We performed 3PCF measurements on the DEMNUni simulation snapshots $z = 0.49$ and $z = 1.05$ using the estimators presented in Sec. 6.2.1 over a wide range of scales: $r_{ij} = [17.5, 152.5] h^{-1} \text{Mpc}$ using the bin-width $\Delta r = 5 h^{-1} \text{Mpc}$. Every symbol in Fig. 6.4 represents a 3PCF measurement for triangles of all sizes, labelled in the X-axis, in the allowed range. From top to bottom, the first and third plots show the 3PCF amplitude multiplied by the triangle side lengths. The second and fourth plot shows the signal to noise, i.e. the 3PCF amplitude in units of the Gaussian error: ζ/σ . Different symbols (and colours) are used for different triplet types. Green dots indicate small-scale triplets, defined as configurations for which $r_{12} \leq 40 h^{-1} \text{Mpc}$. Red triangles identify BAO triplets that encompass the BAO scale, the latter defined as the scale in which BAOs provide their typical wiggling features to the correlation functions. Practically, for BAO configurations, we consider triangle configurations with at least one side in the range $[17.5 - 117.5 h^{-1} \text{Mpc}]$. Yellow stars indicate the squeezed BAO configurations, i.e. triplets with one side much smaller than the other two, whose at least one is in the BAO regime. The grey dot symbol is used for all other cases. The signal to noise is typically above unity, quite insensitive to the triangle size except on small scales where it peaks.

6.4.3 3PCF models vs. data: overview

We compare the SPT-NLO and the LO-2D 3PCF models to the same set of measurements performed on the DEMNUni datasets. The scatterplot in Fig. 6.5 compares the absolute values of normalised residuals in Eq. (6.12) of both models for each triangle configuration for the SPT-NLO (Y-axis) and LO-2D (X-axis) models in the two snapshots. We use the same symbols and colors as in Fig. 6.4 to identify different types of triangle configurations. In Fig. 6.5, we show the normalised absolute value of residuals of SPT-NLO on the Y-axis and LO-2D on the X-axis for both redshift cases. In the left and right panels, respectively $z = 0.49$ and $z = 1.05$, different symbols and colours are used to represent the different

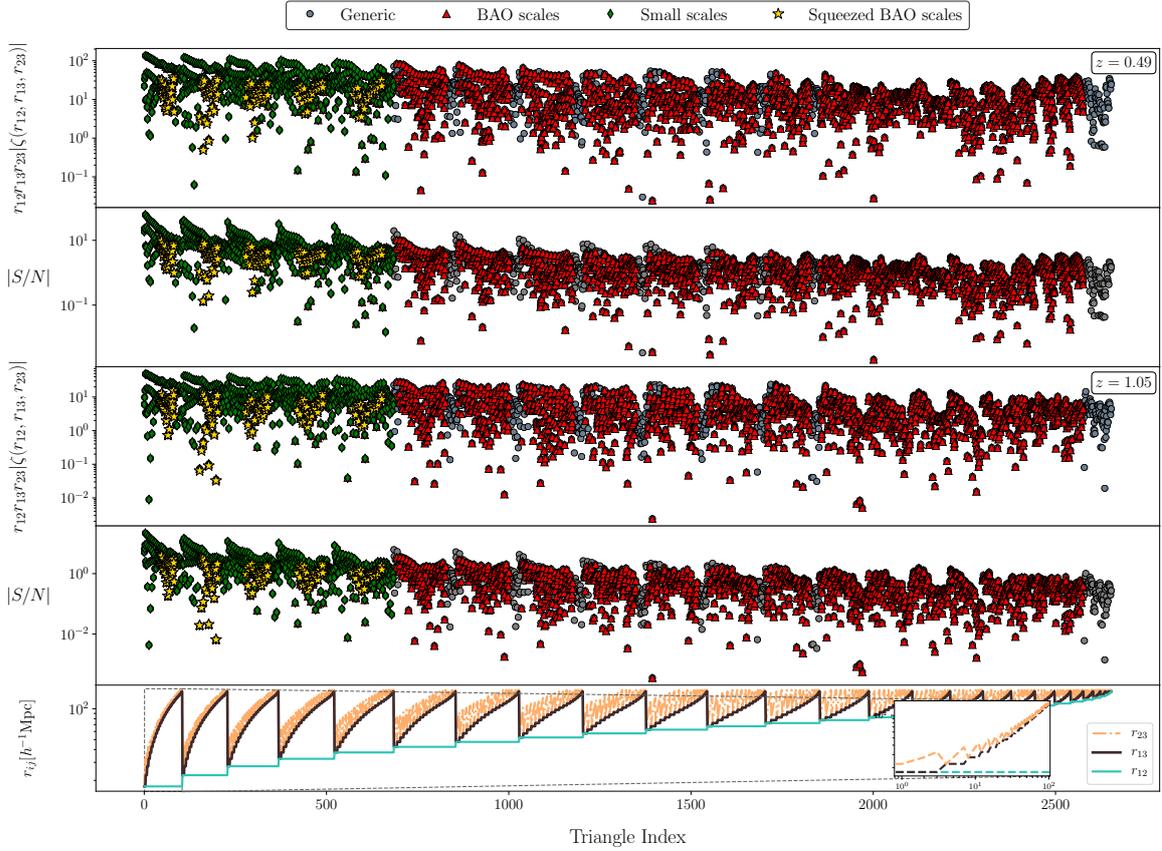


Figure 6.4 Measurements of 3PCF from the DEMNUni snapshots at $z = 0.49$ (top plots) and $z = 1.05$ (bottom plots). In each figure for different redshifts, the upper panel shows the 3PCF amplitude multiplied by the size of the sides of the triangles as a function of the triangle configuration, indicated on the X-axis. The bottom panel shows the signal-to-noise, assuming Gaussian errors. Different colours and symbols are used for different triangle types, as indicated in the labels. In our classification scheme, as described in the text, it is possible for a triangle to have multiple classifications. For instance, a triangle classified as squeezed BAO would also fall into the category of small-scale triangles. First and second panels refer to $z = 0.49$, while third and fourth panels $z = 1.05$. Bottom plot shows the sides of the triangles as a function of the Triangle Index, ranging from $r_{\min} = 17.5h^{-1}\text{Mpc}$ and $r_{\max} = 132.5h^{-1}\text{Mpc}$. The bin-width is $\Delta r = 5h^{-1}\text{Mpc}$.

configurations shown in Fig. 6.4. Deviations under the diagonal indicate the SPT-NLO model behaves better with respect to the LO-2D model. Deviations above the diagonal indicate the opposite case.

We stress that a clear and expected feature is the redshift dependence: the matching between model and data improves with the redshift for both the SPT-NLO and the LO-2D cases. At $z = 1.05$ only for a handful of small triangle configurations, the difference between the model and the measured 3PCF exceeds $3\text{-}\sigma$ significance. And these are mostly LO-2D predictions. At $z = 0.49$, the number of $3\text{-}\sigma$ outliers increases significantly, and,

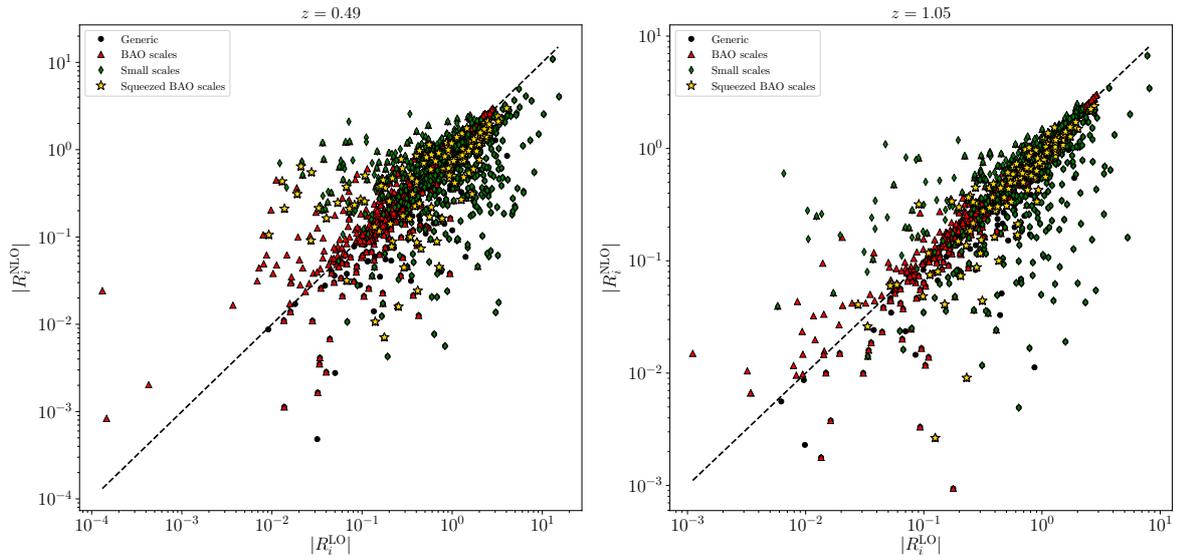


Figure 6.5 Comparison between normalised absolute values of residuals of SPT-NLO and LO-2D models. The LO-2D residuals are displaced in the X-axis, while the SPT-NLO case is displaced in the Y-axis. The left plot refers to $z = 0.49$, while the right plot refers to $z = 1.05$. Different symbols focus on BAO scales (red triangles), small scales (green rhombuses) and small BAO scales (yellow stars).

more generally, the amplitude of the data-model mismatch systematically increases on all scales. We also represented different sub-sets of configurations with different colours and symbols: small scales, BAO configurations, and squeezed BAO scales. To quantify the relative agreement between the LO-2D and SPT-NLO models, we report the average difference between the absolute value of normalised residuals (see Eq. 6.12) - for each considered sub-set - in Tab. 6.2. Negative values of the average difference of the absolute values of normalised residuals indicate the SPT-NLO's predictions are closer to the simulated dataset with respect to the LO-2D's case; positive values indicate the opposite. The whole set of measured triangle configurations shows negative values of $\langle \Delta R \rangle$. Small scales indicate an even more negative value of that quantity, showing the SPT-NLO model exhibits a smaller difference with the data compared to the LO-2D case. Similarly, for the BAO scales, for which the cited difference assumes a slightly negative value despite being consistent with zero. As a preliminary overview, these cases indicate the contribution of the SPT-NLO model better matches the simulated dataset. Considering the squeezed BAO scales, $\langle \Delta R \rangle$ is negative, and it represents the most significant deviation from the zero value considering the error, meaning the SPT-NLO model improves the match with the dataset concerning the LO-2D model. At this stage, we did not consider the correlation between measurements, ignoring the effect of the non-diagonal terms in the covariance matrix. We will properly

quantify the performance of models in Sec 6.4.5. We will focus on the particular squeezed isosceles configurations in the next Sec.

Configurations	$z = 0.49$	$z = 1.05$
All configurations	$\langle \Delta R \rangle = -0.129 \pm 0.038$	$\langle \Delta R \rangle = -0.059 \pm 0.039$
BAO scales	$\langle \Delta R \rangle = -0.009 \pm 0.045$	$\langle \Delta R \rangle = -0.006 \pm 0.045$
Small scales	$\langle \Delta R \rangle = -0.431 \pm 0.076$	$\langle \Delta R \rangle = -0.188 \pm 0.076$
Squeezed BAO	$\langle \Delta R \rangle = -0.262 \pm 0.163$	$\langle \Delta R \rangle = -0.132 \pm 0.163$

Table 6.2 Average difference between the absolute values of normalised residuals of SPT-NLO and LO-2D models. Redshifts are listed in columns two to three, while all configurations, BAO, small scales and squeezed BAO scales are listed in rows two to five.

6.4.4 3PCF models vs. data: the BAO scale

A key feature in clustering statistics is represented by the imprint of the baryons, the large-scale imprinting of primordial baryonic fluctuations that can be used to trace the expansion history of the Universe with per cent precision. BAOs in two-point statistics have become a standard and very effective cosmological probe. However, the BAO feature can be and has been, detected in the galaxy 3-point correlation function, too (72; 174). As for the 2-point statistics, its use for precision cosmology depends on the ability to model nonlinear effects, i.e. to go beyond linear order-based models. Indeed, correct modelling of the BAO is important in full shape fits, as throughout, we are able to get cosmological information.

In this Sec., we, therefore, focus on the BAO scale, and more precisely on squeezed isosceles triangle configurations encompassing the BAO scale, to assess the performance of the NLO-SPT and NLO-EFT models in comparison to LO's predictions.

In Fig.6.6, we compare the measured and the predicted 3PCFs for isosceles triangles with one side length fixed to $\delta r = 22.5 h^{-1} \text{Mpc}$, well within the nonlinear regime, and the other two equal side of increasing length r , shown on the X-axis. 3PCF estimated are represented by the black dots with their 1σ -Gaussian error bars. Models' predictions are shown by the two curves: dashed blue for the LO-2D case, continuous red for the NLO-SPT and dashed-dot green for the NLO-EFT. At low redshift (left panel), the NLO-SPT and NLO-EFT models outperform LO-2D on all scales. In particular, they reproduce the shape of the BAO peak much better, indicating that the nonlinear effects responsible for the widening

Redshift	$c_0 [h^{-2}\text{Mpc}^2]$	$c_1 [h^{-2}\text{Mpc}^2]$	$c_2 [h^{-2}\text{Mpc}^2]$	$c_3 [h^{-2}\text{Mpc}^2]$
$z = 0.49$	2.25 ± 1.68	-4.20 ± 1.73	-1.91 ± 2.37	1.65 ± 3.25
$z = 1.05$	3.68 ± 1.90	-4.81 ± 1.96	-7.38 ± 2.68	11.45 ± 3.65

Table 6.3 Best fit values for the four parameters of the NLO-EFT model for the squeezed BAO configurations, as depicted in Fig. 6.6.

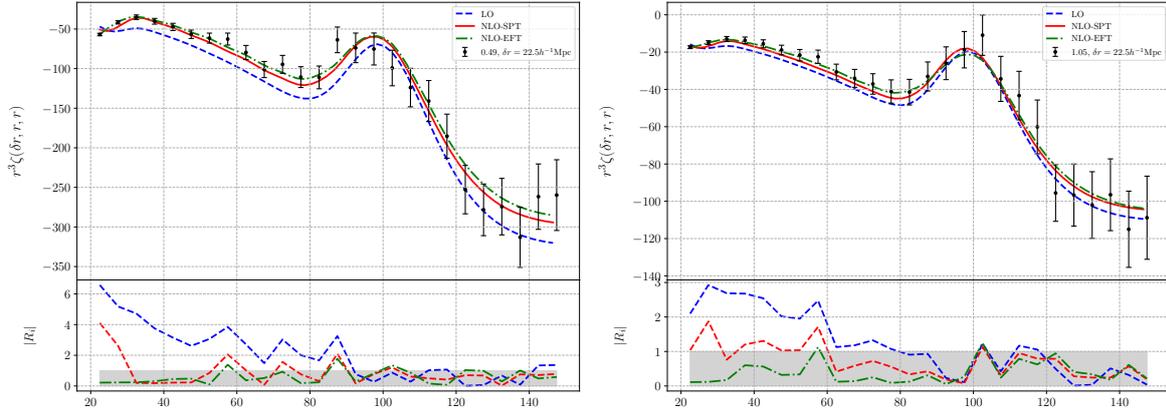


Figure 6.6 *Top*: 3PCF as a function of the two side lengths r of a set of isosceles triangles, with the third side of length $\delta r = 22.5h^{-1}\text{Mpc}$. The 3PCF measurements are shown as black dots, and the error bars indicate the 1σ Gaussian uncertainty. The continuous red curve and the dashed blue ones show the predictions of the NLO-SPT of the LO-2D models, respectively. Left and right panels illustrate the results at $z = 0.49$ and at $z = 1.05$. *Bottom*: normalised residuals for $z = 0.49$ (left), $z = 1.05$ (right), as defined in Eq. 6.11, for the three models depicted by the same colours and shapes of the top panel.

of the peak are correctly accounted for. At larger redshifts, the three models behave almost identically - particularly on larger sides -, which does not come as a surprise. In regard to the comparison between the NLO-EFT model and the NLO-SPT model, incorporating four parameters leads to a significant improvement in data matching, particularly at small scales, as shown by the normalised residual depicted in the bottom panel of Fig. 6.6. We notice that at this redshift and on these scales, the 3PCF signal is very small (notice the difference in the Y-axis scale) and highly correlated, which explains why, as shown in the next Sec., the statistical significance of the mismatch as quantified by the χ^2 analysis is rather small.

6.4.5 3PCF models vs. data: quantitative analysis

We now present the results of the χ^2 analysis introduced in Section 6.3. In addition to LO-2D and NLO-SPT, we also consider the nonlinear NLO-EFT. In these analyses, we have considered all triangles with side length r_{12} in the range $[17.5, 132.5] h^{-1}\text{Mpc}$, splitting our analysis in two cases: $\eta_{\min} = 2$ and $\eta_{\min} = 0$.

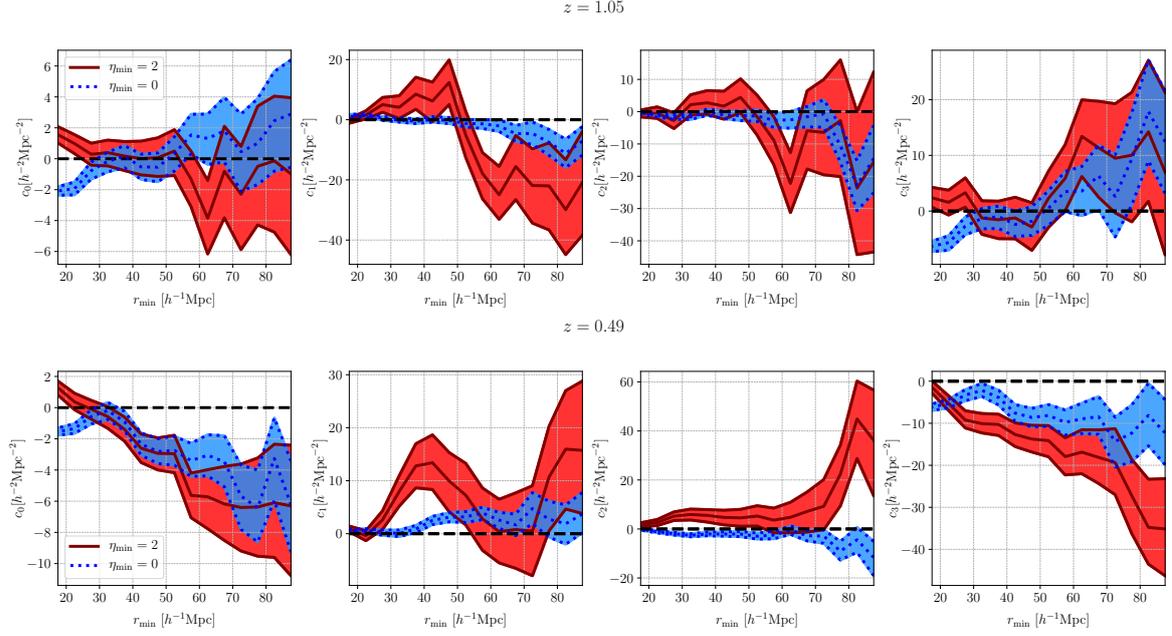


Figure 6.7 Best fit values for the four NLO-EFT free parameters (left to right) from the reduced χ^2 analysis, as a function of r_{\min} . The red continuous band refers to $\eta_{\min} = 2$, the blue dotted band to $\eta_{\min} = 0$. Top and bottom panels show the results for the $z = 0.49$ and $z = 1.05$ snapshots, respectively.

The reason why we ran the analysis with different values of η is motivated by the measurement technique adopted. Indeed, we measure the 3PCF multipoles first, up to $\ell_{\max} = 10$, and then combine them. This is not a problem in principle as long as we repeat the same operation consistently in the model. The complication arises for those combinations $\{r_{12}, r_{13}\}$ that allows the third side r_{23} to span from very large to very small, highly nonlinear scales. For such cases, corresponding to the choice of $\eta = 0, 1$, the multipole series is slowly convergent due to the steep shape of the 3PCF as $r_{23} \rightarrow 0$. All the resummed 3PCFs coming from these multipoles are then dominated by the contribution of these squeezed configurations, which are extremely difficult to properly model. In principle, a standard estimator of the 3PCF would allow to filter out these contributions more efficiently. In order to compare EFT-NLO predictions with the ones from other models, we inferred the four model parameters for both cases of choice of η_{\min} and redshifts as a fixing the

maximum scale to $r_{\max} = 122.5$ and varying the minimum scale of the fit r_{\min} in the interval $[17.5, 62.5] h^{-1} \text{Mpc}$, i.e. the range from mild to strong nonlinear scales in which EFT contributions are supposed to be relevant. The results are shown in Fig. 6.7. The $\eta_{\min} = 0$ case, due to the larger number of triangle configurations, exhibits tighter bands with respect to the $\eta_{\min} = 2$ case. This plot is to be interpreted cumulatively from right to left: on large scales where the nonlinear effects are small, the best-fit values are consistent with zero within the error bars. Moving towards smaller scales and considering an increasingly large number of progressively smaller triangles, the parameters of the best-fitting models significantly depart from zero at both redshifts and η_{\min} choice. The estimation of c_2 benefits from the larger number of configurations in the case $\eta_{\min} = 0$. Using the inferred EFT parameters at each r_{\min} for the EFT-NLO model, we present the performance of the goodness-of-fit of all the models listed in Tab. 6.1. The results are shown in Fig. 6.8 in the form of cumulative χ^2 : from right to left, each point of the curves indicates the reduced χ^2 value obtained by including all triangles with $\eta_{\min} = 2$ and r_{12} in the range $[r_{\min}, 132.5] h^{-1} \text{Mpc}$, where the r_{\min} value is to be read on the X-axis.

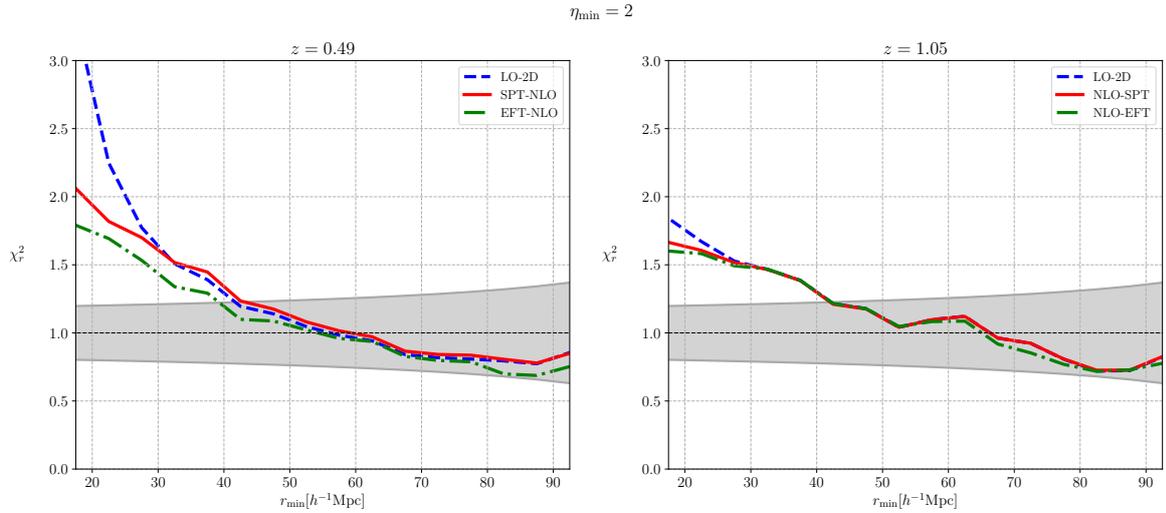


Figure 6.8 Cumulative reduced chi-square χ_r^2 in the range $[r_{\min}, 112.5] h^{-1} \text{Mpc}$, as a function of r_{\min} for the case $\eta_{\min} = 2$. continuous red and dashed blue curves refer, respectively, to the NLO-SPT and LO-2D models. The grey band represents a 99.7% confidence level assuming Gaussian statistics. The left and the right panels refer, respectively, to redshift $z = 0.49$ and $z = 1.05$

The leftmost point of the curves indicates the χ^2 value obtained when considering all triangle configurations with $\eta \geq 2$. The cumulative χ_r^2 value of each model can be compared with the corresponding 99.7% confidence interval for Gaussian statistics, represented by the grey band. The reduced χ^2 is well within the grey strip for most of the scales above $40 h^{-1} \text{Mpc}$, indicating those model uncertainties provide a minor contribution to the 3PCF

total error budget. For this purpose, we normalised the covariance matrix in Eq. 6.7 to have a meaningful goodness-of-fit around unity on large scales in which models are expected to work well. This is not the case for scales below $40h^{-1}\text{Mpc}$, where nonlinear effects are large and cannot be fully captured by perturbative models. We observe that for scales below $30h^{-1}\text{Mpc}$ at both redshifts, the difference between the NLOs models and the LO-2D model is appreciable, particularly for $z = 0.49$, but limited to scales in which uncertainties are known to be underestimated, and the goodness-of-fit is far from being close to unity. These are also the scales, however, on which the SPT-NLO and EFT-NLO provide a better fit to the data than the LO-2D one. The small difference between the models is slightly deceptive, and partly due to the fact that having considered cumulative statistics, the χ_r^2 is significantly contributed by those large triangles in which the two models provide similar (good) predictions. The fact that LO and NLO models perform similarly in the $\eta_{\min} = 2$ case reflects the fact that these are triangle configurations for which the number of scales in the small-scale regime is limited by the constraint on the sides of the triangle.

Therefore we decided to explore the more challenging case of $\eta_{\min} = 0$ for which the number of triangle sides in the small scale range is as large as possible, ranging in $[r_{\min}, 132.5h^{-1}\text{Mpc}]$.

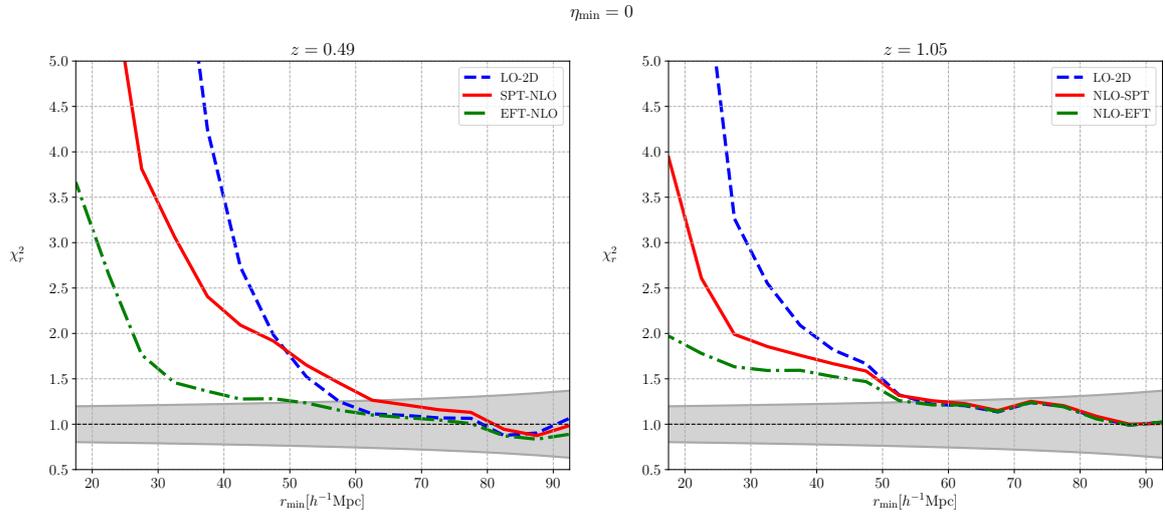


Figure 6.9 Same as Fig. 6.8 fore the case $\eta_{\min} = 0$. Model predictions for NLO-EFT, NLO-SPT and LO-2D are shown with green dot-dashed, reds continuous and blue dashed curves, respectively.

Fig. 6.9, in which cumulative χ_r^2 of all models are represented, shows that while the goodness of the fit for both next-to-leading models, SPT-NLO and EFT-NLO, have significantly worsened with respect to the $\eta_{\min} = 2$ configuration case. And yet they both outperform the LO-2D model on small scales at both redshifts. Here, the improvement coming from the SPT-NLO is more significant and, as in the previous case at $\eta_{\min} = 2$, bounded at scales below

$60h^{-1}\text{Mpc}$, and where goodness-of-fit is far from being close to the expected value. This difference is probably less significant than it appears since model uncertainties are probably underestimated by the Gaussian assumption. The comparison between these different models informs us that the use of nonlinear models is mandatory to analyse the 3PCF signal from $\eta_{\min} < 2$ triangle configurations. And that even on small scales, where the quality of the fit degrades, the systematic errors they introduce are significantly smaller than bias of the LO models.

Although the goodness of the EFT-NLO fit is bound to be superior to the SPT-NLO one, the corresponding reduced χ^2 values are quite similar in the two cases.

In order to quantify the improvement coming from the adoption of the EFT-NLO model with respect to SPT-NLO, we studied the cumulative chi-square difference normalised by SPT-NLO degrees of freedom as presented in Eq. 6.10. Results are shown in Fig. 6.10.

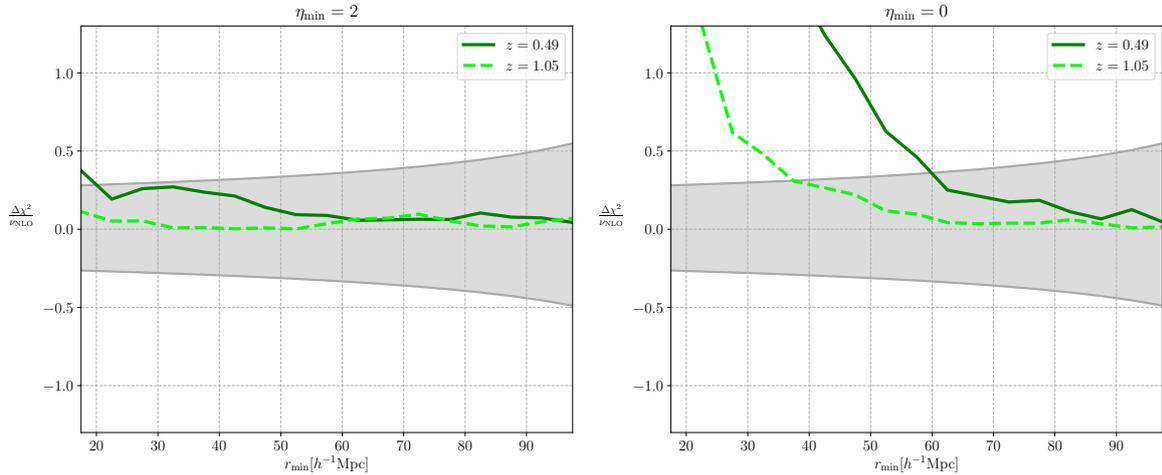


Figure 6.10 Cumulative chi-square difference between NLO-SPT and NLO-EFT models divided by the NLO-SPT degree of freedom. Dotted line refers to $z = 1.05$, the solid green one refers to $z = 0.49$. From left to right, respectively, the case $\eta_{\min} = 2$ and $\eta_{\min} = 0$. The grey band represents a 99.7% confidence level coming from the propagation of errors on Eq. 6.10.

Differences above the grey band mean the improvement coming from NLO-EFT is significant. None of the differences is below the grey band, meaning the model benefits or - at least - is not worsened by the adoption of four extra fitting parameters, as expected. The case $\eta_{\min} = 2$ shows the NLO-EFT models show an almost non-significant improvement over the NLO-SPT model, at both redshifts. On the other hand, the case $\eta_{\min} = 0$ shows the improvement coming from NLO-EFT model is significant in the small-scale regime, confined to scales below $60, 40h^{-1}\text{Mpc}$, respectively at $z = 0.49, 1.05$. However, the improvement holds in the regime in which Gaussian uncertainties are supposed to be underestimated (151) so that the goodness of fit is far from being close to the expected value in this regime. What

is surprising is the contribution from the adoption of EFT counterterms in a configuration space analysis as we focused on this work, and in a Fourier space analysis as done in (60). Bispectrum analyses at next-to-leading order have shown that EFT models are crucial to model the matter signal up to $k = 0.16 - 0.19 \text{Mpc}^{-1}h$ depending on the binning of the data, being a significant improvement with respect to the SPT model (see Fig. 7 of (60)). In our analysis, the improvement of the EFT-NLO model compared to SPT-NLO is much less significant, although the quantification of the improvement is affected by the estimation of the uncertainties that differ in both analyses. This seems to reflect that departures from the SPT model parametrised by adding four extra EFT terms in Fourier space are converted into very small-scale contributions in configuration space, for which both models fail to give a clear representation of the data.

To summarize our results we list, in Tab. 6.4 and 6.5, the signal-to-noise values obtained by comparing the model 3PCF to the data (column 5) at 1) $r_{\min} = 40h^{-1}\text{Mpc}$ for $\eta_{\min} = 0$ and 2) $r_{\min} = 20h^{-1}\text{Mpc}$ for $\eta_{\min} = 2$. For both choices, the NLOs models exhibit smaller χ^2 values and therefore better fit, meaning in the small scale range, the adoption of NLOs models is required to model the dataset properly. Comparing both choices, smaller χ^2 corresponds to smaller S/N values, even if choice 2) considers smaller minimum scales than 1). We stress the agreement between models and data, for all the models considered in this work in the nonlinear regime, is affected by the estimation of uncertainties (151).

Model	Redsfhit	χ_r^2	S/N
LO-2D	0.49	2.73	19.58
EFT/SPT-NLO	0.49	2.09	19.58
LO-2D	1.05	1.82	13.07
EFT/SPT-NLO	1.05	1.66	13.07

Table 6.4 Signal-to-noise [S/N] at $r_{\min} = 40h^{-1}\text{Mpc}$ and $\eta_{\min} = 0$ and the χ^2 comparison between the model 3PCFs considered in this work, and listed in column 1, and the measurements performed on two snapshots of the DEMNUni simulations at the redshifts listed in column 2.

Focusing on the BAO scales, the differences between the leading and next-to-leading order models are also significant. The comparison, described in the previous Section, considers a wide range of scales $r = [22.5, 152.5]h^{-1}\text{Mpc}$ and it is summarised in Tab 6.6. Especially at $z = 0.49$, for the configuration we used as a test, the chi-square value is high, but, along with the case at $z = 1.05$, it is shown that the SPT-NLO model provides a significant improvement on squeezed BAO configuration. We stress the significance of the mismatch between models

Model	Redshift	χ_r^2	S/N
LO-2D	0.49	2.24	9.55
EFT/SPT-NLO	0.49	2.06	9.55
LO-2D	1.05	1.67	12.05
EFT/SPT-NLO	1.05	1.61	12.05

Table 6.5 Same as Tab. 6.4, but $r_{\min} = 20h^{-1}\text{Mpc}$ and $\eta_{\min} = 2$

and data depends on the adopted model, and it is affected by having adopted a Gaussian model for the errors and their covariance.

Model	Redshift	χ_r^2
LO-2D	0.49	27.6
NLO-SPT	0.49	11.9
NLO-EFT	0.49	2.9
LO-2D	1.05	4.4
NLO-SPT	1.05	2.3
NLO-EFT	1.05	1.0

Table 6.6 Reduced χ^2 values obtained when comparing measured and model 3PCF for triangle configurations that encompass the BAO scale (column 3). The models are listed in column 1, and the redshift of the simulation snapshot is shown in column 2. The analysis is described in Sec. 6.4.4.

7

Constraining galaxy bias with the Euclid Spectroscopic Survey

In this Chapter, I illustrate an application of the 3PCF model focusing on its fundamental role in breaking the degeneracy among cosmological parameters that affect clustering analyses based on the use of two-point clustering statistics only. For that, I need to move from the 3PCF model for the matter presented in the previous chapter to a 3PCF model for the galaxies since these are the objects that can actually be observed. Galaxies' positions are typically determined from their measured redshift without accounting for their peculiar velocities. As a result, anisotropic distortions are induced in the 3D mapping of their spatial positions that, however, I will ignore in this chapter and throughout its thesis, limiting my analysis to the real-space case, i.e. assuming that redshift distortions have been somehow removed.

Modelling clustering statistics for galaxies rather than for matter means explicitly accounting for the galaxy bias, i.e. the mapping between their two density fields. Since my goal is to include scales where linear theory does not apply anymore, I will account for nonlinear effects both in the 2PCF model and in the biasing relation. But assume linear order to model 3PCF. The new, next-to-leading order, 3PCF model will be presented in the next Chapter.

7.1 Modeling galaxy bias and clustering statistics

In the analysis presented in this Chapter, I will use the next-to-leading order 2PCF model described in Sec. 5.2. The method I used to implement the model relies on performing the inverse Fourier transform of the matter power spectrum through the 1D-FFTLog algorithm. The method is used to compute the loop terms involved in the power spectrum computation at the next-to-leading order, following (146). This approach represents an efficient and accurate technique to mitigate the computational cost of computing the two-point correlation function model. Having fixed all cosmological parameters, the galaxy two-point correlation function depends on five free parameters to describe the galaxy bias:

$$\{b_1, b_2, \gamma_2, \gamma_{21}, c_0\} \quad (7.1)$$

The first four parameters derive from the adoption of Standard Perturbation Theory formulation (SPT) (see Secs 3.3.2, 4.2.3), while the last one, c_0 , is the counterterm included in the Effective Field Theory formulation in Sec. 4.2.2.

For the galaxy 3PCF, I use the leading order model described in Sec. 5.3. This model is the equivalent to that of (53) but is implemented using a more general 2D-FFTLog procedure, which was previously validated in the context of clustering of dark matter perturbations in the chapter 6. The 3PCF model has three free bias parameters:

$$\{b_1, b_2, \gamma_2\} \quad (7.2)$$

In this analysis, I do not consider the EFT model for the 3PCF. Instead, I rely on 2PCF to constrain the c_0 ETF parameter. The bias model adopted in this Chapter and its parameters have been presented and discussed in Sec. 3.3.2, whereas the parameters of the EFT model have been introduced in Sec. 4.2.2.

7.2 Dataset

The prediction of the galaxy 2PCF and 3PCF models have been compared with the measurements of these statistics performed in *Euclid* Flagship mock galaxy catalogue. This catalogue was generated from a high-resolution N-body simulation of two trillion dark matter particles in a periodic box of $L = 3780 h^{-1}\text{Mpc}$ per side using the PKDGRAV3 algorithm (175). The cosmological parameters used for the simulation were chosen to match those of a flat Λ -cold dark matter (Λ -CDM) cosmology as determined by the *Planck* mission (3), with $\Omega_m = 0.319$, $\Omega_b = 0.049$, $\Omega_{\text{de}} = 0.681$, $\sigma_8 = 0.83$, $n_s = 0.96$ and $h = 0.67$. The simulation outputs were combined to build a light cone within an opening angle covering $1/8^{\text{th}}$ of the full sky. Dark matter halos were identified using the Rockstar halo finder, and galaxies were assigned to the halos using the halo occupation distribution model, whose parameters were set to match the number of galaxies that *Euclid* is supposed to detect at different redshifts. I focused on four redshift snapshots in real space in the range that will be covered by *Euclid*. Particularly, two- and three- point correlation functions models will be tested at redshift $z = 1.79, 1.53, 1.19, 0.90$. These latter, along with the number of galaxies N_g and their number density \bar{n} , are listed in Tab. 7.1.

Redshift	N_g	\bar{n} [10^{-3}]
0.9	110321755	2
1.2	55563490	1
1.5	31613213	0.6
1.8	16926864	0.2

Table 7.1 List of, for each redshift, the number of galaxies N_g and the mean number density \bar{n} in the Flagship galaxy catalogues.

7.2.1 Measurements

The isotropic two-point correlation function has been measured with the simple estimator (153; 176).

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \quad (7.3)$$

where DD and RR are the data-data and random-random pairs of objects, respectively. Since the number density of objects in the box is constant, then the number of random-random counts as a function of the separation has been estimated analytically.

The isotropic three-point correlation function has been measured through the estimator presented in Sec. 5.4 and used in the context of matter perturbation in Sec. 6.2.1. Measurements have been performed considering all configurations for which $\eta \geq \eta_{\min} = 1$. Furthermore, the triplet counting algorithm accounts for the periodicity of the box, and for each snapshot, a random dataset containing $N_R(z)$ times more objects than galaxies has been used to reduce the noise. The values of $N_R(z)$ are listed in Tab. 7.2. This random sample is needed to consistently subtract the disconnected part of the 3PCF.

Redshift	$N_R(z)$
$z = 1.78$	50
$z = 1.53$	50
$z = 1.19$	50
$z = 0.90$	25

Table 7.2 The list of ratios N_R between the random and data datasets

To reduce the computational effort, the random splitting technique has been used (17; 173), repeating the measurement of the same given sample 25 times using different random catalogues and averaging the estimates.

7.2.2 Gaussian covariance

To estimate errors in the measured 2PCF and their covariance, I rely on the Gaussian covariance model \mathbb{C}_ξ defined in Eq. 5.42. To estimate \mathbb{C}_ξ I have used the next-to-leading order power spectrum model for the matter and set the volume and the number density equal to those of the simulations snapshots.

The covariance matrix for the measured 3PCF, \mathbb{C}_ζ , also relies on the Gaussian model proposed in (154). The periodic boundary conditions used to run the parent N-body simulation allow us to use a simplified expression and ignore any mode coupling caused by the sample geometry. I also considered the effect of binning in the computation. Putting it all together I obtained the Gaussian 3PCF covariance matrix for each Legendre coefficient, $T_{\ell,\ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})$, which is then used to form the triangle-binned 3PCF covariance matrix. The expression for $T_{\ell,\ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})$ is given in Eq. 5.45 and details can be found in (17). The 3PCF covariance of the sample depends on the matter power spectrum that I have modelled using the SPT next-to-leading order and on the number of objects and the volume of the sample, which I took from the simulation snapshots.

To speed up the 3PCF estimate, I have adopted the so-called splitting method, which introduces an additional error, σ_r , to be added on top of the shot noise term, corresponding to adding a term to the diagonal elements of the matrix:

$$\mathbb{C}_\zeta = \mathbb{T} + \sigma_r^2 \mathbb{I} \quad (7.4)$$

where \mathbb{C}_ζ is the full covariance matrix considered in the likelihood analysis concerning the 3PCF.

7.3 Results

7.3.1 Clustering measurements

The measurements of 2PCF are shown in Fig. 7.1 for each snapshot. The figure also shows, in the bottom panel, the absolute value of the signal-to-noise ratio defined as $|\xi(r)/\sigma|$, where σ is the error from the diagonal of the covariance matrix \mathbb{C}_ξ . The monopole of the binned ($\Delta r = 10h^{-1}\text{Mpc}$) 2PCF as a function of the pair separation is shown in the top panel. It has the characteristic power-law shape and the BOA peak at $\ell_{\text{BAO}} \sim 110 h^{-1}\text{Mpc}$. The signal-to-noise ratio decreases almost monotonically, being null at the 2PCF zero crossing scale, and then flattens on scales in which errors are dominated by the cosmic variance.

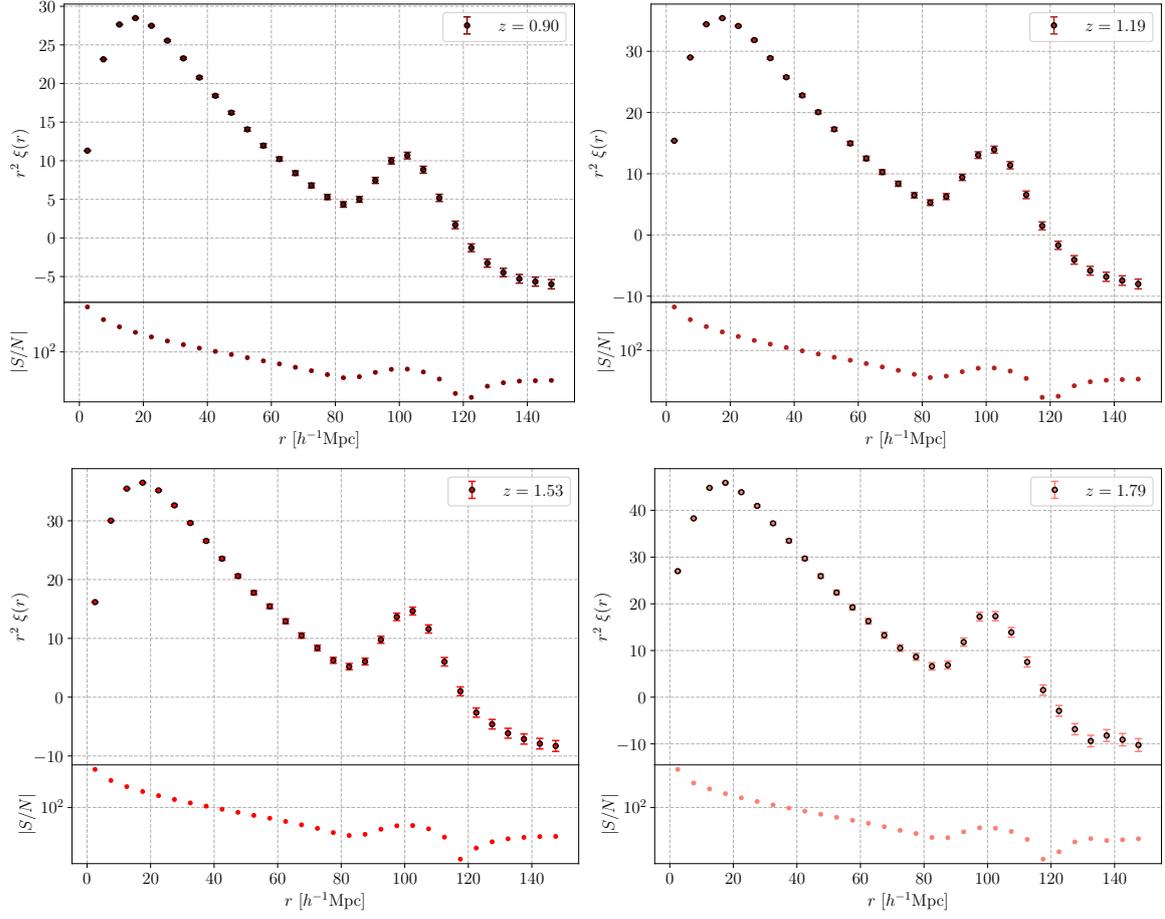


Figure 7.1 Measurements of 2PCF extracted from four Flashes snapshots. For each panel, the upper and the bottom plots refer to the measurement and the signal-to-noise ratio (S/N). *Top-left, top-right, bottom-left, bottom-left* plots refer respectively to $z = 0.90, 1.19, 1.53, 1.78$, the bin-width is $\Delta r = 5h^{-1}\text{Mpc}$.

The four panels of Fig. 7.2 are analogous to those of Fig. 7.1 and show the absolute value of the 3PCF (top panels) and their corresponding signal-to-noise ratio values (bottom panels), as a function of the triangle index. 3PCF measurements span a wide range of scale: $r_{ij} = [5, 145]h^{-1}\text{Mpc}$ and were performed using the bin-width $\Delta r = 10h^{-1}\text{Mpc}$. I notice that the signal-to-noise, which is larger than unit on small scale triangles with $r_{12} \leq 40h^{-1}\text{Mpc}$ but then drops below unit on larger scales. The relation between triangle index and its sides is not trivial and is shown at the very bottom of the figure.

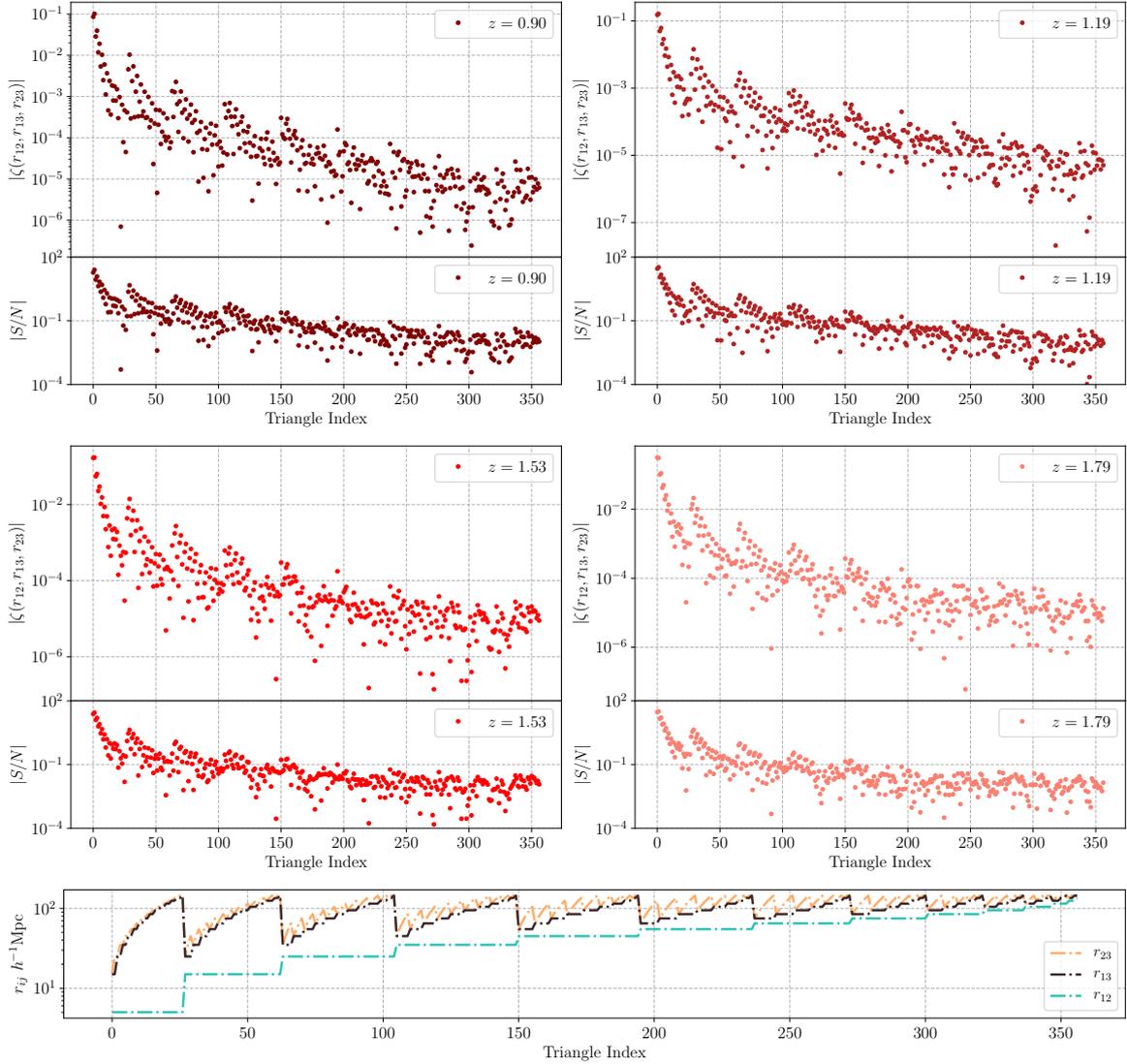


Figure 7.2 Measurements of 3PCF extracted from four Flashhip snapshots. *Top-left, top-right, middle-left, middle-left* panels show measurements and signal-to-noise ratio (S/N), respectively at $z = 0.90, 1.19, 1.53, 1.78$. *Bottom* panel shows the sides of the triangles as a function of the Triangle Index, ranging from $r_{min} = 5 h^{-1}\text{Mpc}$ to $r_{max} = 145 h^{-1}\text{Mpc}$. The bins-width is $\Delta r = 10 h^{-1}\text{Mpc}$.

7.3.2 Parameter inference

To estimate the free parameters, θ , that define the bias model I used the posterior probability $P(\mu(\theta)|d)$, where $\mu(\theta)$ is the model and d is the data vector, respectively. The posterior probability is obtained by multiplying the prior probability, $P(\theta)$, and the likelihood \mathcal{L}

$$\ln \mathcal{L} = -\frac{1}{2} \chi^2(\mu, \mathbb{C}) \quad (7.5)$$

where the chi-square $\chi^2(\theta)$ is

$$\chi^2(\mu, \mathbb{C}) = (d - \mu)^T \mathbb{C}^{-1} (d - \mu) \quad (7.6)$$

and \mathbb{C} is the covariance matrix.

In this analysis, I explore three choices of data-vectors. In the first I considered only the two-point correlation function using the likelihood

$$\ln \mathcal{L}_\xi = -\frac{1}{2} [\hat{\xi} - \xi(\theta)]^T \mathbb{C}_\xi^{-1} [\hat{\xi} - \xi(\theta)], \quad (7.7)$$

where \mathbb{C}_ξ is the Gaussian 2PCF covariance matrix.

In the second one I considered only the three-point correlation functions, with a likelihood

$$\ln \mathcal{L}_\zeta = -\frac{1}{2} [\hat{\zeta} - \zeta(\theta)]^T \mathbb{C}_\zeta^{-1} [\hat{\zeta} - \zeta(\theta)], \quad (7.8)$$

with Gaussian covariance \mathbb{C}_ζ .

Finally, I also considered a joint fit by combining the two- and the three- point correlation functions. The corresponding likelihood, in which I did not include the cross-correlation between the two statistics is

$$\ln \mathcal{L}_{\zeta+\xi} = \ln \mathcal{L}_\zeta + \ln \mathcal{L}_\xi = -\frac{1}{2} [\hat{\zeta} - \zeta(\theta)]^T \mathbb{C}_\zeta^{-1} [\hat{\zeta} - \zeta(\theta)] - \frac{1}{2} [\hat{\xi} - \xi(\theta)]^T \mathbb{C}_\xi^{-1} [\hat{\xi} - \xi(\theta)]. \quad (7.9)$$

To explore the posterior probability of each bias parameter, I used the `emcee` software, a Python implementation of the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler (177). The flat prior distributions used to run the chains are listed in their minimum and maximum values in Tab. 7.3.

Parameter	Uniform prior ranges
b_1	[1, 6]
b_2	[-8, 8]
γ_2	[-10, 10]
γ_{21}	[-10, 10]
$c_0 [h^{-2} \text{Mpc}^2]$	[-100, 100]

Table 7.3 The list of assumed uniform prior ranges.

In order to assess the performance and the range of validity of the model it is useful to consider the phenomenological relations, obtained from N-body experiments, among some of the bias parameters. The adoption of these relations help breaking some degeneracy among

the parameters themselves and increases the precision of their estimate. The relations I considered are the bias relations $b_2(b_1)$, $\gamma_2(b_1)$ and $b_3(b_1)$ presented in Eqs. 3.72, 3.73, 4.76:

7.3.3 Bias constraints from the next-to-leading order 2PCF and leading order 3PCF alone

The triangle plots in Fig. 7.3 show the 2D, and 1D marginalized posterior distributions of the bias parameters used to model the 2PCF. Different colours indicate, instead, results obtained by considering pairs above a minimum separation $r_{\min} = 20, 30, 40 h^{-1} \text{Mpc}$. The colour code is indicated in the plots. Light and dark shading of the same colour indicate the $1\text{-}\sigma$ and $2\text{-}\sigma$ error intervals, respectively. The four panels show the results in the four z -snapshots. As expected, the parameter b_1 has been tightly constrained accurately at each redshift. On the other hand, constraints on the higher-order bias parameters b_2 , γ_2 , and γ_{21} are considerably weaker and degenerate, as revealed by the broad contours and even characterized, in the case of b_2 , by a double peak in the 1D posterior. The EFT parameter $c_0(z)$ is consistent with zero at all redshifts.

In the 3PCF-only analysis I focused on the bias parameters b_1 , b_2 , and γ_2 . Their posterior probability has been estimated using three different minimum triangle sizes, ($r_{\min} = 20 h^{-1} \text{Mpc}$, $30 h^{-1} \text{Mpc}$, $40 h^{-1} \text{Mpc}$) but the same maximum size $r_{\max} = 150 h^{-1} \text{Mpc}$. Also, I did focus on the $\eta_{\min} = 3$ case only. The results for usual four snapshots are displayed in Fig. 7.4. The black and orange dashed lines drawn on top of the 2D contours represent, respectively, the relations $b_2(b_1)$ and $\gamma_2(b_1)$ in Eq. 3.72 and Eq. 3.73. We notice that the agreement between these two relations and the probability contours increases with the redshift, also for small r_{\min} values. This is not surprising since the nonlinear dynamical effects and deviations from the linear bias prescriptions increase over time. The visual inspection of the plots reveals that the agreement with the phenomenological relation at $z = 0.90, 1.19$ and 1.53 is found for $r_{\min} = 30 h^{-1} \text{Mpc}$ whereas at $z = 1.78$ the agreement is maintained down to scale as small as $r_{\min} = 20 h^{-1} \text{Mpc}$.

I have repeated the analysis focusing on the η_{\min} dependence. Fig. 7.5 shows the probability contours for different values of $\eta_{\min} = 3, 2, 1$ at a fixed value of r_{\min} . The r_{\min} values in the examples shown in the plots are those that better match the phenomenological bias relationships. Their values are shown in the plots. In Tab. 7.5 I show, instead, the η_{\min}, r_{\min} combinations that guarantee the best match to those relationships. The agreement, especially for the γ_2 parameter, is particularly good at $z = 1.78$ where one can push the analysis down to $r_{\min} = 20 h^{-1} \text{Mpc}$ and for $\eta = 2$, i.e. for a larger number of triangle configurations than at lower redshift.

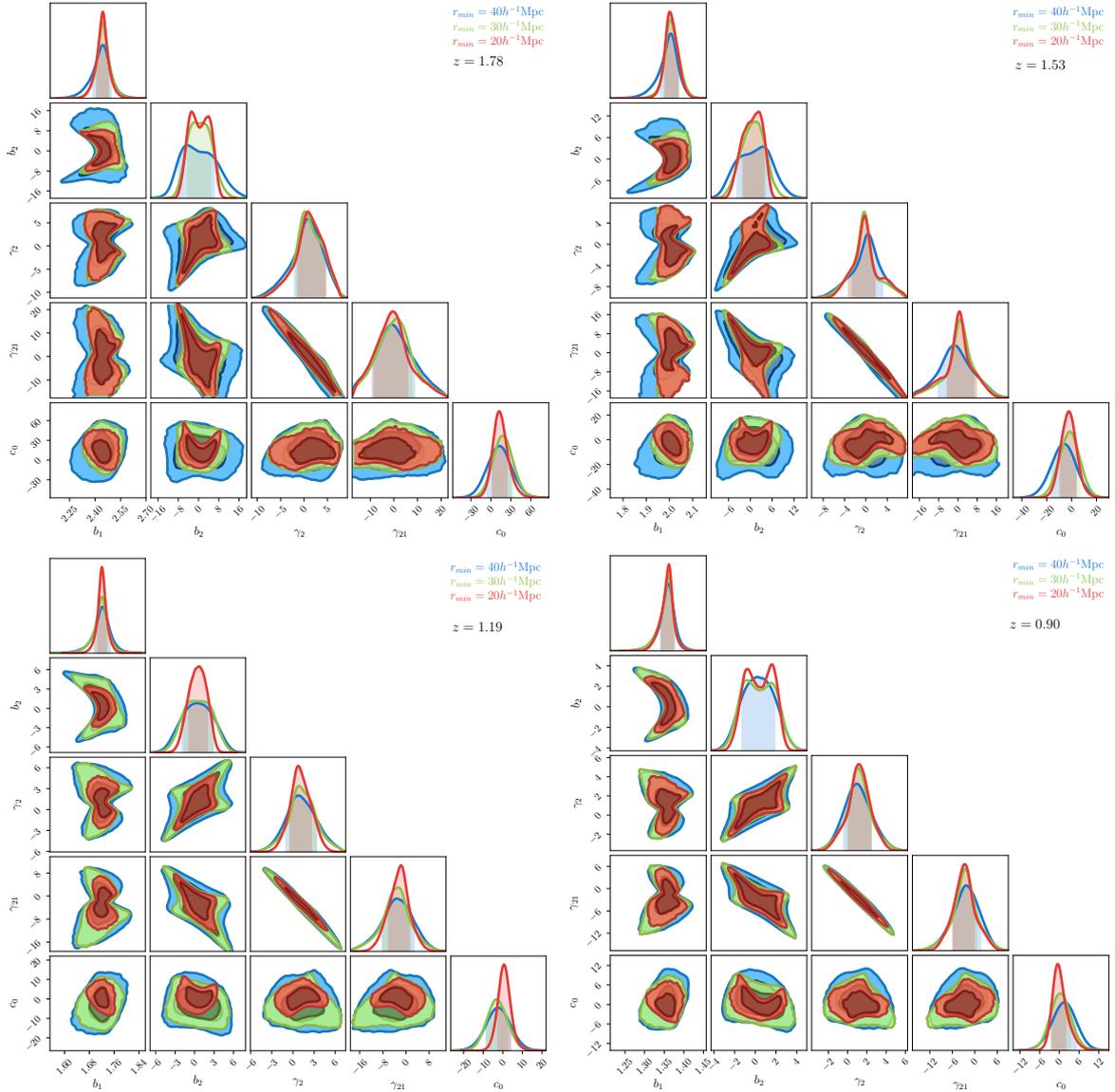


Figure 7.3 Contour plots of next-to-leading order 2PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (blue, green, red). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 1.78, 1.53, 1.19, 0.90$. The maximum scale for 2PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$.

7.3.4 Bias constraints from the joint 2PCF + 3PCF analysis

Having evaluated the constraining power on galaxy biasing from a 3PCF-only and a 2PCF-only analysis, joint modelling of both the 2PCF and 3PCF models being considered is now explored, following what has been discussed in Sec. 7.3.2.

Fig. 7.6 displays the probability contour plots of the bias and ETF parameters inferred from the joint 2PCF and 3PCF. The four panels refer to the same four redshift snapshots

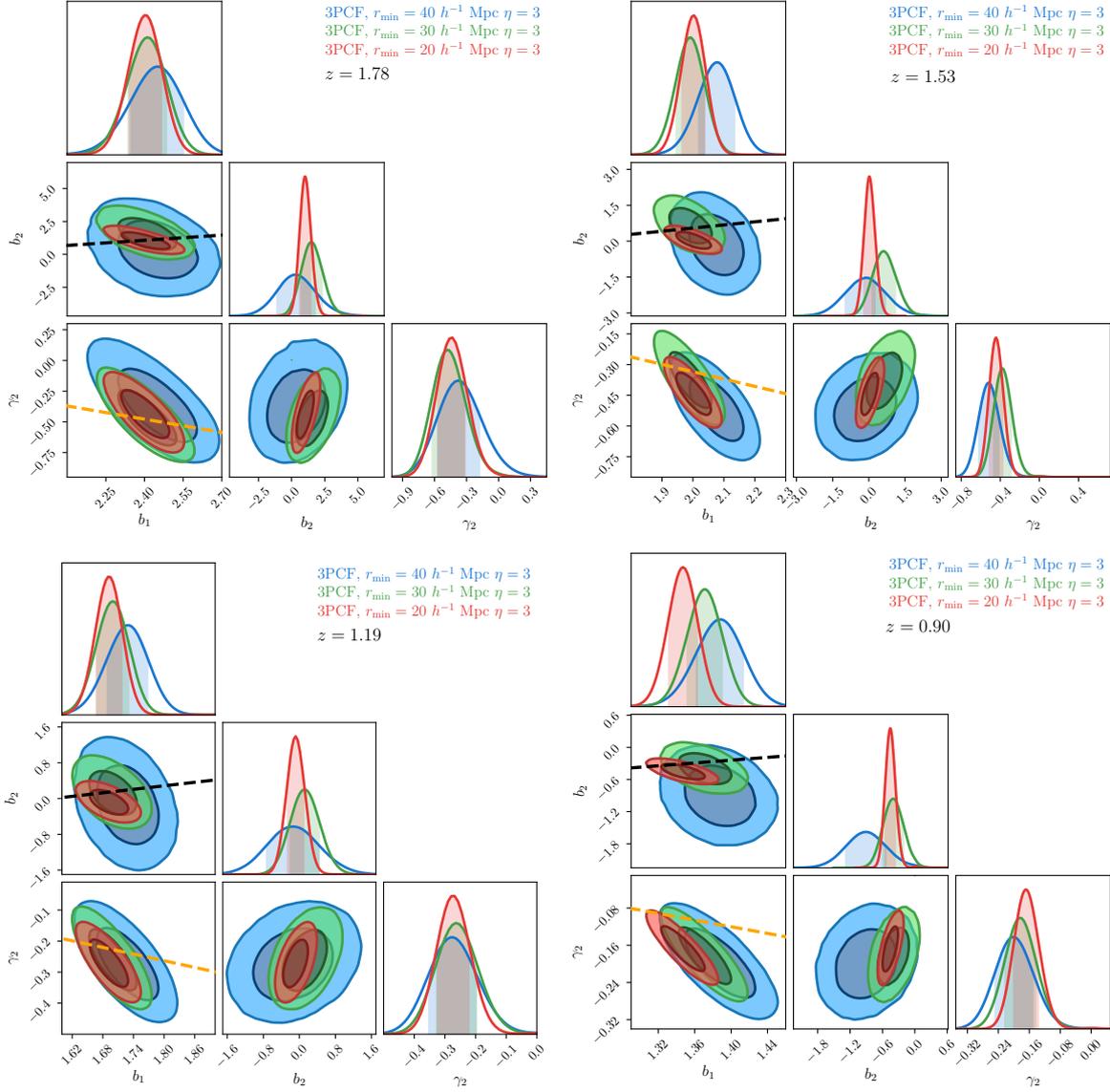


Figure 7.4 Contour plots of leading order 3PCF fixing $\eta_{\min} = 3$ and varying $r_{\min} = 40, 30, 20 h^{-1} \text{Mpc}$ (blue, green, red). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 0.90, 1.19, 1.53, 1.78$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150 h^{-1} \text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73

as in the previous figures. We also show the probability contours obtained previously from the 3PCF-only analysis (yellow contours). The colour brown is used for the 2PCF + 3PCF results. We have explored pairs and triplets with minimum and maximum sizes (separations) $r_{\min} = 20 h^{-1} \text{Mpc}$ and $r_{\max} = 150 h^{-1} \text{Mpc}$, respectively. Dashed lines show the two phenomenological relations among the bias parameters (Eq. 3.72 and 3.73). Tab. 7.5 compares the best-fit values inferred, at the four snapshots' redshifts, from the 3PCF and

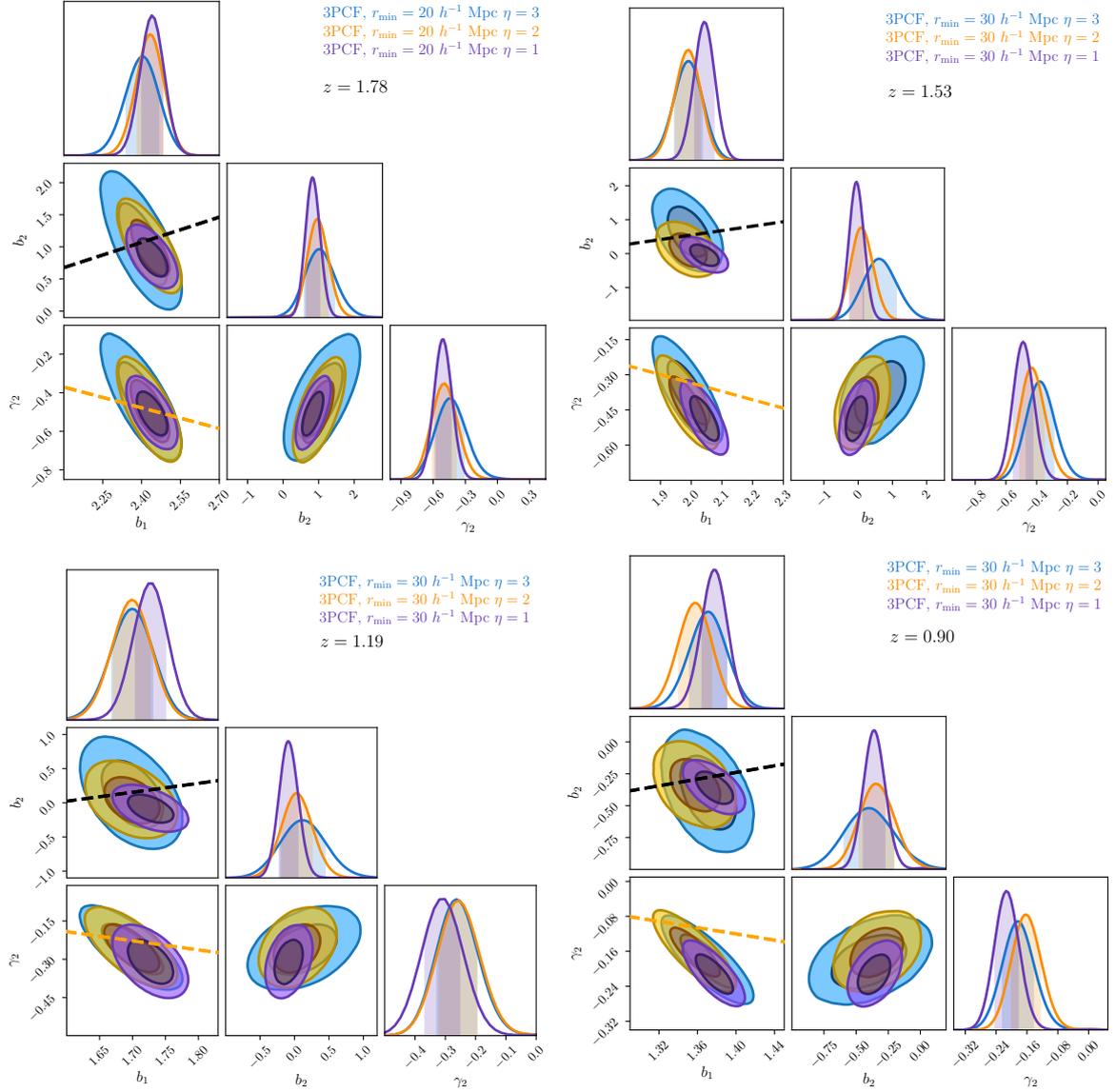


Figure 7.5 Contour plots of leading order 3PCF fixing $r_{\min} = 20, 30, 30, 30 h^{-1} \text{Mpc}$ varying $\eta_{\min} = 3, 2, 1$ (blue, orange, purple). *Bottom-right, bottom-left, top-right* and *top-left* panels refer respectively to $z = 1.78, 1.53, 1.19, 0.90$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150 h^{-1} \text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73

2PCF + 3PCF analyses. Clearly the combinations of the statistics significantly increase the precision of the parameters' estimates and removes some degeneracy that characterizes the analysis that uses one statistic only, particularly the 2PCF one.

Tab. 7.6 shows the ratio of estimated errors from a joint analysis and a 3PCF-only analysis. Results indicate that, as expected, the precision improves significantly for the measurement if b_1 , especially for $z = 1.53$ and $z = 1.78$. Additionally, the better estimate

Redshift	r_{\min}	η_{\min}
$z = 1.78$	20	2
$z = 1.53$	30	3
$z = 1.19$	30	3
$z = 0.90$	30	3

Table 7.4 List of minimum validity scales r_{\min} and η_{\min} or 3PCF models for each snapshot

of b_1 improves the constraints on b_2 and γ_2 , having removed the parameter degeneracy that affects the 3PCF-only or a 2PCF-only analysis.

Analysis	Redshift	b_1	b_2	γ_2
3PCF	$z = 0.90$	$1.370^{+0.020}_{-0.019}$	-0.40 ± 0.19	$-0.182^{+0.039}_{-0.041}$
2PCF + 3PCF	$z = 0.90$	$1.370^{+0.006}_{-0.005}$	$-0.43^{+0.14}_{-0.15}$	$-0.181^{+0.022}_{-0.023}$
3PCF	$z = 1.19$	1.699 ± 0.031	$0.11^{+0.33}_{-0.31}$	$-0.261^{+0.065}_{-0.066}$
2PCF + 3PCF	$z = 1.19$	$1.733^{+0.012}_{-0.095}$	$-0.18^{+0.24}_{-0.23}$	$-0.336^{+0.059}_{-0.046}$
3PCF	$z = 1.53$	$1.990^{+0.046}_{-0.044}$	$0.61^{+0.50}_{-0.47}$	$-0.378^{+0.090}_{-0.089}$
2PCF + 3PCF	$z = 1.53$	$2.015^{+0.016}_{-0.014}$	$0.49^{+0.39}_{-0.35}$	$-0.421^{+0.049}_{-0.046}$
3PCF	$z = 1.78$	$2.432^{+0.050}_{-0.049}$	0.97 ± 0.29	$-0.507^{+0.119}_{-0.099}$
2PCF + 3PCF	$z = 1.78$	$2.474^{+0.019}_{-0.025}$	$0.77^{+0.18}_{-0.17}$	$-0.561^{+0.082}_{-0.094}$

Table 7.5 List of inferred values of bias parameters of the leading order 3PCF model estimated throughout a 3PCF-only and a joint 2PCF and 3PCF analysis.

Redshift	$\sigma_{\text{joint}}/\sigma_{\text{3PCF}}(b_1)$	$\sigma_{\text{joint}}/\sigma_{\text{3PCF}}(b_2)$	$\sigma_{\text{joint}}/\sigma_{\text{3PCF}}(\gamma_2)$
$z = 0.90$	0.282	0.763	0.501
$z = 1.19$	0.579	0.734	0.824
$z = 1.53$	0.334	0.763	0.531
$z = 1.78$	0.445	0.603	0.807

Table 7.6 List of ratios between errors on bias parameters of the leading order 3PCF model estimated throughout a 3PCF and a joint 2PCF and 3PCF analysis.

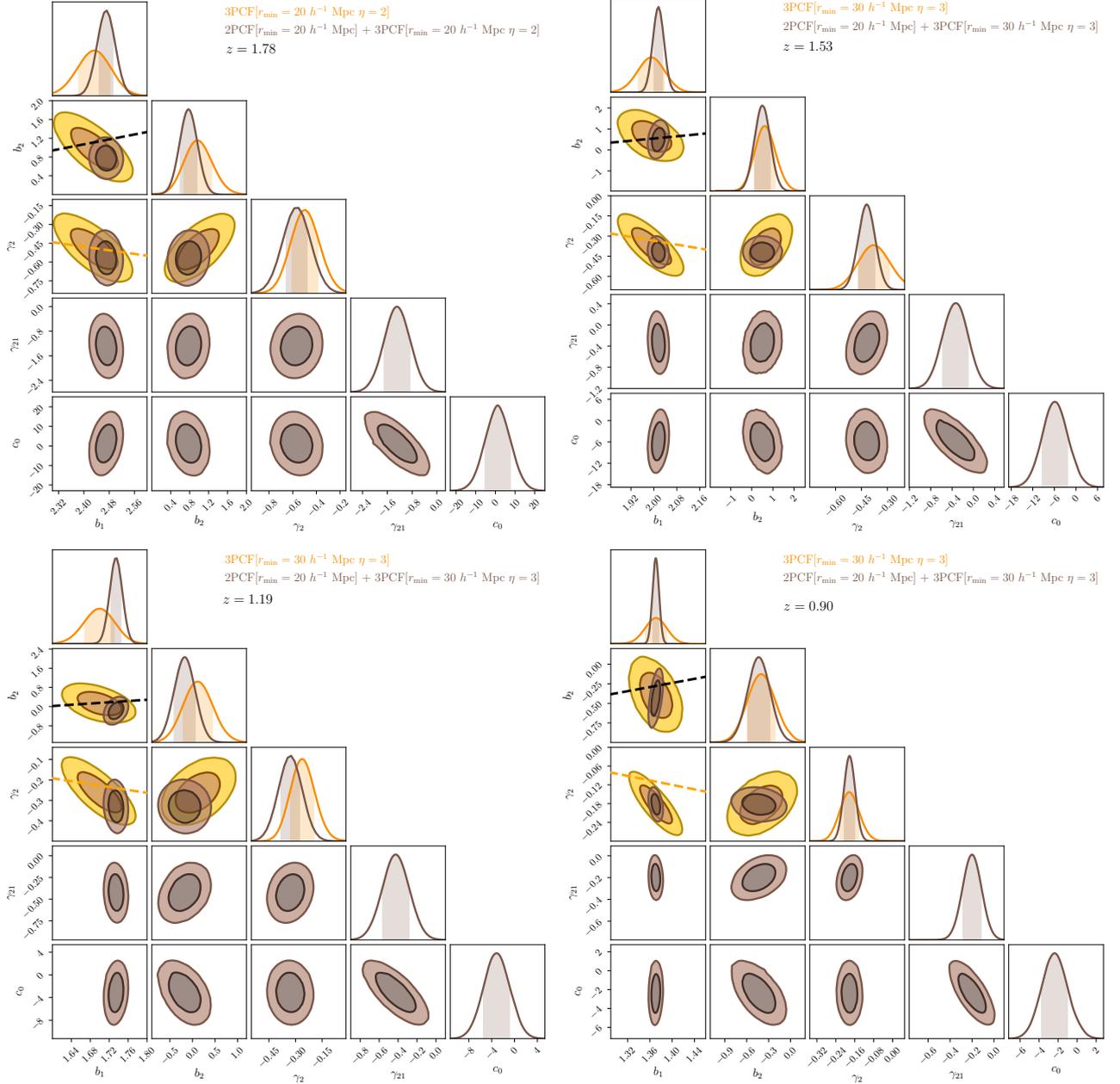


Figure 7.6 Contour plots from a 3PCF-only (yellow) and a joint 2PCF and 3PCF analysis (brown). The minimum scales for the 3PCF are the one listed in Tab. 7.5. For the 2PCF, it is fixed to $r_{\min} = 20h^{-1}\text{Mpc}$. The maximum scale both for 3PCF and 2PCF are the maximum clustering measured lengths, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation presented in Eq. 3.72 and Eq. 3.73

8

Bias loop corrections in configuration space: a joint 2PCF + 3PCF analysis

In previous chapters, I have presented the next-to-leading order model for the 3PCF of the matter and haloes and compared them with measurements obtained from simulated data. In this Chapter, I will instead consider the next-to-leading order model for the 3PCF of the galaxies in a joint analysis with the analogous 2PCF model for galaxies. This represents a step forward with respect to current joint analyses that, as we did in the last Chapter, only consider the leading order model for the 3PCF. Our ability to estimate them well into the nonlinear regime against prediction from halos from N-body simulations is the focus of this Chapter.

8.1 Models

For the halo 2PCF, I have used the same next-to-leading order model as in Chapter 7, which is obtained by 1D-FFTL_{og} transforming the analogous one-loop power spectrum model from Fourier space. This model has four free parameters that specify the biasing relation:

$$\{b_1, b_2, \gamma_2, \gamma_{21}\}. \quad (8.1)$$

For the 3PCF, I have used two models. The first one is the same leading-order model used in 7 and is characterized by three free parameters:

$$\{b_1, b_2, \gamma_2\}. \quad (8.2)$$

Then I also considered the new, considerably more elaborated, galaxy 3PCF model based on next-to-leading order expansion. This model is built upon the bispectrum model of (126) described in Sec. 4.2.3, and is meant to provide a more accurate description of the relation between galaxy/haloes distribution and the underlying dark matter fluctuations

in the small scale regime. Its implementation relies on the 2D-FFTLog, applied, however, much more intensively due to the larger number of loop terms in the bispectrum model (see Appendix A.2 for details). It is worth noting that, compared to Fourier space modelling, the computational burden is much higher due to the large number of points used to sample Fourier space quantities before 2D-FFTLog transforming them to configuration space. The actual computational cost is a trade-off between computing memory and speed on one end and precision and accuracy on the other. The number of free bias parameters for this 3PCF model is also larger:

$$\{ \underbrace{b_1}_{1st}; \underbrace{b_2, \gamma_2}_{2nd}; \underbrace{b_3, \gamma_2^\times, \gamma_3, \gamma_{21}}_{3rd}; \underbrace{\gamma_{21}^\times, \gamma_{211}, \gamma_{22}, \gamma_{31}}_{4th} \}. \quad (8.3)$$

where I have labelled the parameters as first, second third and fourth order in the associated perturbative expansion. First and second-order bias parameters are in common with the leading-order model. The third and fourth-order bias parameters only appear at this level of the expansion, apart from for γ_{21} , which is also included in the next-to-leading order 2PCF model. The bias parameters listed in Eq. 8.3 represents a subset of those used to model 3-point statistics in Fourier space (126) since I have neglected all the stress-tensor parameters, which are specific to the EFT treatment and that we did not consider in this analysis. The EFT parameter c_0 that was part of the 2PCF model in the previous Chapter is not included in the next-to-leading 2PCF model either to avoid sampling too high-dimensional space in the likelihood analysis. When it comes to comparing models with real data, though, all the bias parameters are free to vary and may absorb some of the nonlinear dynamics that are present on nonlinear scales. The rationale for introducing this sophisticated 3PCF next-to-leading order model is to probe deeper into the nonlinear regime and to improve the match between modelling and data on smaller scales for better precision and accuracy in estimating the cosmological parameters.

8.2 Dataset

For our measurements, we use the MINERVA set of 298 N-body simulations (155) performed with the GADGET-2 code (178). Each experiment follows the evolution of 1000^3 dark matter particles, each one with a mass $2.67 \times 10^{11} h^{-1} M_\odot$, in a periodic cubic box of side length $L = 1500 h^{-1}$ Mpc, assuming a flat background cosmology with $h = 0.695$, $\Omega_m = 0.285$, and $\Omega_b = 0.046$, matching the results obtained by combining the analyses of the WMAP CMB maps and the spatial distribution of galaxies in the BOSS DR9 survey (see Table I of (179) for details). The initial density and velocity fields were generated at a redshift of $z_{in} = 63$

by displacing the simulation particles from a regular grid using second-order Lagrangian perturbation theory. The transfer function for the Gaussian linear fluctuations in the matter density was computed using the CAMB code (118) with a primordial scalar spectral index of $n_s = 0.9632$ and an *rms* matter density fluctuation of $\sigma_8 = 0.828$. Dark matter halos were identified using a standard friends-of-friends algorithm with a linking length of 0.2 times the mean one-dimensional inter-particle separation. Gravitationally unbound particles were removed using the SUBFIND code (178). I only considered halos that contained at least 42 particles, corresponding to a minimum mass of $M \simeq 1.12 \times 10^{13} h^{-1} M_\odot$. Also, I focus here on the $z = 1$ snapshot only, which overlaps with the redshift range probed by spectroscopic galaxy surveys such as Euclid (40; 41) or DESI (42). The mean number density for the resulting halo in $z = 1$ box is $2.13 \times 10^{-4} h^3 \text{Mpc}^{-3}$.

8.2.1 Measurements

To estimate the isotropic two-point correlation function, the natural estimator presented in Sec. 7.2.1 has been used.

Concerning the three-point correlation function, it has been used the estimator presented in Sec. 5.4 and used in Sec. 6.2.1, 7.2.1. The estimator uses two catalogues as input: one containing haloes and the other containing randomly distributed objects, which lack clustering properties. To measure the 3PCF, it has been employed the random splitting method, in which data are divided into small random samples, measured the 3PCF for each sample, and then took the average. To ensure that the data-random ratio is consistent, we set $N_R = 1$ after splitting. Since the number of objects in each sample varies, we adjusted the number of random objects accordingly. For the 3PCF measurements, I used 50 times more random objects than haloes to ensure that the estimation of the 3PCF is not biased but only contributes to its variance. However, this has a negligible impact on the overall error budget.

8.2.2 Covariance

Unlike in the previous analyses presented in this Thesis, the covariance matrix that I use in the likelihood analysis is not based on the Gaussian model, but it is estimated directly from the halo two- and three- point correlation functions measured in the 298 MINERVA boxes:

$$\hat{C}_{i,j} = \frac{1}{N_{mocks} - 1} \sum_{n=1}^{N_{mocks}} (\hat{d}_i^n - \bar{d}_i) (\hat{d}_j^n - \bar{d}_j) \quad (8.4)$$

where d_i^m is the data vector containing the estimates of ξ and ζ , i and j identify data vector bins, while n identifies the n -th halo catalog. \bar{d} represents the average among the mocks'

data vectors i.e.

$$\bar{d}_i = \frac{1}{N_{mocks}} \sum_{n=1}^{N_{mocks}} \hat{d}_i^n \quad (8.5)$$

The precision of this numerically derived covariance matrix has been analysed and discussed in (151). In this work, the matrix was applied in an analysis focused on the leading-order 3PCF of halo distributions. The 3PCF model predictions were compared with the same dataset at the redshift described in this Chapter. The numerical covariance derived from a single mock was compared with its theoretical counterpart, revealing that the latter underestimated the uncertainties compared to the numerical estimate. Moreover, using the numerical covariance, the minimum scale of validity at which the predictions of the leading-order 3PCF model match the dataset was determined to be $r_{\min} = 40 h^{-1}\text{Mpc}$, $\eta = 3$.

8.3 Results

8.3.1 Clustering measurements

Fig. 8.1 shows the average halo 2PCF measured, at $z = 1$, over the separation range $[5, 145h^{-1}\text{Mpc}]$ in a bin width of $10h^{-1}\text{Mpc}$. The bottom panel shows the signal-to-noise ratio, previously defined, calculated using the diagonal terms of the covariance matrix \mathbb{C}_ξ estimated from the mocks.

It is worth noting that the signal-to-noise ratio is higher than in the galaxy 2PCF Flagship case considered in the previous Chapter since we are computing it using the error on the mean 2PCF rather than that on the 2PCF of the individual simulation box. And possibly, also because we are using a numerical covariance instead of adopting the Gaussian model.

The measurements of the 3PCF at a $z = 1$ are displayed in Fig 8.2. They have been obtained by considering triangles with sizes ranging from $5h^{-1}$ to $145h^{-1}\text{Mpc}$. The corresponding signal-to-noise ratio, shown in the middle panel, is also larger than in the Flagship case explored in the previous Chapter, due to the same reason advocated for the 2PCF. The mapping between the triangle index and sides is shown in the bottom panel.

8.3.2 Parameter inference

Similarly to what I have presented in Sec. 7.3.2, to estimate the bias parameters of the models, collectively grouped in the vector θ , I estimate their posterior probability $P(\mu(\theta)|d_\alpha)$. This probability takes into account the model prediction, represented by $\mu(\theta)$, and the data vector,

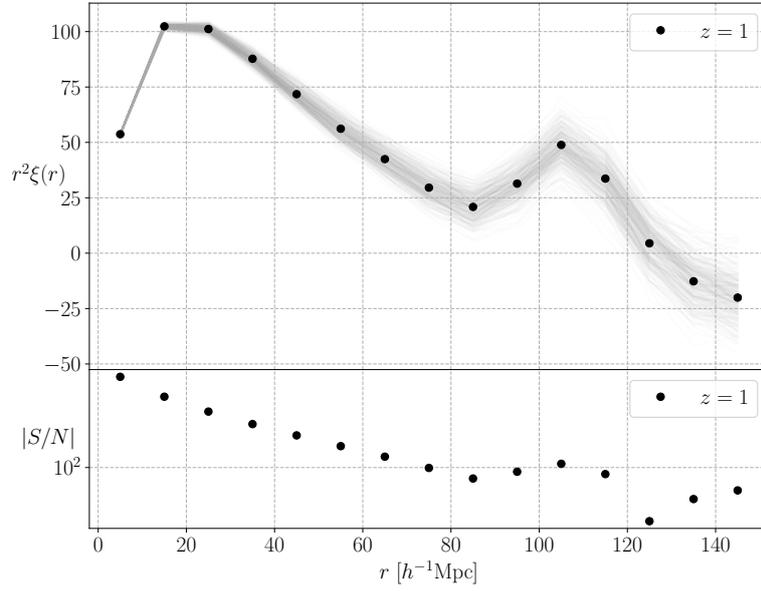


Figure 8.1 Average (black circle) of 2PCF extracted from measuring 2PCF from 298 mock catalogues (thin grey lines). Error bars correspond to the diagonal of the numerical covariance extracted from the 298 mocks. The bin is $\Delta r = 5 h^{-1}\text{Mpc}$

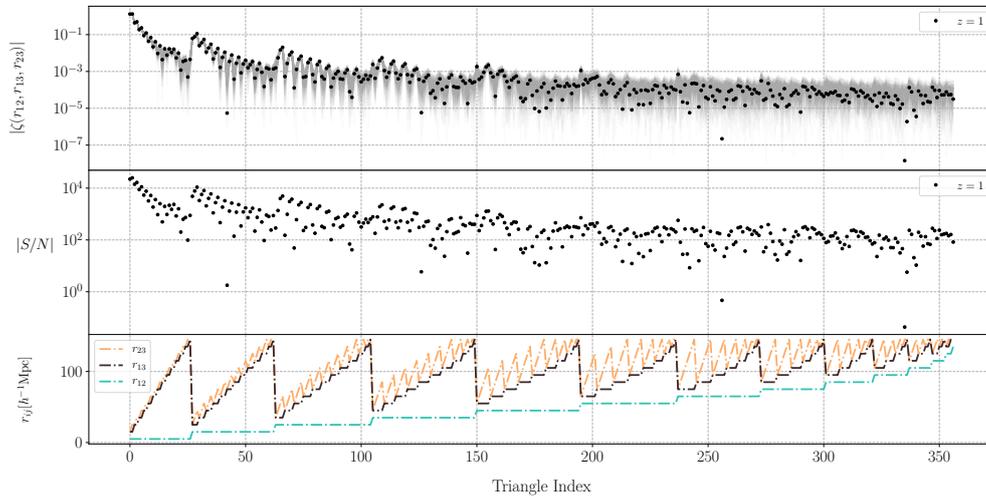


Figure 8.2 *Top panel*: mean average (black circle) of 3PCF extracted from measuring 3PCF from 298 mock catalogues (thin grey lines). *Middle panel*: signal-to-noise ratio. *Bottom panel*: sides of triangles as a function of the Triangle Index. The bin is $\Delta r = 5 h^{-1}\text{Mpc}$

represented by d_α . It is calculated by multiplying the prior probability, $P(\theta)$, with the

likelihood \mathcal{L}_α where α identifies the mock index.

$$\ln \mathcal{L}_\alpha = -\frac{1}{2} \chi^2(\mu, \mathbb{C}) \quad (8.6)$$

where the chi-square $\chi^2(\theta)$ is given by

$$\chi^2(\mu, \mathbb{C}) = (d_\alpha - \mu)^T \mathbb{C}^{-1} (d_\alpha - \mu) \quad (8.7)$$

where \mathbb{C} is the numerical covariance. The product of all individual likelihood of mocks obtains the total likelihood adopted in my evaluation

$$\log \mathcal{L}_{\text{tot}} = \sum_{\alpha=1}^{N_{\text{mock}}} \log \mathcal{L}_\alpha. \quad (8.8)$$

In this study, I investigate three kinds of analyses. The first one focuses solely on 2PCF, examining the next-to-leading order model for galaxy using the likelihood presented in Eq. 7.7. The second one focuses solely on 3PCF, examining both the leading and next-to-leading order models for galaxy using the likelihood presented in Eq. 7.8. Last, I also consider a joint fit considering the combination of two- and three- point correlation functions, using the likelihood discussed in Eq. 7.9. I used the `emcee` software, a Python implementation of the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler (177), to explore the posterior probability of each bias parameter. I defined flat prior distributions for the parameters, with their minimum and maximum values listed in Tab.8.1.

Parameter	Uniform prior ranges
b_1	[1, 6]
b_2	[-8, 8]
γ_2	[-10, 10]
γ_{21}	[-10, 10]
b_3	[-80, 80]
γ_2^\times	[-80, 80]
γ_3	[-60, 60]
γ_{21}^\times	[-50, 50]
γ_{211}	[-50, 50]
γ_{22}	[-50, 50]
γ_{31}	[-150, 150]

Table 8.1 The list of assumed uniform prior ranges.

As in the previous Chapter, I will adopt, in some steps of the analysis, the same bias relation $b_2(b_1)$, $\gamma_2(b_1)$ presented in Eqs. 3.72, 3.73. However, since I am now considering third-order bias parameters, I also consider the relation the bias relation $b_3(b_1)$ presented in Eq. 4.76 and calibrated, like the others, on N-body data (100).

8.3.3 Bias constraints from 2PCF and 3PCF measurements alone

Let us start the analysis with the case in which only the 2PCF measurements are available. Fig 8.3 displays the probability contour plots for the bias parameters of the 2PCF model (Eq. 8.1) obtained by progressively increasing the minimum pair separation considered in the analysis: from $r_{\min} = 20 h^{-1}\text{Mpc}$ to $30 h^{-1}\text{Mpc}$ and finally $40 h^{-1}\text{Mpc}$, represented by red, green, and blue, respectively. The maximum separation was set equal to $r_{\max} = 145 h^{-1}\text{Mpc}$ in all cases. The black and orange dashed lines in the figure represent the bias relationships $b_2(b_1)$ and $\gamma_2(b_1)$ specified by Eqs 3.72, 3.73, respectively. The comparison between the contours and the lines shows that when the analysis is pushed down to $r_{\min} = 20 h^{-1}\text{Mpc}$, then one underestimates the parameter γ_2 . The mismatch is progressively reduced by increasing the minimum scale and disappears for $r_{\min} = 40 h^{-1}\text{Mpc}$. We use this result to set $r_{\min} = 40 h^{-1}\text{Mpc}$ as the minimum pair separation considered in this analysis presented in this Chapter that uses the halo 2PCF.

Next, I have considered the 3PCF case only. In this case, I do not only vary the value of r_{\max} , the minimum size of the triangles used in the analysis, but I also consider two different 3PCF models, the leading-order and the next-to-leading-order ones, each one characterized by its own set of bias parameters. In Fig 8.4, I show results obtained using the leading-order 3PCF model with $r_{\min} = 40h^{-1}\text{Mpc}$ and $\eta_{\min} = 3$ (blue contours) and with the next-to-leading-order 3PCF model with $\eta_{\min} = 3$ and $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (purple, red, and green, respectively). The choice of limiting the analysis to triangle configurations $\eta_{\min} = 3$ is a conservative one and inspired by the results of the (151) analysis. Indeed, as I have checked, the results obtained with my leading-order model obtained via 2D-FFTLLog match those of the 3PCF model of (151) generated using 1D-FFTLLog. In addition to the black and orange dashed lines, we add the green dashed line that represents the third bias relation adopted in the analysis (Eq 4.76).

The visual inspection of the figure reveals that

- The relation $b_2(b_1)$ intersects with all contours except the next-to-leading-order one with $r_{\min} = 20h^{-1}\text{Mpc}$;
- The relation $\gamma_2(b_1)$ intersects the contour of the next-to-leading-order 3PCF with $r_{\min} = 30h^{-1}\text{Mpc}$ and is barely consistent with the $r_{\min} = 40h^{-1}\text{Mpc}$ case. It is

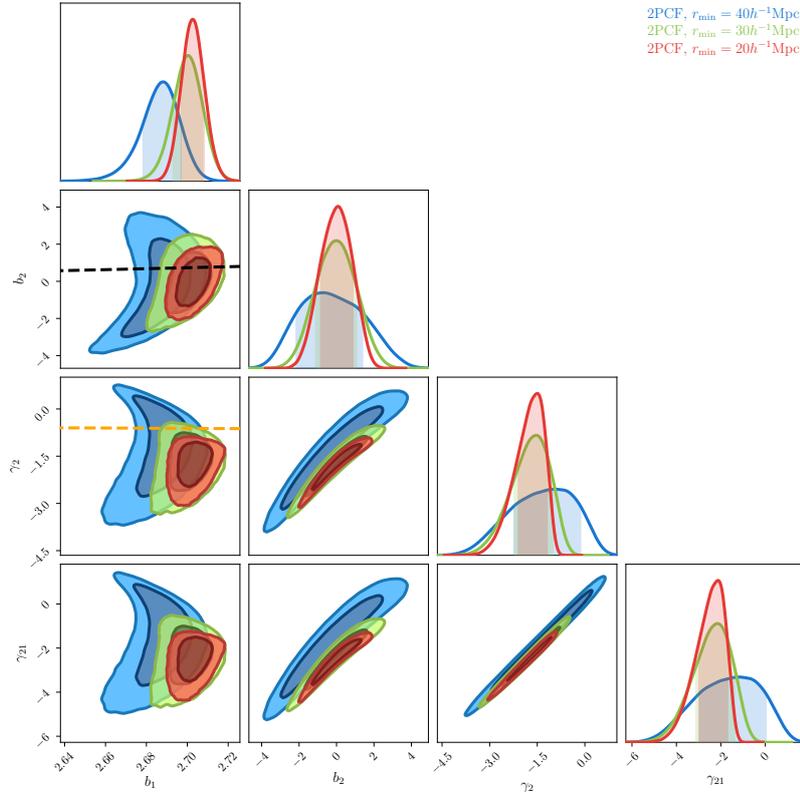


Figure 8.3 Contour plots of next-to-leading order 2PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (blue, green, red). The maximum scale for 2PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73.

inconsistent with both the 3PCF leading-order model result and the next-to-leading-order 3PCF case with $r_{\min} = 20h^{-1}\text{Mpc}$,

- The relation $b_3(b_1)$ intersects all next-to-leading-order 3PCF models but the $r_{\min} = 20h^{-1}\text{Mpc}$ one.

To summarize, the next-to-leading order model does not seem to be able to reproduce the halo 3PCF down to $r_{\min} = 20h^{-1}\text{Mpc}$. It does, instead, fit the data reasonably well (assuming that the phenomenological biasing relations are an accurate description of the halo bias on all scales) for $r_{\min} \geq 30h^{-1}\text{Mpc}$. Both these models perform better than the leading-order one with $r_{\min} = 40h^{-1}\text{Mpc}$, although their constraining power is weaker due to the larger number of free parameters. This is not always the case, though. The peak in the 1D posterior distribution for b_2 is sharper in the next-to-leading case than in the leading order one, suggesting that its ability to constrain that parameter is intrinsically better.

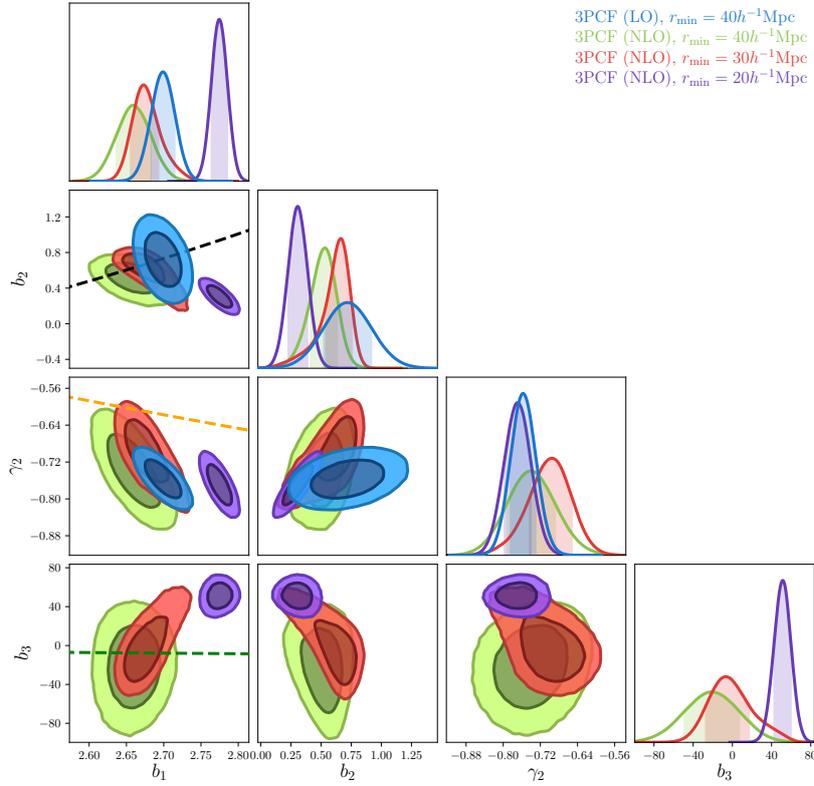


Figure 8.4 Contour plots of next-to-leading order 3PCF varying $r_{\min} = 40, 30, 20h^{-1}\text{Mpc}$ (green, red, purple) and the leading order 3PCF at $r_{\min} = 20h^{-1}\text{Mpc}$ (blue) fixing $\eta_{\min} = 3$. The maximum scale for 3PCF is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black, orange and green lines refer to the bias relation reported in Eqs. 3.72 , 3.73, 4.76.

To increase the precision in estimating the bias parameters of the next-to-leading order 3PCF model, we can use the Lagrangian bias relations that exist among the bias parameters (126) to reduce the dimension of the parameter space. The outcome of this exercise is shown in Fig 8.5 for the conservative case $r_{\min} = 40h^{-1}\text{Mpc}$ and $\eta_{\min} = 3$. The blue contour shows the leading-order 3PCF model results, plotted for reference. The green contour represents the probability contours obtained when all the bias parameters are free to vary. The red contours show the same case in which, however, the Lagrangian relations $\gamma_{31}(b_1)$ and $\gamma_{22}(b_1)$ have been adopted. Finally, the purple contours represent the same model but with the additional inclusion of Lagrangian relations $\gamma_{211}(b_1)$, $\gamma_{21}^{\times}(b_1)$, $\gamma_{21}(b_1)$, and $\gamma_3(b_1)$.

The visual inspection of Fig 8.5 shows that adopting six Lagrangian bias relations significantly reduces the uncertainty on the parameter b_3 , whose value is consistent with that expected from the phenomenological relations. The parameters b_1 , b_2 , and γ_2 are also measured more precisely. However, the best fit values of b_2 , and, especially, γ_2 , do not appear

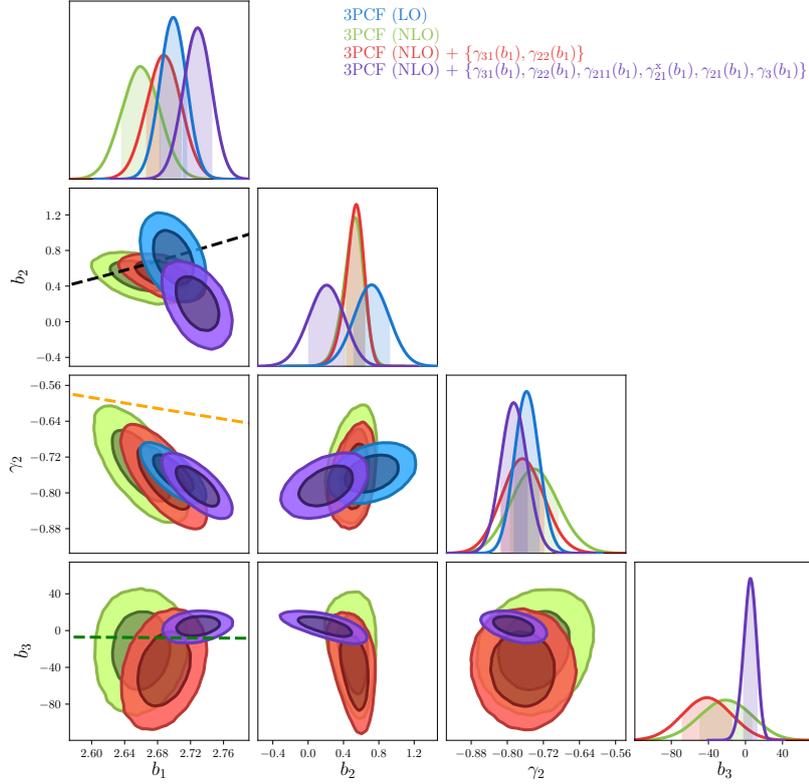


Figure 8.5 Contour plots of leading and next-to-leading order 3PCF fixing $r_{\min} = 40h^{-1}\text{Mpc}$. The blue contour represents constraints from the leading order 3PCF. Green from the next-to-leading order 3PCF with free parameters. Purple and red from the next-to-leading order 3PCF model using the Lagrangian bias relation specified in the legend. The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black, orange and green lines refer to the bias relation reported in Eqs. 3.72 , 3.73, 4.76.

to be consistent with the expected relations, suggesting that the Lagrangian relations may not accurately describe all aspects of the bias relation.

I stress that when the next-to-leading order model does not perform well, there is a noticeable increase in the discrepancy between the inferred values of the linear bias parameter b_1 . This suggests an internal degeneracy of the parameters that must be closely monitored. The presence of degeneracy between parameters can lead to difficulties in accurately constraining them individually, as well as in interpreting the results. To address this issue, we perform, in the next section, a joint 3PCF and the 2PCF analysis that adopts the next-to-leading order models for both statistics.

8.3.4 Bias constraints from the joint 2PCF and 3PCF analysis: next-to-leading order models

In Fig. 8.6, I show the constraints on the bias parameters b_1 , b_2 , and γ_2 obtained from the joint two- and three- point correlation function analysis with $r_{\min} = 40h^{-1}\text{Mpc}$ and $\eta_{\min} = 3$.

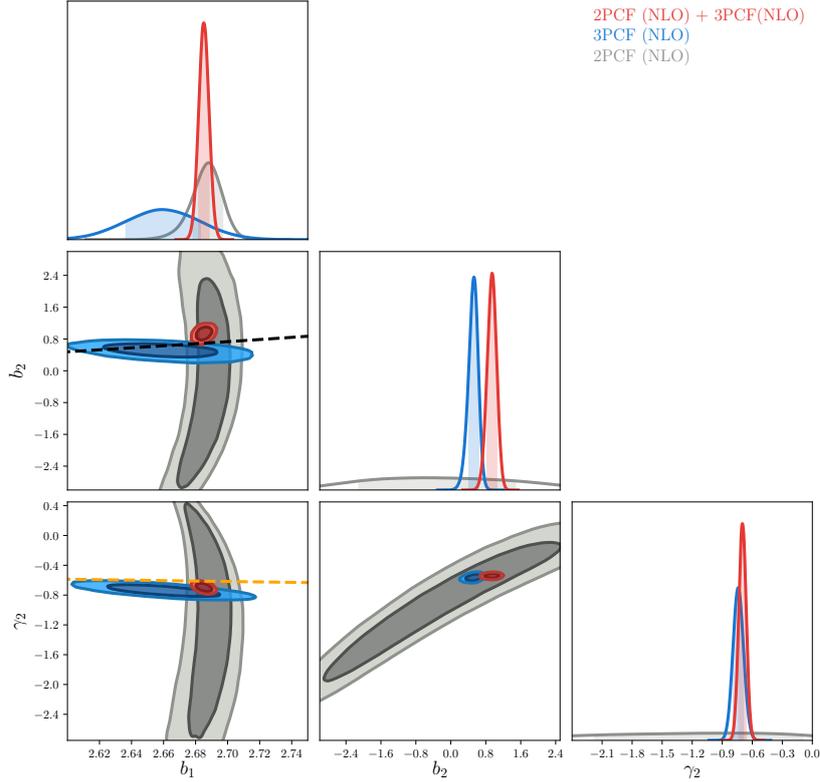


Figure 8.6 Contour plots from the next-to-leading order 2PCF (grey), the next-to-leading order 3PCF (blue) and the joint 2PCF + 3PCF (red) at $r_{\min} = 40h^{-1}\text{Mpc}$ at $\eta_{\min} = 3$. The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73.

The grey and blue probability contours represent, respectively, the results obtained by considering the 2PCF and the 3PCF measurements alone and their next-to-leading order models. The red contours are obtained from the joint analysis. The ability of the 2PCF to estimate b_1 is combined with that of the 3PCF to constrain b_2 and γ_2 to obtain a precise estimate for all of them that are consistent with the phenomenological relations (dashed lines). There is a mild tension between the b_2 values obtained from the joint (red) and from the 3PCF-only (blue) analysis.

Next, I have compared the results just described and obtained using next-to-leading order models for both functions, two- and three- point correlation function, to those obtained from an analogous joint analysis that, however, adopts leading order to model the 3PCF.

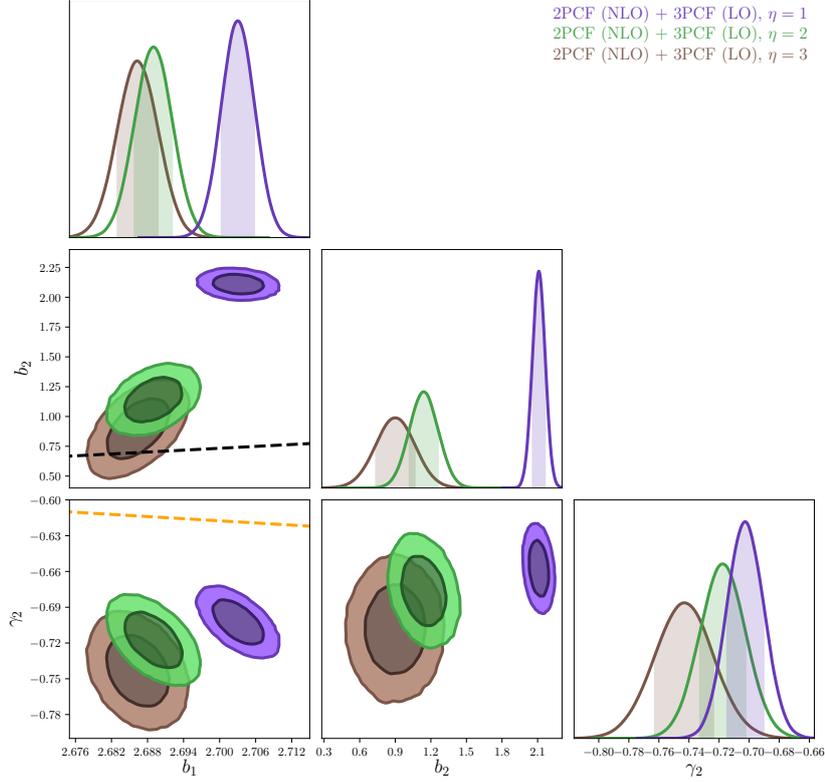


Figure 8.7 Contour plots from the joint 2PCF (next-to-leading order) + 3PCF (leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ for $\eta_{\min} = 3, 2, 1$ (brown, green, blue). The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73.

Fig. 8.7 shows the results obtained using the leading order model for the 3PCF. In all analyses I set $r_{\min} = 40h^{-1}\text{Mpc}$. Probability contours with different colours are used for different choices of triangle configurations: brown, green, and violet indicate $\eta_{\min} = 3, 2, 1$, respectively. In no case do the results match expectations for γ_2 , whereas for b_2 the only case in which the measured value agrees with the phenomenological relation is $\eta_{\min} = 3$. Moreover, the results obtained when the quasi-isosceles configurations $\eta_{\min} = 1$ are included in the analysis are inconsistent with those obtained for all the other configurations.

A more explicit comparison between leading order and next-to-leading order results is shown in Fig. 8.8. The brown contours represent the leading order case with $\eta_{\min} = 3$ and $r_{\min} = 40h^{-1}\text{Mpc}$ (the value of r_{\min} is the same in all case considered). The dark orange

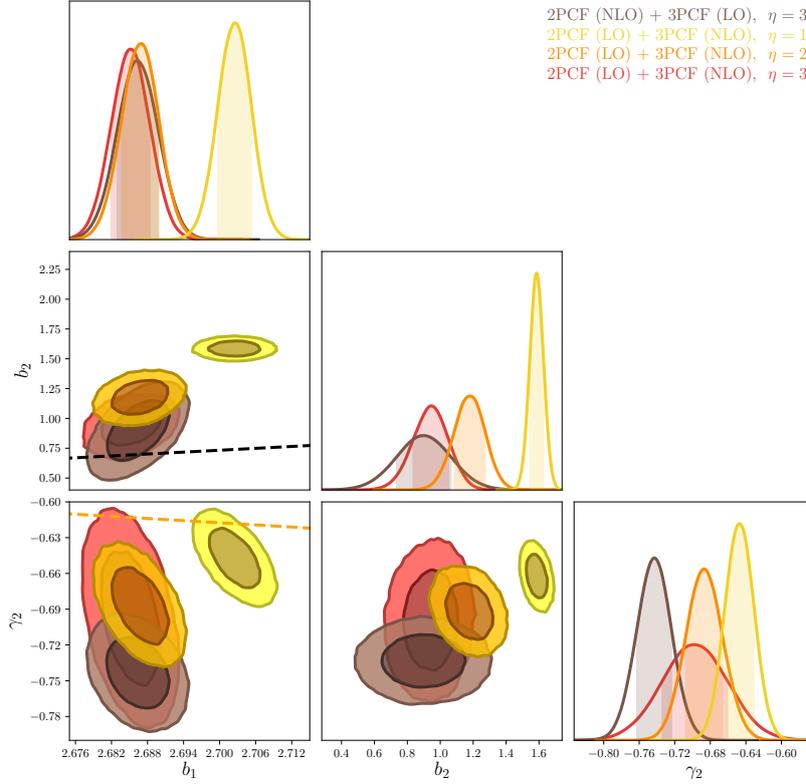


Figure 8.8 Contour plots from the joint 2PCF (next-to-leading order) + 3PCF (next-to-leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ for $\eta_{\min} = 3, 2, 1$ (red, orange, yellow) and the joint 2PCF (next-to-leading order) + 3PCF (leading order) at $r_{\min} = 40h^{-1}\text{Mpc}$ at $\eta_{\min} = 3$ (red, orange, yellow). The maximum scale is the maximum clustering measured length, i.e. $r_{\max} = 150h^{-1}\text{Mpc}$. Dashed black and orange lines refer to the bias relation reported in Eq. 3.72 and Eq. 3.73.

contours show the results obtained using the next-to-leading model and the same set of triangles $\eta_{\min} = 3$. The number of triangle configurations has been progressively increased by setting $\eta_{\min} = 2$ (light orange) and $\eta_{\min} = 1$ (yellow). The main results of this comparison are:

- The results obtained with $\eta_{\min} = 3$ using the leading order model are consistent with those obtained using the next-to-leading model. The adoption of the next-to-leading order model increases the magnitude of the uncertainty for γ_2 because of the larger number of free parameters involved. However, it also eliminates the tension with the phenomenological relation (orange dashed line). Errors on b_1 and b_2 are quite similar for both models.
- Increasing the number of triangle configurations by setting $\eta_{\min} = 2$ improves the precision but generates a systematic mismatch with the expected values.

- Like in the leading order case, setting $\eta_{\min} = 1$ produces results that are inconsistent with this obtained with higher η values.

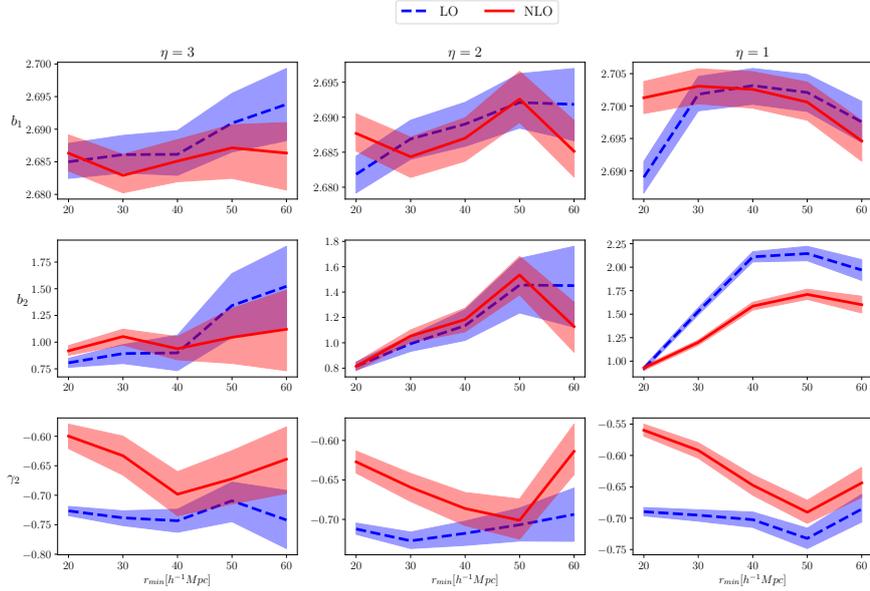


Figure 8.9 Inferred bias parameters b_1 , b_2 and γ_2 as a function of the minimum scale r_{\min} and parameter η_{\min} . Solid red and dashed blue lines refer respectively to the 3PCF next-to-leading and leading order.

In Fig. 8.9, I show the trends of the three bias parameters of Fig.8.8 (Y-axis) as a function of r_{\min} (X-axis) for different sets of triangles $\eta_{\min} = 3, 2, 1$ (left, middle and central column panel, respectively). The blue and red colours indicate the results obtained with the leading order and the next-to-leading order moments. The two models generally provide consistent prediction except in the case of γ_2 for almost all r_{\min} values and η choices. For $\eta_{\min} = 1$, the two models provide different results also for b_2 but consistent results for b_1 (except when scales smaller than $30h^{-1}\text{Mpc}$ are included in the analysis. It is remarkable that, despite having many more free parameters, the next-to-leading order model provides constraints for b_2 that are tighter than those obtained with the leading order model.

Conclusions

This PhD Thesis has addressed the modelling of three-point statistics of Large Scale Structure (LSS) in configuration space, the three-point correlation function. Over the years, high-order statistics have been recognized as a fundamental tool, when used jointly with two-point statistics, in inferring cosmological parameters and removing their degeneracy from the spatial distribution of galaxies observed in spectroscopic redshift surveys. The analysis of the galaxy clustering properties is performed either in configuration space or in Fourier space, the latter approach being so far more popular for higher-order statistics. Measurements of the power spectrum and the bispectrum of galaxies in the BOSS surveys confronted with theoretical predictions have been used to increase the precision of the estimated cosmological parameters (180; 181; 182). Furthermore, the BAO signature has been detected in bispectrum analyses with high (4.1σ) statistical significance (183).

On the other hand, analyses in configuration space are less advanced. The BAO feature has indeed been detected in the galaxy 3PCF and used to derive cosmological parameters by considering scales that are evolving in the quasi-linear regime (72; 174). However, to fully exploit the potential of the analysis, one should really compare the measured 3PCF with the theoretical predictions on nonlinear scales. And these were not available yet before this Thesis project. This is regrettable since performing clustering analysis in configuration space offers some advantage over Fourier space since survey geometry and selection effects can be accounted for at the estimator level, with no need to perform difficult 3D convolution integrals. On the other hand, modelling higher-order statistics in configuration space is more challenging than in Fourier space due to the lack of a modelling approach directly in configuration space, given the operators that rule the equations of motions. Lacking this, modelling in configuration space can be done by inverse Fourier transforming existing models in Fourier space. These were precisely the goal and the strategy adopted in this Thesis, which aims at closing the gap between bispectrum and 3PCF models.

For this, I relied on the 2D-FFTLog algorithm (148) to evaluate the two-dimensional Hankel transform between the models of the bispectrum multipoles and the three-point

correlation function multipoles. Performing 2D-FFTL_{log} transform is feasible but computationally intensive since functions must be evaluated on a large 2D grid for the resulting 3PCF multipole models to match the accuracy of the bispectrum multipoles. To assess the adequacy of my approach and assess the accuracy of the 3PCF model, I compared the tree-level 3PCF model obtained using the 2D-FFTL_{log} technique to the reference tree-level model of (53) that only uses 1D-FFTL_{log} making use of analytical simplifications. I found that using a 256×256 2D-grid for the 2D-FFTL_{log} transform guarantees a remarkably good match between the two techniques, leading to a percentage error around 10^{-5} .

The 2D-FFTL_{log} represents a general method to convert any bispectrum into the corresponding counterpart in configuration space. In Chapter 6, this method has been applied to map the next-to-leading order matter bispectrum model into the new next-to-leading order three-point correlation function of dark matter fluctuations in real space. These model predictions have been tested against measurements extracted from the DEMNUni simulation (156) at $z = 0.49$ and $z = 1.05$ associated with Gaussian theoretical uncertainties. The main results of this analysis are:

- The differences between the next-to-leading and the leading order 3PCF model - dubbed SPT-NLO and LO, respectively - depend on the scale and the redshifts, as expected. The two models agree with each other on large scales, as they should. The matching scale, defined as the one in which the differences between model predictions are smaller than the expected Gaussian error, decreases with the redshift, as we have verified by considering the two DEMNUni snapshots at $z = 0.49$ and $z = 1.05$. Below this scale, the SPT-NLO model outperforms better than the LO one. To quantify the significance of the improvement, I have estimated the residuals of the two model predictions with respect to the 3PCF measured in the DEMNUni snapshots and compared them. At $z = 0.49$, the SPT-NLO residuals, averaged over all triangle configurations, are significantly smaller, at the $4\text{-}\sigma$ level, than the LO ones. A similar improvement is seen at $z = 1.05$, though with a lower ($2\text{-}\sigma$) statistical significance. The superiority of the SPT-NLO model over LO increases when moving toward smaller scales. For example, when considering triplets in which $r_{12} \leq 40 h^{-1}\text{Mpc}$ the improvement is as large as $6\text{-}\sigma$ at $z = 0.49$,
- The 3PCF next-to-leading model outperforms the leading-order one also on the BAO scale, whose importance for cosmological analyses cannot be overstressed. In particular, I focused on isosceles squeezed BAO configurations in which the smallest triangle side is $r_{12} = 17.5 h^{-1}\text{Mpc}$ that are sensitive to nonlinear effects. The reduced χ^2 of the LO-2D model vs. simulated data is 6(7) times larger than the SPT-NLO one at $z = 0.49(1.05)$.

- The 3PCF NLO-SPT and NLO-EFT models outperform the LO one also on the BAO scale, whose importance for cosmological analyses cannot be overstressed. In particular, we focused on isosceles squeezed BAO configurations in which the smallest triangle side is $r_{12} = 22.5 h^{-1}\text{Mpc}$ to probe the nonlinear regime. The quantitative comparison reveals that the reduced χ^2 of the LO model is 2 times larger than the NLO-SPT one at $z = 0.49, 1.05$, and 4 and 10 at $z = 0.49, 1.05$, respectively. However, from a visual inspection, neither models seem to perfectly match the shape (and the position) of the BAO peak in the 3PCF, but the correlation between measurements and errors makes it difficult to quantify the significance of the mismatch
- Two kinds of next-to-leading order models have been considered. The first one is based on standard perturbation theory (SPT-NLO). The second one is based instead on the effective field theory of the large-scale structure (EFT-NLO). This second model depends on four free parameters that have been estimated by minimizing the χ^2 function separately for both snapshots and for each choice of r_{\min} , the smallest size in each triangle configuration where the other two sizes span the range $[17.5, 132.5]h^{-1}\text{Mpc}$. Moreover, in our analysis, two sets of triangle configurations have been explored: the first one labelled $\eta_{\min} = 0$ includes all triangles, and the second one, $\eta_{\min} = 2$, in which $|r_{13} - r_{12}| \geq 10h^{-1}\text{Mpc}$. EFT-free parameters are estimated for these two cases separately too. It turns out that the best-fit EFT parameters are significantly different from zero only on the small scales range and larger scales starting approaching the BAO regime, and the precision of their estimate improves from $\eta_{\min} = 2$ to $\eta_{\min} = 0$,
- To assess the relative performance of the three models, SPT-NLO, EFT-NLO, and LO-SPT, we have estimated and compared their reduced χ^2 difference with respect to the DEMNUni measurements as a function of r_{\min} . For $\eta_{\min} = 2$ configurations, all models perform similarly on scales larger than $30h^{-1}\text{Mpc}$, whereas on smaller scales, both next-to-leading order models, SPT-NLO and EFT-NLO, outperform LO-SPT. However, this occurs on scales in which no model seems to provide a satisfactory good fit to the data, although the validity of this statement relies on the errors that, in the analysis, are assumed to be Gaussian. This hypothesis, valid on large scales, probably breaks down on the scales of the analysis, leading to an underestimate of the actual errors. A similar conclusion holds true for $\eta_{\min} = 0$, although in this case, the scale at which the linear order model under-performs with respect to the NLO ones is as large as $40h^{-1}\text{Mpc}$,
- Focusing on the relative performance of NLO-SPT and NLO-EFT, we find that the two models perform almost identically at $\eta_{\min} = 2$. For $\eta_{\min} = 0$, when all triangle

configurations are considered, the NLO-EFT model provides a better fit to the data than NLO-SPT in the range below $r_{\min} = 50h^{-1}\text{Mpc}$.

In Chapter 7, the focus of the project has been shifted to the distribution of luminous tracers and their biasing relation with the underlying matter. The same 2D-FFTLog transform strategy has been applied to model the leading-order three-point correlation function of the galaxies. The ability of these 3PCF models to constrain the bias parameters in a joint analysis with the next-to-leading order 2PCF models has been gauged against measurements performed in different snapshots extracted from the proprietary Euclid *Flagship* simulation mimicking the property of the upcoming Euclid galaxy survey at $z = 1.79, 1.53, 1.19, 0.90$. Gaussian errors were assumed in the comparison. The main results of the analysis are:

- The use of the galaxy 2PCF alone provides accurate estimates on the linear bias parameter b_1 only (having fixed the *rms* amplitude of the matter density fluctuations, σ_8) whose accuracy increases when decreasing the minimum scale of the analysis r_{\min} . However, the other bias parameters are weakly constrained and, in some cases, highly degenerate. The only one EFT parameter c_0 involved in the 2PCF model in this analysis is found to be well constrained, despite being consistent with zero,
- Constraining galaxy bias using the three-point correlation function alone provides one to constrain the higher-order bias parameters b_2, γ_2 . As expected, the agreement with bias relations being considered holds a larger number of triangle configurations as redshift increases,
- The combination of the two-point correlation function with the three-point correlation function in a joint analysis helps breaking degeneracy among some of the bias parameters in addition to improving the precision of their measurement. The expected agreement with theoretical expectations is verified on large scales. It is particularly good for $\gamma_2(b_1)$ and when the redshift increases.

Chapter 8 is dedicated to validating the new next-to-leading order 3PCF model for the halos rather than matter. Validation tests consisted in comparing the leading order and the next-to-leading order model predictions and in comparing both to the measured 3PCF of the dark matter halos extracted from the $z = 1$ snapshots extracted from the suite of MINERVA N-body simulations. To build the model, we have used the same 2D-FFTLog transform approach adopted for the matter case used, this time, to map the bias loops halos bispectrum model into the 3PCF. The accuracy of the bispectrum model we started with has been verified by comparing it to the reference model of (126) (details of the comparison are provided in Appendix A.2). We also included the next-to-leading order 2PCF model in the analysis since

it comes as a byproduct of the 3PCF one. Unlike in the other validation tests, we did not assume Gaussian errors. Instead, we have computed it numerically from the halo 2- and 3PCF measurements performed in all 298 MINERVA mocks. The main results of the analysis are:

- Like in the previous case, the 2PCF successfully constrain the linear bias parameters but not the higher-order bias parameters. The adoption of a large number of mocks together with the usage of numerical uncertainties associated to the average of the 2PCF estimates provide us with a better constraining power with respect to the case explored in Chapter 7, where the analytical covariance was adopted for considering uncertainties of a single mock. Scales such as $r_{\min} = 20, 30 h^{-1}\text{Mpc}$ affect the quality of the agreement between bias constraints and bias relations, which is conservatively assumed to be up $r_{\min} = 40 h^{-1}\text{Mpc}$,
- The value of the bias parameter $\gamma_2(b_1)$ inferred from the analysis that adopts next-to-leading order 3PCF model is in better agreement with expectations than the leading-order one on the scales in which both are reliable (151). Similarly, the precision with which the b_2 is measured is higher in the next-to-leading order case than is the leading order one, since it exhibits a narrower peak in the marginalized posterior distributions of b_2 , despite the large number of bias parameters that are left free to vary in the analysis. In the nonlinear regime ($r_{\min} \leq 30 h^{-1}\text{Mpc}$, $\eta_{\min} = 1$), the degeneracy among the parameters significantly reduces the benefit of using a next-to-leading order model,
- Lagrangian bias relations on the third and the four-order bias parameters allow us to reduce the dimensions of the parameter space and to break some degeneracy. Despite of this, the quality of the matching between $\gamma_2(b_1)$ and its expected value is degraded to a level similar to that obtained with the leading order model. On the other hand, their adoption helps tighten the constraints on the third order bias parameter b_3 , which turns out to be in agreement with the expected relation with b_1 $b_3(b_1)$,
- The combination of next-to-leading order two- and three- point correlation function models, as we have seen, breaks some degeneracy among the bias parameters. The analysis confirms the expectations and shows that b_1 can be measured with good precision, and definitively outperforms the leading order model when it comes to measuring $\gamma_2(b_1)$. It is a firm conclusion that a joint analysis provides significantly better constraints than using 2PCF and 3PCF alone.

The results presented in this Thesis represent an important step towards a nonlinear model for higher-order clustering statistics in configuration space, matching the quality of the

analogous models already available in Fourier space and precious tools to extract scientific information from next-generation spectroscopic galaxy catalogues containing tens of millions of objects. The 2D-FFTlog technique has been shown to be a fundamental tool for modelling the 3PCF of matter and galaxies. Its application is computationally intensive but can be dealt with by using standard computing facilities.

The expansion of the existing 3PCF models to the next-to-leading order of the perturbation expansion allows us to access small scales where precious information on the cosmological parameters and on the galaxy biasing is there to be collected. It also opens up the possibility to perform joint analyses in configuration space as well as in Fourier space also for 3-point statistics in order to mitigate the impact of systematic errors that affect either approach.

Moreover, clustering analyses based on spectroscopic galaxy catalogues use measured redshifts as a distant proxy. This generates "redshift space distortions" in the spatial mapping of the galaxies that, in turn, induces anisotropy in clustering statistics, including the 3PCF.

Finally, even with the availability of efficient estimators and, now, accurate nonlinear models, the use of the 3PCF for clustering analysis remains computationally challenging, preventing us from rapidly exploring the multidimensional space of the free parameters. For this, the usual strategy of probing the likelihood function using Monte Carlo Markov Chain techniques is too inefficient and will have to be complemented with emulators. Alternatively, a full machine learning approach, similar to the ones that are currently being implemented for two-point clustering analyses, should be explored.

These latter represent a very challenging aspect that needs to be properly accounted for and included in the 3PCF model proposed in this Thesis.

A

A.1 Theoretical covariance: two- and three- point correlation functions

Let D be a catalogue of data points and R a random set of points over the same volume. A general estimator can be defined as

$$D^p R^q = \sum_{i!=j} \Phi(x_1, \dots, x_p, y_1, \dots, y_p) \quad (\text{A.1})$$

where Φ is a generic function symmetric in its arguments and $p + q = N$ with N refers to the N -point statistics taken into consideration. For instance, taking the two-point case corresponds to

$$\Phi(x, y) = [x, y \in D, r \leq d(x, y) \leq r + dr], \quad (\text{A.2})$$

where $d(x, y)$ is the distance between two points. The general covariance of a pair of estimators is (153)

$$\langle D_a^{p_1} R_a^{q_1} D_b^{p_2} R_b^{q_2} \rangle = \sum_{i,j} \binom{p_1}{i} \binom{p_2}{i} i! \binom{q_1}{j} \binom{q_2}{j} j! S_{i+j} \lambda^{p_1+p_2-i} \rho^{q_1+q_2-j}, \quad (\text{A.3})$$

where

$$S_k = \int \Phi_a(x_1, \dots, x_k, y_{k+1}, \dots, y_N) \Phi_b(x_1, \dots, x_k, z_{k+1}, \dots, z_N) \mu_{2N-k} \quad (\text{A.4})$$

where a and b denote two possible different radial bins or even different statistics, and μ_s is the s dimensional Lebesgue measure. It has been shown (76; 152; 154) that the binned covariance matrix for the two-point correlation function, reducing the result to real space

monopole - our case of interest -, can be expressed as follows

$$\begin{aligned} \text{cov}(\hat{\xi}^a, \hat{\xi}^b) &= C^{ab} = \langle \hat{\xi}^a, \hat{\xi}^b \rangle - \langle \hat{\xi}^a \rangle \langle \hat{\xi}^b \rangle \\ &= {}^4C^{ab} + {}^3C^{ab} + {}^2C^{ab} \end{aligned} \quad (\text{A.5})$$

where the single terms read as

$$\begin{aligned} {}^4C_{cd}^{ab} &= \frac{1}{RR_c^a RR_d^b} \sum_{i \neq j \neq k \neq l} n_i n_j n_k n_l w_i w_j w_k w_l \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \Theta^b(r_{kl}) \Theta^d(\mu_{kl}), \\ &\quad \times \left[\xi_{ijkl}^{(4)} + 2\xi_{ik}\xi_{jl} \right] \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} {}^3C_{cd}^{ab} &= \frac{4}{RR_c^a RR_d^b} \sum_{i \neq j \neq k} n_i n_j n_k w_i (w_j)^2 w_k \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) \Theta^b(r_{jk}) \Theta^d(\mu_{jk}), \\ &\quad \times \left[\zeta_{ijk} + \xi_{ik} \right] \\ {}^2C_{cd}^{ab} &= \frac{2\delta^{ab}\delta^{cd}}{RR_c^a RR_d^b} \sum_{i \neq j} n_i n_j (w_i w_j)^2 \Theta^a(r_{ij}) \Theta^c(\mu_{ij}) [1 + \xi_{ij}], \end{aligned} \quad (\text{A.7})$$

with n_i , w_i the mean number density and weight and Θ^a being a binning function, i.e. unity if $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is inside the bin a and zero elsewhere. Also, the pre-factor involves the following quantities

$$RR_a = \sum_{i \neq j} R^{ij}, \quad (\text{A.8})$$

$$R_a^{ij} = \Theta^a(r_{ij}) n_i n_j w_i w_j, \quad (\text{A.9})$$

and the Gaussian terms have been highlighted in red. It has been shown in (155) that, under the Gaussian hypothesis, the computation of two-point correlation function covariance reduces to Eq. 5.42.

Concerning the three-point correlation function case, similarly to the previous case, it has been shown (184) that 3PCF covariance can be expressed as

$$\begin{aligned} \text{cov}(\hat{\zeta}_{\ell_1}^{ab}, \hat{\zeta}_{\ell_2}^{cd}) &= C_{\ell_1, \ell_2}^{ab, cd} = \langle \hat{\zeta}^{ab}, \hat{\zeta}^{cd} \rangle - \langle \hat{\zeta}^{ab} \rangle \langle \hat{\zeta}^{cd} \rangle \\ &= {}^6C^{ab, cd} + {}^5C^{ab, cd} + {}^4C^{ab, cd} + {}^3C^{ab, cd}, \end{aligned} \quad (\text{A.10})$$

where

$$\begin{aligned}
{}^6C_{\ell_1, \ell_2}^{ab, cd} &= \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{[6V(\overline{nw})^3]^2} \sum_{i \neq j \neq k \neq l \neq m \neq n} n_i n_j n_k n_l n_m n_n w_i w_j w_k w_l w_m w_n K_{ijk}^{ab, \ell_1} K_{lmn}^{cd, \ell_2} \\
&\times \left[\xi_{ijklmn}^{(6)} + 3\xi_{ij} \xi_{klmn}^{(4)} + 9\xi_{il} \xi_{jkmn}^{(4)} + 3\xi_{mn} \xi_{ijkl}^{(4)} + 9\xi_{ijl} \zeta_{kmn} + \right. \\
&\left. 9\xi_{ij} \xi_{kl} \xi_{mn} + 6\xi_{il} \xi_{jm} \xi_{kn} \right], \tag{A.11}
\end{aligned}$$

$$\begin{aligned}
{}^5C_{\ell_1, \ell_2}^{ab, cd} &= 9 \times \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{[6V(\overline{nw})^3]^2} \sum_{i \neq j \neq k \neq l \neq m} n_i n_j n_k n_l n_m w_i w_j (w_k)^2 w_l w_m K_{ijk}^{ab, \ell_1} K_{klm}^{cd, \ell_2} \\
&\times \left[\xi_{ijklm}^{(5)} + \xi_{ijlm}^{(4)} + 2\xi_{ik} \zeta_{jlm} + 2\xi_{kl} \zeta_{ijm} + \xi_{ij} \zeta_{klm} + 4\xi_{il} \zeta_{jkm} + \right. \\
&\left. \xi_{lm} \zeta_{ijk} + \xi_{ij} \xi_{lm} + 2\xi_{il} \xi_{jm} \right], \tag{A.12}
\end{aligned}$$

$$\begin{aligned}
{}^4C_{\ell_1, \ell_2}^{ab, cd} &= 18 \times \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{[6V(\overline{nw})^3]^2} \sum_{i \neq j \neq k \neq l} n_i n_j n_k n_l w_i (w_j w_k)^2 w_l K_{ijk}^{ab, \ell_1} K_{jkl}^{cd, \ell_2} \\
&\times \left[\xi_{ijkl}^{(4)} + 2\zeta_{ikl} + \xi_{il} \xi_{jk} + 2\xi_{ij} \xi_{kl} + \xi_{il} \right], \tag{A.13}
\end{aligned}$$

$$\begin{aligned}
{}^3C_{\ell_1, \ell_2}^{ab, cd} &= 6 \times \frac{(2\ell_1 + 1)(2\ell_2 + 1)}{[6V(\overline{nw})^3]^2} \sum_{i \neq j \neq k} n_i n_j n_k (w_i w_j w_k)^2 K_{ijk}^{ab, \ell_1} K_{ijk}^{cd, \ell_2} \\
&\times \left[\zeta_{ijk} + 3\xi_{ij} + 1 \right], \tag{A.14}
\end{aligned}$$

where, again, the terms highlighted in red are the Gaussian ones, and they contribute to all terms. The quantity $\overline{(nw)^3}$ is the survey-averaged value of $(nw)^3$, $v_a = 4\pi(r_{a, \max}^3 - r_{a, \min}^3)$ and the kernels involved are defined as

$$K_{ijk}^{ab, \ell} = \left[\Theta^a(r_{ij}) \Theta^b(r_{ik}) \mathcal{L}_\ell(\chi_{jk}) \Phi(r_a, r_b, \chi_{jk}) + 5 \text{ perms.} \right] \tag{A.15}$$

with χ_{jk} being the opening angle of a triangle of sides in bin i and j , and $\Phi(r_a, r_b, \chi_{jk})$ being the so-called *survey correction factor*, a function for each radial bin being able to encapsulate the effects of non-uniform sampling and the survey boundaries, usually approximated as a

two-dimensional smooth function (185). Under the Gaussian hypothesis, it is possible to derive simplified expressions as presented in Eq. 5.45.

A.2 Bias loop corrections to the Bispectrum: kernels

The kernels involved in the loop terms presented in Eqs 4.62, 4.63, and 4.64 are shown below

$$K(\mathbf{k}_1, \mathbf{k}_2) = \mu_{12}^2 - 1, \quad (\text{A.16})$$

$$L(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2\mu_{12}\mu_{23}\mu_{31} - \mu_{12}^2 - \mu_{23}^2 - \mu_{31}^2 + 1, \quad (\text{A.17})$$

where

$$\mathcal{K}_{\delta\mathcal{G}_2(\phi_2, \phi_1)}^{(4,F)} \equiv \mathcal{K}_{\delta\mathcal{G}_2(\phi_2, \phi_1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv \frac{1}{6} \left[K(\mathbf{k}_1, \mathbf{k}_2) + \text{sym}(6) \right], \quad (\text{A.18})$$

$$\mathcal{K}_{\mathcal{G}_3(\phi_3, \phi_1, \phi_1)}^{(4)} \equiv \mathcal{K}_{\mathcal{G}_3(\phi_3, \phi_1, \phi_1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv \frac{1}{6} \left[L(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_{34})K(\mathbf{k}_3, \mathbf{k}_4) + \text{sym}(6) \right], \quad (\text{A.19})$$

$$\mathcal{K}_{\mathcal{G}_2(\phi_2, \phi_2)}^{(4)} \equiv \mathcal{K}_{\mathcal{G}_2(\phi_2, \phi_2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv \frac{1}{3} \left[K(\mathbf{k}_{12}, \mathbf{k}_{34})K(\mathbf{k}_1, \mathbf{k})K(\mathbf{k}_3, \mathbf{k}_4) + \text{sym}(6) \right], \quad (\text{A.20})$$

$$\mathcal{K}_{\mathcal{G}_2(\phi_3, \phi_1)}^{(4)} \equiv \mathcal{K}_{\mathcal{G}_2(\phi_3, \phi_1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv \frac{1}{12} \left[\frac{1}{18} K(\mathbf{k}_1, \mathbf{k}_{234}) \right] \quad (\text{A.21})$$

$$\times \left(\frac{15}{7} K(\mathbf{k}_{23}, \mathbf{k}_4)K(\mathbf{k}_2, \mathbf{k}_3) - L(\mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \right) + \quad (\text{A.22})$$

$$\frac{1}{14} \left(M(\mathbf{k}_1, \mathbf{k}_{23}, \mathbf{k}_4, \mathbf{k}_{234}) - M(\mathbf{k}_1, \mathbf{k}_{234}, \mathbf{k}_{23}, \mathbf{k}_4) \right) K(\mathbf{k}_2, \mathbf{k}_3) + \text{sym}(12) \Big], \quad (\text{A.23})$$

where $\text{sym}(n)$ denoted the number of terms to be symmetrised.

A.3 Bias loop corrections to the Bispectrum: numerical integration

The next-to-leading order galaxy bispectrum model employed to calculate the corresponding 3PCF model in Chapter 8 was derived through the numerical implementation of loop integrals. The model comprises 46 integrals, each corresponding to a bias parameter triplet. These integrals are typically divided into four classes of integrals, as explained in Secs 4.2.1, 4.2.3. These classes of integrals exhibit a structural difference in the way they depend on

the linear power spectrum P_L . The P_L in B_{222} depends on both angular coordinates in a three-dimensional reference system. The P_L in B_{321-I} is dependent upon a single angle, and by analytically integrating along one angular coordinate, the integral to be numerically evaluated is reduced to two dimensions. Similarly, the P_L in B_{321-I} and B_{411} do not depend upon any angles. It is thus possible to analytically integrate along the two angular coordinates, resulting in one-dimensional integrals that can be numerically evaluated. In Figs A.1 A.2, A.3, and A.4, I show the computations of the aforementioned integrals, which have been computed using CUBA¹ and GSL-Gnu Scientific Library², and compared against a bispectrum reference (126) at the $z = 0$ for cosmology, as determined by the *Planck* mission citePlanck2020, with $\Omega_m = 0.319$, $\Omega_b = 0.049$, $\Omega_{de} = 0.681$, $\sigma_8 = 0.83$, $n_s = 0.96$ and $h = 0.67$. Values of k_1 , k_2 and k_3 as a function of the Index are shown in A.1.

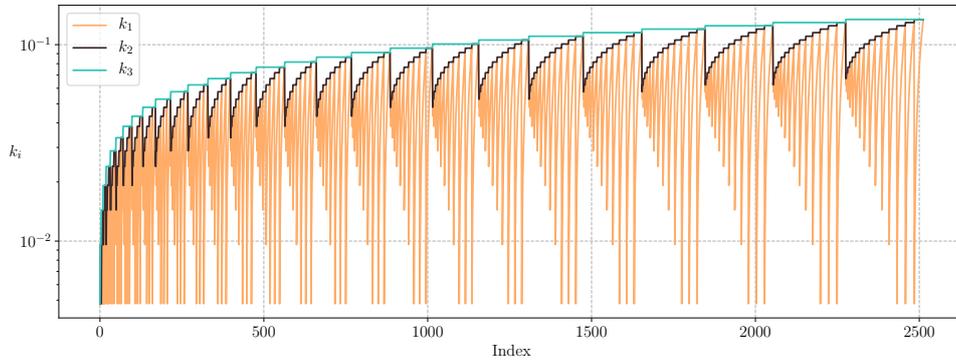
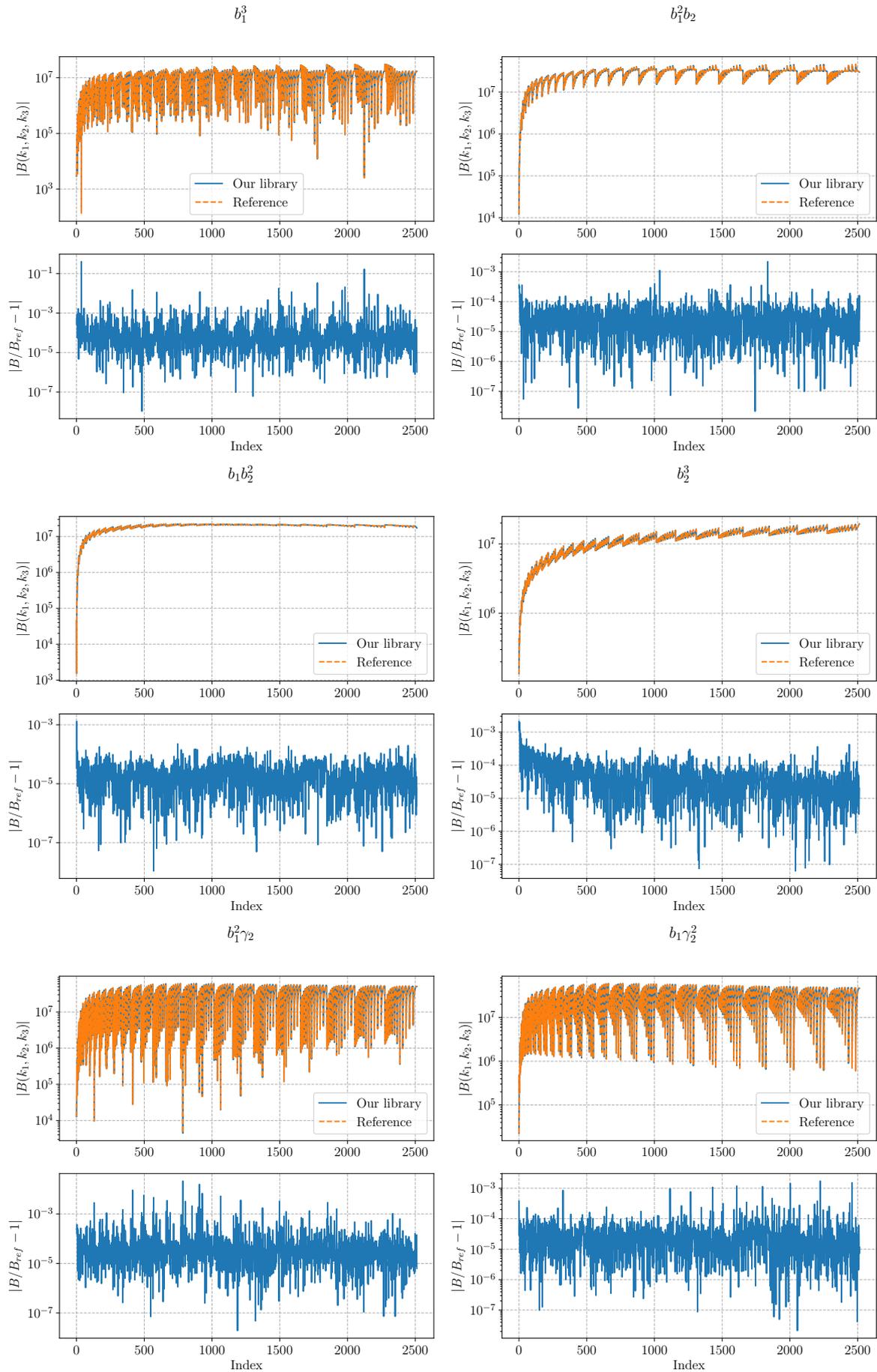


Figure A.1 k_1 , k_2 and k_3 - where $k_1 \leq k_2 \leq k_3$ - as a function of the Index.

¹feynarts.de/cuba

²gnu/software/gsl



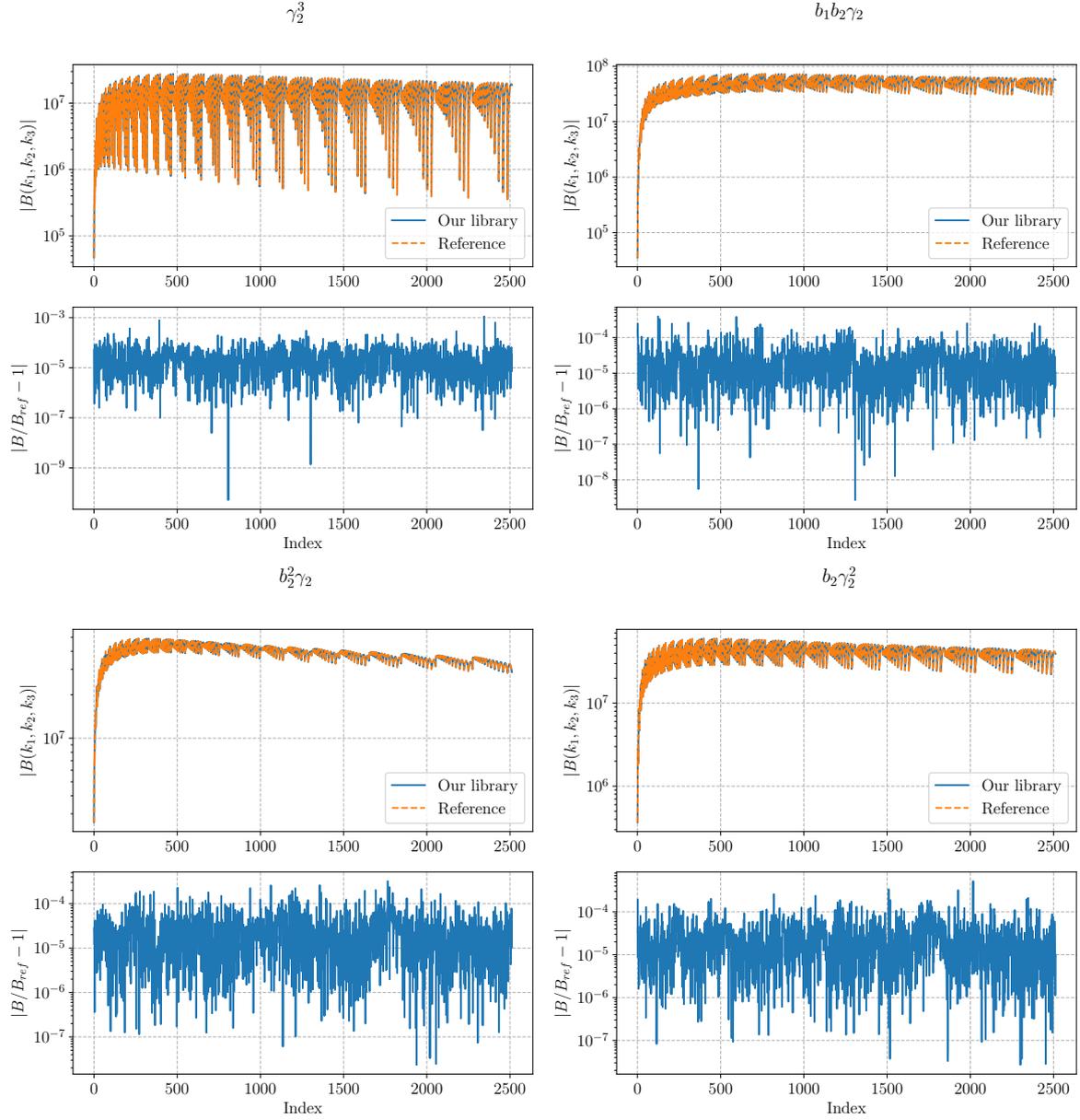
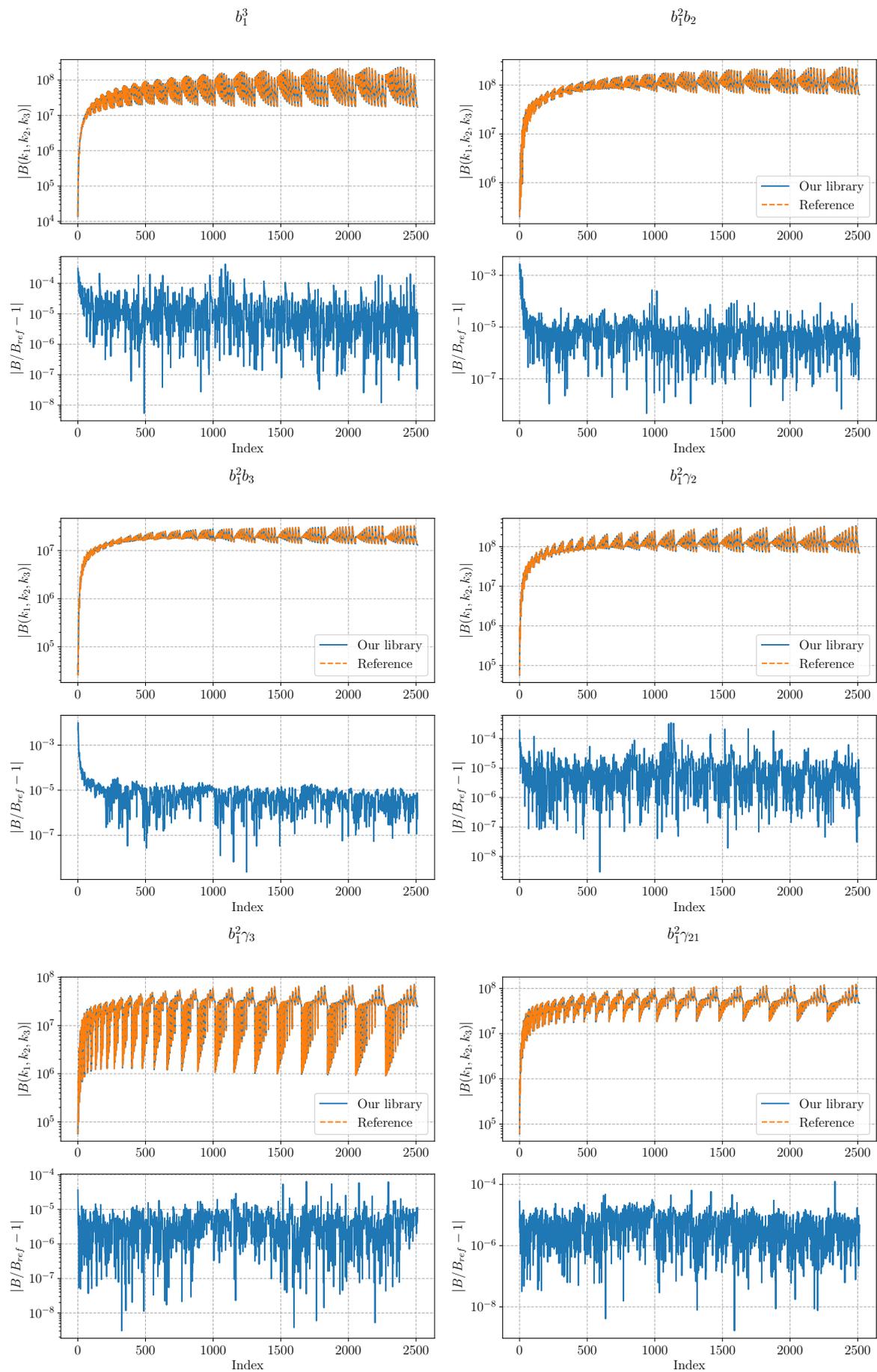
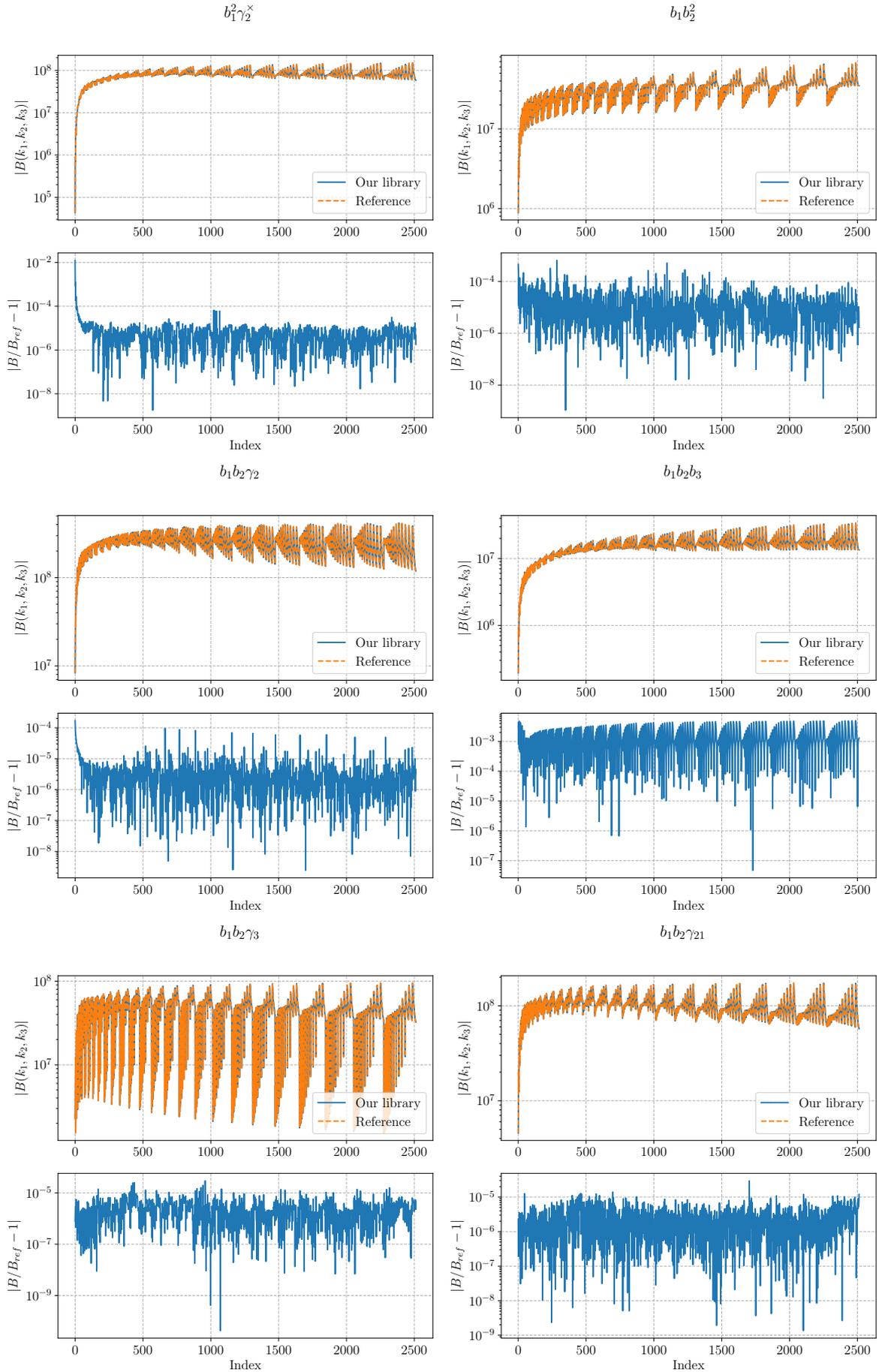
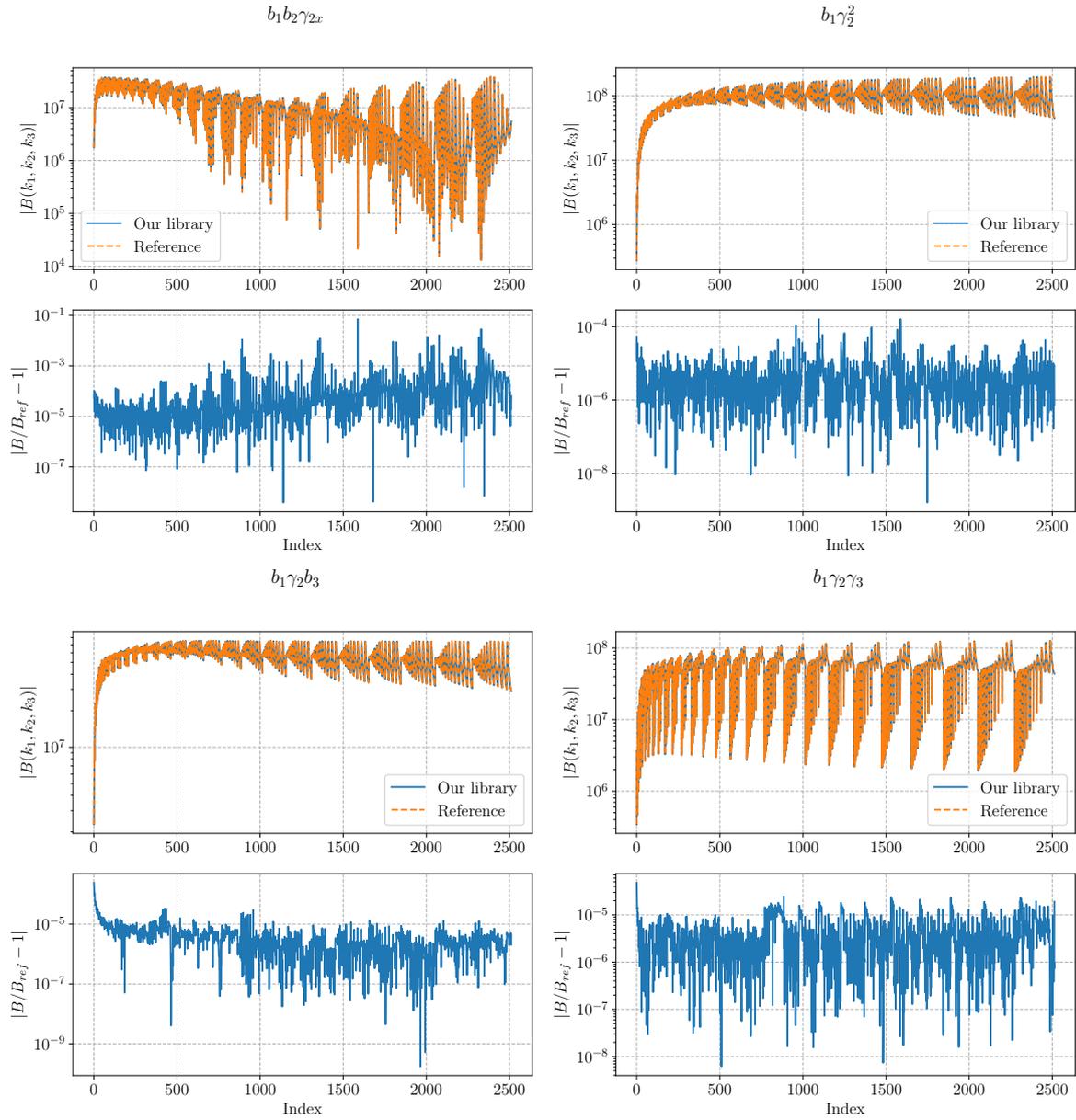


Figure A.1 Computation of individual contributions B_{222} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations.







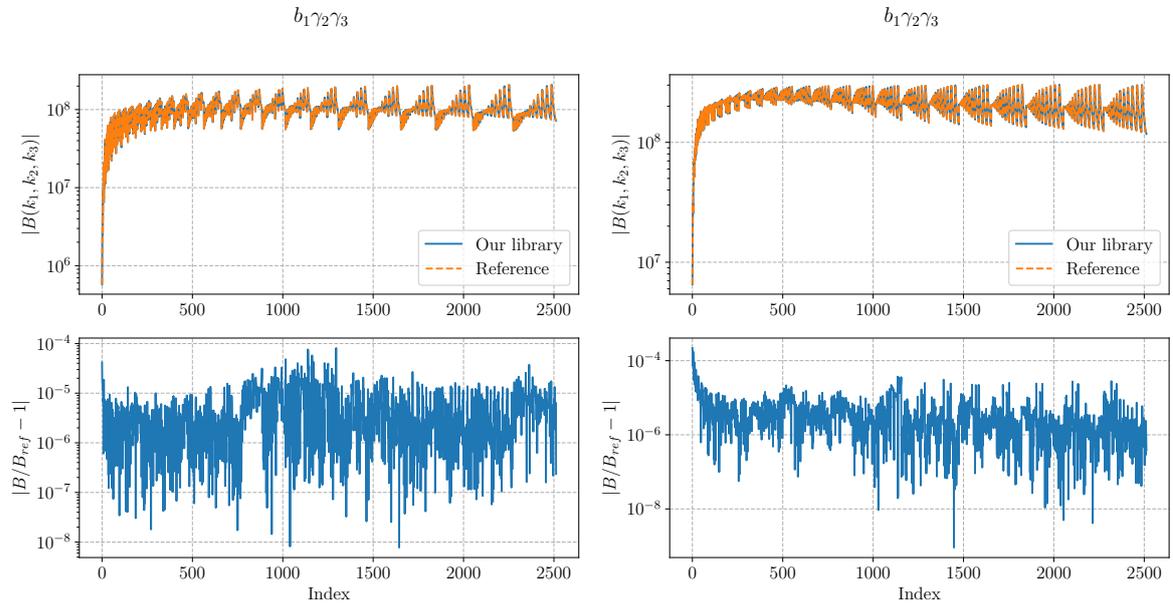
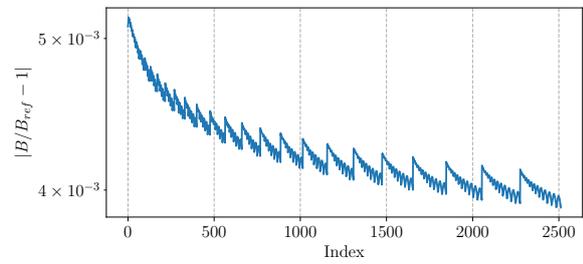
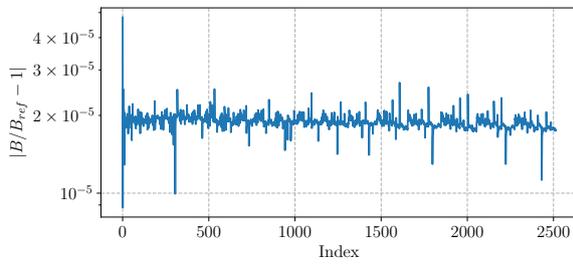
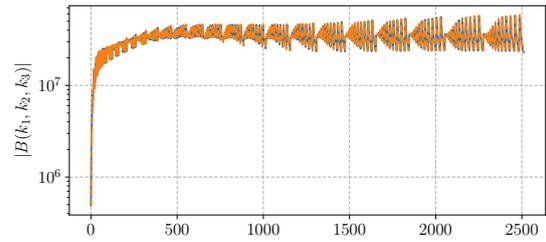
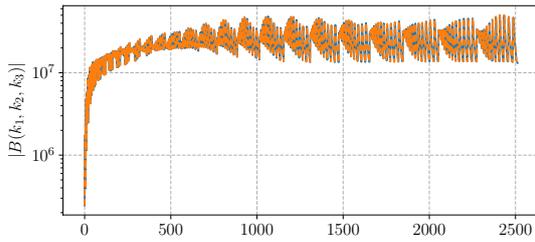
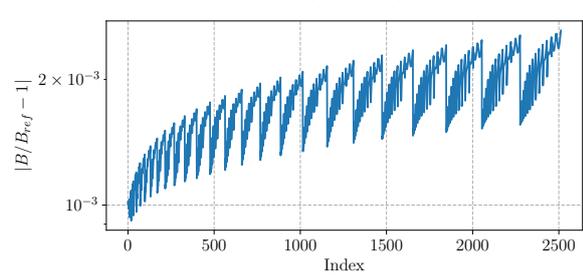
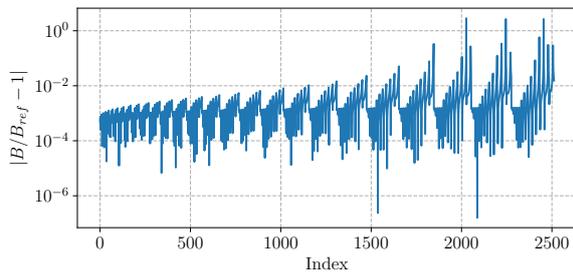
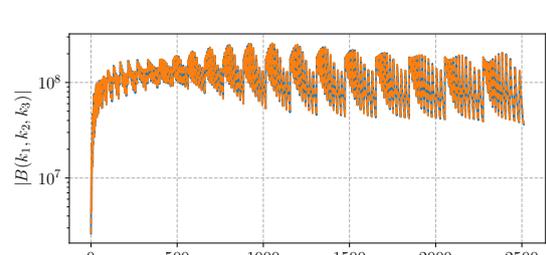
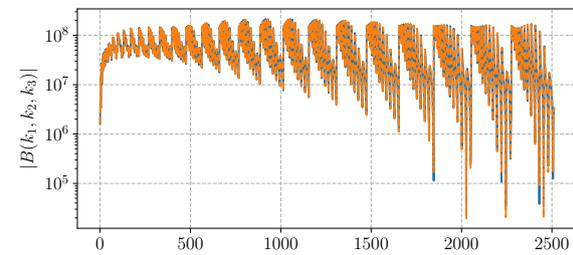


Figure A.2 Computation of individual contributions B_{321-1} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations.

$b_1^2 b_2$



$b_1^2 \gamma_2$



$b_1^2 \gamma_{21}$

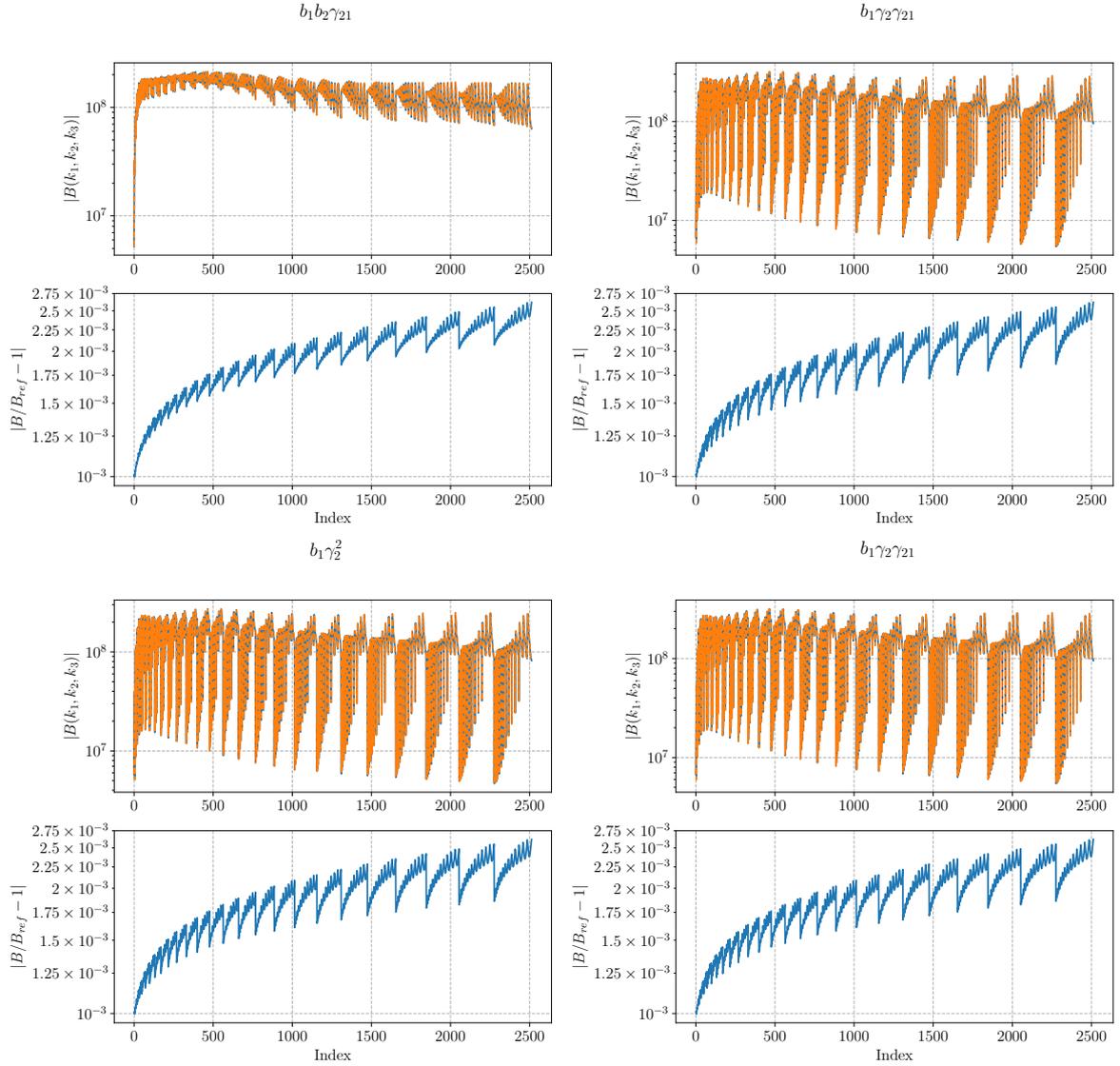
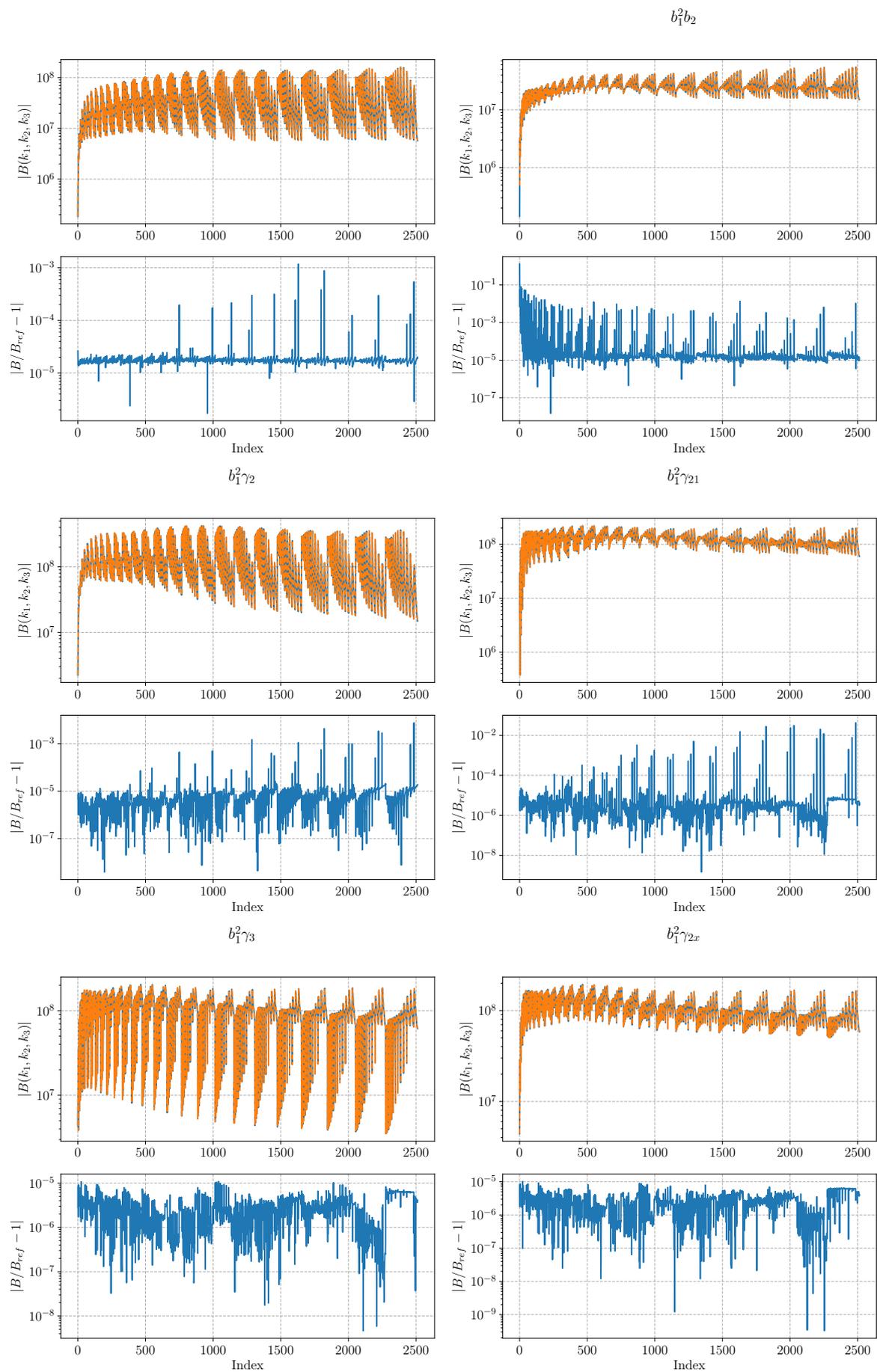


Figure A.3 Computation of individual contributions B_{321-II} of galaxy bispectrum at the next-to-leading order. For each plot: the *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while the *bottom plot* displays in blue the percentage differences between the two evaluations.



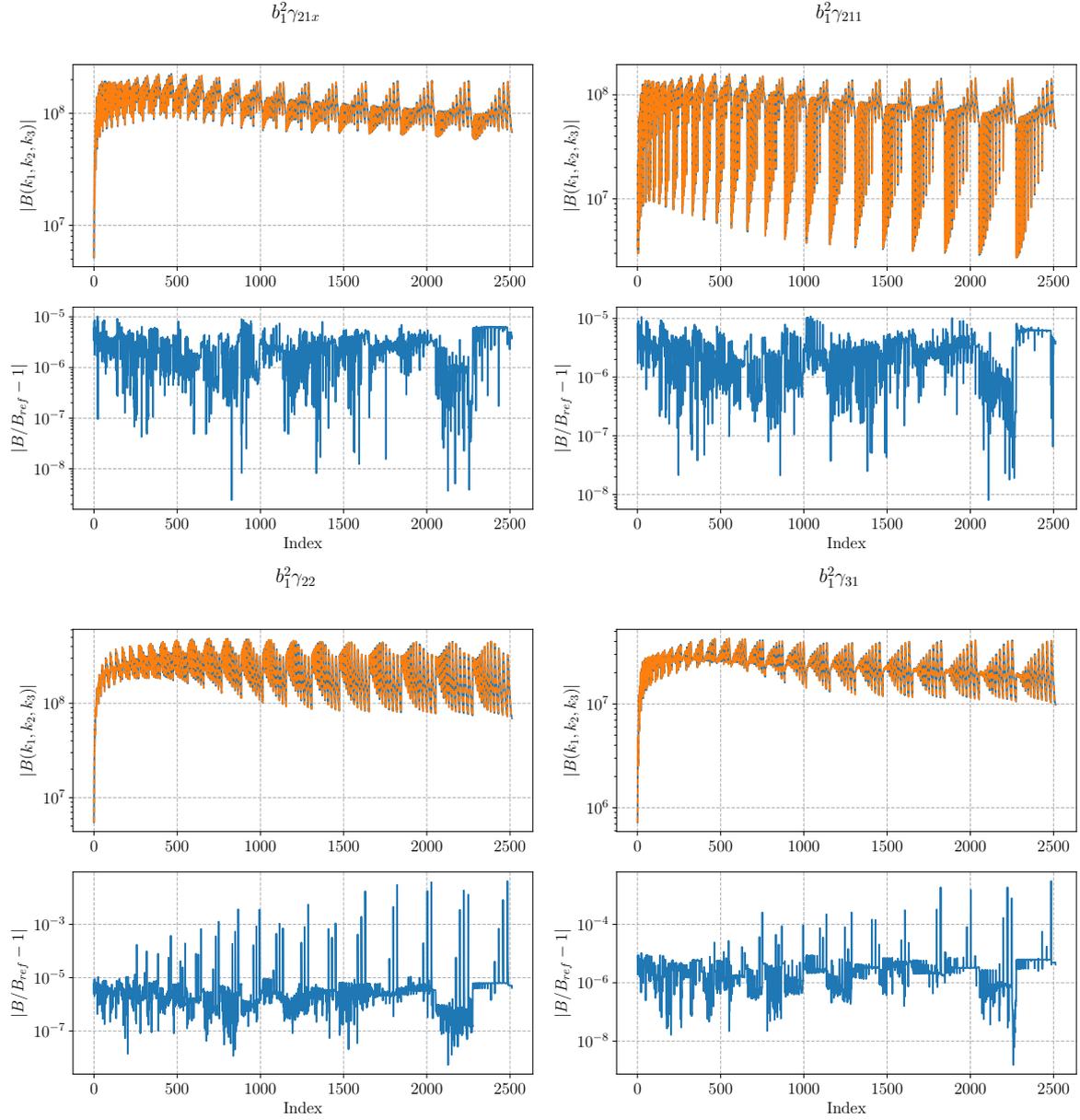


Figure A.4 Computation of individual contributions B_{411} of galaxy bispectrum at the next-to-leading order. For each plot: *top plot* refers to the absolute value of the individual contribution, where the blue line represents our computation, and the orange line represents the reference, while *bottom plot* displays in blue the percentage differences between the two evaluations.

References

- [1] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *Astrophys. J. Suppl.*, 208(2):20, October 2013.
- [2] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. Bobin, J. J. Bock, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, X. Chen, H. C. Chiang, L. Y. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, T. C. Gaier, S. Galeotta, S. Galli, K. Ganga, M. Giard, G. Giardino, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, D. Hanson, D. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, K. M. Huffmanberger, A. H. Jaffe, T. R. Jaffe, J. Jewell, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, R. J. Laureijs, C. R. Lawrence, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, J. Lesgourgues, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Cañiegos, P. M. Lubin,

- J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mandolesi, M. Maris, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, S. Osborne, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, D. Pearson, T. J. Pearson, H. V. Peiris, O. Perdureau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, J. L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, November 2014.
- [3] Planck Collaboration:, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff,

- M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* , 641:A6, September 2020.
- [4] Matthew Colless, Gavin Dalton, Steve Maddox, Will Sutherland, Peder Norberg, Shaun Cole, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Chris Collins, Warrick Couch, Nicholas Cross, Kathryn Deeley, Roberto De Propris, Simon P. Driver, George Efstathiou, Richard S. Ellis, Carlos S. Frenk, Karl Glazebrook, Carole Jackson, Ofer Lahav, Ian Lewis, Stuart Lumsden, Darren Madgwick, John A. Peacock, Bruce A. Peterson, Ian Price, Mark Seaborne, and Keith Taylor. The 2dF Galaxy Redshift Survey: spectra and redshifts. *Mon. Not. R. Astron. Soc.* , 328(4):1039–1063, December 2001.
- [5] Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey, Robert H. Barkhouser, Julian E. Bautista, Alessandra Beifiori, Andreas A. Berlind, Vaishali Bhardwaj, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, Adam S. Bolton, Arnaud Borde, Jo Bovy, W. N. Brandt, Howard Brewington, Jon Brinkmann, Peter J. Brown, Joel R. Brownstein, Kevin Bundy, N. G. Busca, William Carithers, Aurelio R. Carnero, Michael A. Carr, Yanmei Chen, Johan Comparat, Natalia Connolly, Frances Cope, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Timothée Delubac, Roland de Putter, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Daniel J. Eisenstein, S. Escoffier, Xiaohui Fan, N. Filiz Ak, Hayley Finley, Andreu Font-Ribera, R. Génova-

- Santos, James E. Gunn, Hong Guo, Daryl Haggard, Patrick B. Hall, Jean-Christophe Hamilton, Ben Harris, David W. Harris, Shirley Ho, David W. Hogg, Diana Holder, Klaus Honscheid, Joe Huehnerhoff, Beatrice Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal A. Kazin, David Kirkby, Mark A. Klaene, Jean-Paul Kneib, Jean-Marc Le Goff, Khee-Gan Lee, Daniel C. Long, Craig P. Loomis, Britt Lundgren, Robert H. Lupton, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Marc Manera, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena, Jordi Miralda-Escudé, Antonio D. Montero-Dorta, Francesco Montesano, Demitri Muna, Adam D. Myers, Tracy Naugle, Robert C. Nichol, Pasquier Noterdaeme, Sebastián E. Nuza, Matthew D. Olmstead, Audrey Oravetz, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Rafael Rebolo, James Rich, Gordon T. Richards, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, J. A. Rubiño-Martin, Lado Samushia, Ariel G. Sánchez, Conor Sayres, Sarah J. Schmidt, Donald P. Schneider, C. G. Scóccola, Hee-Jong Seo, Alaina Shelden, Erin Sheldon, Yue Shen, Yiping Shu, Anže Slosar, Stephen A. Smee, Stephanie A. Snedden, Fritz Stauffer, Oliver Steele, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Tomer Tal, Masayuki Tanaka, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Christy A. Tremonti, M. Vargas Magaña, Licia Verde, Matteo Viel, David A. Wake, Mike Watson, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Andrew A. West, Martin White, W. M. Wood-Vasey, Christophe Yèche, Idit Zehavi, Gong-Bo Zhao, and Zheng Zheng. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *Astron. J.* , 145(1):10, January 2013.
- [6] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, G. Allen, A. Allocca, P. A. Altin, A. Amato, A. Ananyeva, S. B. Anderson, W. G. Anderson, S. V. Angelova, S. Antier, S. Appert, K. Arai, M. C. Araya, J. S. Areeda, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, M. Ast, S. M. Aston, P. Astone, D. V. Atallah, P. Aufmuth, C. Aulbert, K. AultONeal, C. Austin, A. Avila-Alvarez, S. Babak, P. Bacon, M. K. M. Bader, S. Bae, M. Bailes, P. T. Baker, F. Baldaccini, G. Ballardín, S. W. Ballmer, S. Banagiri, J. C. Barayoga, S. E. Barclay, B. C. Bar-

ish, D. Barker, K. Barkett, F. Barone, B. Barr, L. Barsotti, M. Barsuglia, D. Barta, S. D. Barthelmy, J. Bartlett, I. Bartos, R. Bassiri, A. Basti, J. C. Batch, M. Bawaj, J. C. Bayley, M. Bazzan, B. Bécsy, C. Beer, M. Bejger, I. Belahcene, A. S. Bell, B. K. Berger, G. Bergmann, S. Bernuzzi, J. J. Bero, C. P. L. Berry, D. Bersanetti, A. Bertolini, J. Betzwieser, S. Bhagwat, R. Bhandare, I. A. Bilenko, G. Billingsley, C. R. Billman, J. Birch, R. Birney, O. Birnholtz, S. Biscans, S. Biscoveanu, A. Bisht, M. Bitossi, C. Biwer, M. A. Bizouard, J. K. Blackburn, J. Blackman, C. D. Blair, D. G. Blair, R. M. Blair, S. Bloemen, O. Bock, N. Bode, M. Boer, G. Bogaert, A. Bohe, F. Bondu, E. Bonilla, R. Bonnand, B. A. Boom, R. Bork, V. Boschi, S. Bose, K. Bossie, Y. Bouffanais, A. Bozzi, C. Bradaschia, P. R. Brady, M. Branchesi, J. E. Brau, T. Briant, A. Brillet, M. Brinkmann, V. Brisson, P. Brockill, J. E. Broida, A. F. Brooks, D. A. Brown, D. D. Brown, S. Brunett, C. C. Buchanan, A. Buikema, T. Bulik, H. J. Bulten, A. Buonanno, D. Buskulic, C. Buy, R. L. Byer, M. Cabero, L. Cadonati, G. Cagnoli, C. Cahillane, J. Calderón Bustillo, T. A. Callister, E. Calloni, J. B. Camp, M. Canepa, P. Canizares, K. C. Cannon, H. Cao, J. Cao, C. D. Capano, E. Capocasa, F. Carbognani, S. Caride, M. F. Carney, G. Carullo, J. Casanueva Diaz, C. Casentini, S. Caudill, M. Cavaglià, F. Cavalier, R. Cavalieri, G. Cella, C. B. Cepeda, P. Cerdá-Durán, G. Cerretani, E. Cesarini, S. J. Chamberlin, M. Chan, S. Chao, P. Charlton, E. Chase, E. Chassande-Mottin, D. Chatterjee, K. Chatziioannou, B. D. Cheeseboro, H. Y. Chen, X. Chen, Y. Chen, H. P. Cheng, H. Chia, A. Chincarini, A. Chiummo, T. Chmiel, H. S. Cho, M. Cho, J. H. Chow, N. Christensen, Q. Chu, A. J. K. Chua, S. Chua, A. K. W. Chung, S. Chung, G. Ciani, R. Ciolfi, C. E. Cirelli, A. Cirone, F. Clara, J. A. Clark, P. Clearwater, F. Cleva, C. Cocchieri, E. Coccia, P. F. Cohadon, D. Cohen, A. Colla, C. G. Collette, L. R. Cominsky, M. Constancio, L. Conti, S. J. Cooper, P. Corban, T. R. Corbitt, I. Cordero-Carrión, K. R. Corley, N. Cornish, A. Corsi, S. Cortese, C. A. Costa, M. W. Coughlin, S. B. Coughlin, J. P. Coulon, S. T. Countryman, P. Couvares, P. B. Covas, E. E. Cowan, D. M. Coward, M. J. Cowart, D. C. Coyne, R. Coyne, J. D. E. Creighton, T. D. Creighton, J. Cripe, S. G. Crowder, T. J. Cullen, A. Cumming, L. Cunningham, E. Cuoco, T. Dal Canton, G. Dálya, S. L. Danilishin, S. D'Antonio, K. Danzmann, A. Dasgupta, C. F. Da Silva Costa, V. Dattilo, I. Dave, M. Davier, D. Davis, E. J. Daw, B. Day, S. De, D. DeBra, J. Degallaix, M. De Laurentis, S. Deléglise, W. Del Pozzo, N. Demos, T. Denker, T. Dent, R. De Pietri, V. Dergachev, R. De Rosa, R. T. DeRosa, C. De Rossi, R. DeSalvo, O. de Varona, J. Devenson, S. Dhurandhar, M. C. Díaz, T. Dietrich, L. Di Fiore, M. Di Giovanni, T. Di Girolamo, A. Di Lieto, S. Di Pace, I. Di Palma, F. Di Renzo, Z. Doctor, V. Dolique, F. Donovan, K. L. Dooley, S. Doravari,

- I. Dorrington, R. Douglas, M. Dovale Álvarez, T. P. Downes, M. Drago, C. Dreissigacker, J. C. Driggers, Z. Du, M. Ducrot, R. Dudi, P. Dupej, S. E. Dwyer, T. B. Edo, M. C. Edwards, A. Effler, H. B. Eggenstein, P. Ehrens, J. Eichholz, S. S. Eikenberry, R. A. Eisenstein, R. C. Essick, D. Estevez, Z. B. Etienne, T. Etzel, M. Evans, T. M. Evans, M. Factourovich, V. Fafone, H. Fair, S. Fairhurst, X. Fan, S. Farinon, B. Farr, W. M. Farr, E. J. Fauchon-Jones, M. Favata, M. Fays, C. Fee, H. Fehrmann, J. Feicht, M. M. Fejer, A. Fernandez-Galiana, I. Ferrante, E. C. Ferreira, F. Ferrini, F. Fidecaro, D. Finstad, I. Fiori, D. Fiorucci, M. Fishbach, R. P. Fisher, M. Fitz-Axen, R. Flaminio, M. Fletcher, H. Fong, J. A. Font, P. W. F. Forsyth, S. S. Forsyth, J. D. Fournier, S. Frasca, F. Frasconi, Z. Frei, A. Freise, R. Frey, V. Frey, E. M. Fries, P. Fritschel, V. V. Frolov, P. Fulda, M. Fyffe, H. Gabbard, B. U. Gadre, S. M. Gaebel, J. R. Gair, L. Gammaitoni, M. R. Ganija, S. G. Gaonkar, C. Garcia-Quiros, F. Garufi, B. Gateley, S. Gaudio, G. Gaur, V. Gayathri, N. Gehrels, G. Gemme, E. Genin, A. Gennai, D. George, J. George, L. Gergely, V. Germain, S. Ghonge, Abhirup Ghosh, Archisman Ghosh, S. Ghosh, J. A. Giaime, K. D. Giardina, A. Giazotto, K. Gill, L. Glover, E. Goetz, R. Goetz, S. Gomes, B. Goncharov, G. González, J. M. Gonzalez Castro, A. Gopakumar, M. L. Gorodetsky, S. E. Gossan, M. Gosselin, R. Gouaty, A. Grado, C. Graef, M. Granata, A. Grant, S. Gras, C. Gray, G. Greco, A. C. Green, E. M. Gretarsson, P. Groot, H. Grote, S. Grunewald, P. Gruning, G. M. Guidi, X. Guo, A. Gupta, M. K. Gupta, K. E. Gushwa, E. K. Gustafson, R. Gustafson, O. Halim, B. R. Hall, E. D. Hall, E. Z. Hamilton, G. Hammond, M. Haney, M. M. Hanke, J. Hanks, C. Hanna, M. D. Hannam, O. A. Hannuksela, J. Hanson, T. Hardwick, J. Harms, G. M. Harry, I. W. Harry, M. J. Hart, C. J. Haster, K. Haughian, J. Healy, A. Heidmann, M. C. Heintze, H. Heitmann, P. Hello, G. Hemming, M. Hendry, I. S. Heng, J. Hennig, A. W. Heptonstall, M. Heurs, S. Hild, T. Hinderer, W. C. G. Ho, D. Hoak, D. Hofman, K. Holt, D. E. Holz, P. Hopkins, C. Horst, J. Hough, E. A. Houston, E. J. Howell, A. Hreibi, Y. M. Hu, E. A. Huerta, D. Huet, B. Hughey, S. Husa, S. H. Huttner, T. Huynh-Dinh, N. Indik, R. Inta, G. Intini, H. N. Isa, J. M. Isac, M. Isi, B. R. Iyer, K. Izumi, T. Jacqmin, K. Jani, P. Jaranowski, S. Jawahar, F. Jiménez-Forteza, W. W. Johnson, N. K. Johnson-McDaniel, D. I. Jones, R. Jones, R. J. G. Jonker, L. Ju, J. Junker, C. V. Kalaghatgi, V. Kalogera, B. Kamai, S. Kandhasamy, G. Kang, J. B. Kanner, S. J. Kapadia, S. Karki, K. S. Karvinen, M. Kasprzack, W. Kastaun, M. Katoлик, E. Katsavounidis, W. Katzman, S. Kaufer, K. Kawabe, F. Kéfélian, D. Keitel, A. J. Kemball, R. Kennedy, C. Kent, J. S. Key, F. Y. Khalili, I. Khan, S. Khan, Z. Khan, E. A. Khazanov, N. Kijbunchoo, Chunglee Kim, J. C. Kim, K. Kim, W. Kim, W. S. Kim, Y. M. Kim, S. J. Kimbrell, E. J. King, P. J. King, M. Kinley-Hanlon, R. Kirchhoff,

J. S. Kissel, L. Kleybolte, S. Klimenko, T. D. Knowles, P. Koch, S. M. Koehlenbeck, S. Koley, V. Kondrashov, A. Kontos, M. Korobko, W. Z. Korth, I. Kowalska, D. B. Kozak, C. Krämer, V. Kringel, B. Krishnan, A. Królak, G. Kuehn, P. Kumar, R. Kumar, S. Kumar, L. Kuo, A. Kutynia, S. Kwang, B. D. Lackey, K. H. Lai, M. Landry, R. N. Lang, J. Lange, B. Lantz, R. K. Lanza, S. L. Larson, A. Lartaux-Vollard, P. D. Lasky, M. Laxen, A. Lazzarini, C. Lazzaro, P. Leaci, S. Leavey, C. H. Lee, H. K. Lee, H. M. Lee, H. W. Lee, K. Lee, J. Lehmann, A. Lenon, E. Leon, M. Leonardi, N. Leroy, N. Letendre, Y. Levin, T. G. F. Li, S. D. Linker, T. B. Littenberg, J. Liu, X. Liu, R. K. L. Lo, N. A. Lockerbie, L. T. London, J. E. Lord, M. Lorenzini, V. Lorrlette, M. Lormand, G. Losurdo, J. D. Lough, C. O. Lousto, G. Lovelace, H. Lück, D. Lumaca, A. P. Lundgren, R. Lynch, Y. Ma, R. Macas, S. Macfoy, B. Machenschalk, M. MacInnis, D. M. Macleod, I. Magaña Hernandez, F. Magaña-Sandoval, L. Magaña Zertuche, R. M. Magee, E. Majorana, I. Maksimovic, N. Man, V. Mandic, V. Mangano, G. L. Mansell, M. Manske, M. Mantovani, F. Marchesoni, F. Marion, S. Márka, Z. Márka, C. Markakis, A. S. Markosyan, A. Markowitz, E. Maros, A. Marquina, P. Marsh, F. Martelli, L. Martellini, I. W. Martin, R. M. Martin, D. V. Martynov, J. N. Marx, K. Mason, E. Massera, A. Masserot, T. J. Massinger, M. Masso-Reid, S. Mastrogiovanni, A. Matas, F. Matichard, L. Matone, N. Mavalvala, N. Mazumder, R. McCarthy, D. E. McClelland, S. McCormick, L. McCuller, S. C. McGuire, G. McIntyre, J. McIver, D. J. McManus, L. McNeill, T. McRae, S. T. McWilliams, D. Meacher, G. D. Meadors, M. Mehmet, J. Meidam, E. Mejuto-Villa, A. Melatos, G. Mendell, R. A. Mercer, E. L. Merilh, M. Merzougui, S. Meshkov, C. Messenger, C. Messick, R. Metzdorff, P. M. Meyers, H. Miao, C. Michel, H. Middleton, E. E. Mikhailov, L. Milano, A. L. Miller, B. B. Miller, J. Miller, M. Millhouse, M. C. Milovich-Goff, O. Minazzoli, Y. Minenkov, J. Ming, C. Mishra, S. Mitra, V. P. Mitrofanov, G. Mitselmakher, R. Mittleman, D. Moffa, A. Moggi, K. Mogushi, M. Mohan, S. R. P. Mohapatra, I. Molina, M. Montani, C. J. Moore, D. Moraru, G. Moreno, S. Morisaki, S. R. Morriss, B. Mours, C. M. Mow-Lowry, G. Mueller, A. W. Muir, Arunava Mukherjee, D. Mukherjee, S. Mukherjee, N. Mukund, A. Mullavey, J. Munch, E. A. Muñiz, M. Muratore, P. G. Murray, A. Nagar, K. Napier, I. Nardecchia, L. Naticchioni, R. K. Nayak, J. Neilson, G. Nelemans, T. J. N. Nelson, M. Nery, A. Neunzert, L. Nevin, J. M. Newport, G. Newton, K. K. Y. Ng, P. Nguyen, T. T. Nguyen, D. Nichols, A. B. Nielsen, S. Nissanke, A. Nitz, A. Noack, F. Nocera, D. Nolting, C. North, L. K. Nuttall, J. Oberling, G. D. O’Dea, G. H. Ogin, J. J. Oh, S. H. Oh, F. Ohme, M. A. Okada, M. Oliver, P. Oppermann, Richard J. Oram, B. O’Reilly, R. Ormiston, L. F. Ortega, R. O’Shaughnessy, S. Ossokine, D. J. Ottaway, H. Overmier, B. J. Owen, A. E.

Pace, J. Page, M. A. Page, A. Pai, S. A. Pai, J. R. Palamos, O. Palashov, C. Palomba, A. Pal-Singh, Howard Pan, Huang-Wei Pan, B. Pang, P. T. H. Pang, C. Pankow, F. Panarale, B. C. Pant, F. Paoletti, A. Paoli, M. A. Papa, A. Parida, W. Parker, D. Pascucci, A. Pasqualetti, R. Passaquieti, D. Passuello, M. Patil, B. Patricelli, B. L. Pearlstone, M. Pedraza, R. Pedurand, L. Pekowsky, A. Pele, S. Penn, C. J. Perez, A. Perreca, L. M. Perri, H. P. Pfeiffer, M. Phelps, O. J. Piccinni, M. Pichot, F. Piergiovanni, V. Pierro, G. Pillant, L. Pinard, I. M. Pinto, M. Pirello, M. Pitkin, M. Poe, R. Poggiani, P. Popolizio, E. K. Porter, A. Post, J. Powell, J. Prasad, J. W. W. Pratt, G. Pratten, V. Predoi, T. Prestegard, M. Prijatelj, M. Principe, S. Privitera, R. Prix, G. A. Prodi, L. G. Prokhorov, O. Puncken, M. Punturo, P. Puppo, M. Pürner, H. Qi, V. Quetschke, E. A. Quintero, R. Quitzow-James, F. J. Raab, D. S. Rabeling, H. Radkins, P. Raffai, S. Raja, C. Rajan, B. Rajbhandari, M. Rakhmanov, K. E. Ramirez, A. Ramos-Buades, P. Rapagnani, V. Raymond, M. Razzano, J. Read, T. Regimbau, L. Rei, S. Reid, D. H. Reitze, W. Ren, S. D. Reyes, F. Ricci, P. M. Ricker, S. Rieger, K. Riles, M. Rizzo, N. A. Robertson, R. Robie, F. Robinet, A. Rocchi, L. Rolland, J. G. Rollins, V. J. Roma, J. D. Romano, R. Romano, C. L. Romel, J. H. Romie, D. Rosińska, M. P. Ross, S. Rowan, A. Rüdiger, P. Ruggi, G. Rutins, K. Ryan, S. Sachdev, T. Sadecki, L. Sadeghian, M. Sakellariadou, L. Salconi, M. Saleem, F. Salemi, A. Samajdar, L. Sammut, L. M. Sampson, E. J. Sanchez, L. E. Sanchez, N. Sanchis-Gual, V. Sandberg, J. R. Sanders, B. Sassolas, B. S. Sathyaprakash, P. R. Saulson, O. Sauter, R. L. Savage, A. Sawadsky, P. Schale, M. Scheel, J. Scheuer, J. Schmidt, P. Schmidt, R. Schnabel, R. M. S. Schofield, A. Schönbeck, E. Schreiber, D. Schuette, B. W. Schulte, B. F. Schutz, S. G. Schwalbe, J. Scott, S. M. Scott, E. Seidel, D. Sellers, A. S. Sengupta, D. Sentenac, V. Sequino, A. Sergeev, D. A. Shaddock, T. J. Shaffer, A. A. Shah, M. S. Shahriar, M. B. Shaner, L. Shao, B. Shapiro, P. Shawhan, A. Sheperd, D. H. Shoemaker, D. M. Shoemaker, K. Siellez, X. Siemens, M. Sieniawska, D. Sigg, A. D. Silva, L. P. Singer, A. Singh, A. Singhal, A. M. Sintes, B. J. J. Slagmolen, B. Smith, J. R. Smith, R. J. E. Smith, S. Somala, E. J. Son, J. A. Sonnenberg, B. Sorazu, F. Sorrentino, T. Souradeep, A. P. Spencer, A. K. Srivastava, K. Staats, A. Staley, M. Steinke, J. Steinlechner, S. Steinlechner, D. Steinmeyer, S. P. Stevenson, R. Stone, D. J. Stops, K. A. Strain, G. Stratta, S. E. Strigin, A. Strunk, R. Sturani, A. L. Stuver, T. Z. Summerscales, L. Sun, S. Sunil, J. Suresh, P. J. Sutton, B. L. Swinkels, M. J. Szczepańczyk, M. Tacca, S. C. Tait, C. Talbot, D. Talukder, D. B. Tanner, M. Tápai, A. Taracchini, J. D. Tasson, J. A. Taylor, R. Taylor, S. V. Tewari, T. Theeg, F. Thies, E. G. Thomas, M. Thomas, P. Thomas, K. A. Thorne, K. S. Thorne, E. Thrane, S. Tiwari, V. Tiwari, K. V. Tokmakov, K. Toland, M. Tonelli, Z. Tornasi, A. Torres-Forné, C. I. Torrie, D. Töyrä,

- F. Travasso, G. Traylor, J. Trinastic, M. C. Tringali, L. Trozzo, K. W. Tsang, M. Tse, R. Tso, L. Tsukada, D. Tsuna, D. Tuyenbayev, K. Ueno, D. Ugolini, C. S. Unnikrishnan, A. L. Urban, S. A. Usman, H. Vahlbruch, G. Vajente, G. Valdes, M. Vallisneri, N. van Bakel, M. van Beuzekom, J. F. J. van den Brand, C. Van Den Broeck, D. C. Vander-Hyde, L. van der Schaaf, J. V. van Heijningen, A. A. van Veggel, M. Vardaro, V. Varma, S. Vass, M. Vasúth, A. Vecchio, G. Vedovato, J. Veitch, P. J. Veitch, K. Venkateswara, G. Venugopalan, D. Verkindt, F. Vetrano, A. Viceré, A. D. Viets, S. Vinciguerra, D. J. Vine, J. Y. Vinet, S. Vitale, T. Vo, H. Vocca, C. Vorvick, S. P. Vyatchanin, A. R. Wade, L. E. Wade, M. Wade, R. Walet, M. Walker, L. Wallace, S. Walsh, G. Wang, H. Wang, J. Z. Wang, W. H. Wang, Y. F. Wang, R. L. Ward, J. Warner, M. Was, J. Watchi, B. Weaver, L. W. Wei, M. Weinert, A. J. Weinstein, R. Weiss, L. Wen, E. K. Wessel, P. Weßels, J. Westerweck, T. Westphal, K. Wette, J. T. Whelan, S. E. Whitcomb, B. F. Whiting, C. Whittle, D. Wilken, D. Williams, R. D. Williams, A. R. Williamson, J. L. Willis, B. Willke, M. H. Wimmer, W. Winkler, C. C. Wipf, H. Wittel, G. Woan, J. Woehler, J. Wofford, K. W. K. Wong, J. Worden, J. L. Wright, D. S. Wu, D. M. Wysocki, S. Xiao, H. Yamamoto, C. C. Yancey, L. Yang, M. J. Yap, M. Yazback, Hang Yu, Haocun Yu, M. Yvert, A. Zadrożny, M. Zanolin, T. Zelenova, J. P. Zendri, M. Zevin, L. Zhang, M. Zhang, T. Zhang, Y. H. Zhang, C. Zhao, M. Zhou, Z. Zhou, S. J. Zhu, X. J. Zhu, A. B. Zimmerman, M. E. Zucker, J. Zweizig, LIGO Scientific Collaboration, and Virgo Collaboration. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.* , 119(16):161101, October 2017.
- [7] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas Macri, Beatrice Bucciarelli, Mario G. Lattanzi, John W. MacKenty, J. Bradley Bowers, WeiKang Zheng, Alexei V. Filippenko, Caroline Huang, and Richard I. Anderson. Milky way cepheid standards for measuring cosmic distances and application to Implications for the hubble constant. *The Astrophysical Journal*, 861(2):126, jul 2018.
- [8] Kevork N. Abazajian, Peter Adshead, Zeeshan Ahmed, Steven W. Allen, David Alonso, Kam S. Arnold, Carlo Baccigalupi, James G. Bartlett, Nicholas Battaglia, Bradford A. Benson, Colin A. Bischoff, Julian Borrill, Victor Buza, Erminia Calabrese, Robert Caldwell, John E. Carlstrom, Clarence L. Chang, Thomas M. Crawford, Francis-Yan Cyr-Racine, Francesco De Bernardis, Tijmen de Haan, Sperello di Serego Alighieri, Joanna Dunkley, Cora Dvorkin, Josquin Errard, Giulio Fabbian, Stephen Feeney, Simone Ferraro, Jeffrey P. Filippini, Raphael Flauger, George M. Fuller, Vera Gluscevic, Daniel Green, Daniel Grin, Evan Grohs, Jason W. Henning, J. Colin

- Hill, Renee Hlozek, Gilbert Holder, William Holzzapfel, Wayne Hu, Kevin M. Huffenberger, Reijo Keskitalo, Lloyd Knox, Arthur Kosowsky, John Kovac, Ely D. Kovetz, Chao-Lin Kuo, Akito Kusaka, Maude Le Jeune, Adrian T. Lee, Marc Lilley, Marilena Loverde, Mathew S. Madhavacheril, Adam Mantz, David J. E. Marsh, Jeffrey McMahon, Pieter Daniel Meerburg, Joel Meyers, Amber D. Miller, Julian B. Munoz, Ho Nam Nguyen, Michael D. Niemack, Marco Peloso, Julien Peloton, Levon Pogosian, Clement Pryke, Marco Raveri, Christian L. Reichardt, Graca Rocha, Aditya Rotti, Emmanuel Schaan, Marcel M. Schmittfull, Douglas Scott, Neelima Sehgal, Sarah Shandera, Blake D. Sherwin, Tristan L. Smith, Lorenzo Sorbo, Glenn D. Starkman, Kyle T. Story, Alexander van Engelen, Joaquin D. Vieira, Scott Watson, Nathan Whitehorn, and W. L. Kimmy Wu. CMB-S4 Science Book, First Edition. *arXiv e-prints*, page arXiv:1610.02743, October 2016.
- [9] Samaya Nissanke, Daniel E. Holz, Scott A. Hughes, Neal Dalal, and Jonathan L. Sievers. Exploring Short Gamma-ray Bursts as Gravitational-wave Standard Sirens. *Astrophys. J.* , 725(1):496–514, December 2010.
- [10] Stefano Casertano, Adam G. Riess, Jay Anderson, Richard I. Anderson, J. Bradley Bowers, Kelsey I. Clubb, Aviv R. Cukierman, Alexei V. Filippenko, Melissa L. Graham, John W. MacKenty, Carl Melis, Brad E. Tucker, and Gautam Upadhyya. Parallax of Galactic Cepheids from Spatially Scanning the Wide Field Camera 3 on the Hubble Space Telescope: The Case of SS Canis Majoris. *Astrophys. J.* , 825(1):11, July 2016.
- [11] Jose María Ezquiaga and Miguel Zumalacárregui. Dark Energy in light of Multi-Messenger Gravitational-Wave astronomy. *Frontiers in Astronomy and Space Sciences*, 5:44, December 2018.
- [12] Catherine Heymans, Tilman Tröster, Marika Asgari, Chris Blake, Hendrik Hildebrandt, Benjamin Joachimi, Konrad Kuijken, Chieh-An Lin, Ariel G. Sánchez, Jan Luca van den Busch, Angus H. Wright, Alexandra Amon, Maciej Bilicki, Jelte de Jong, Martin Crocce, Andrej Dvornik, Thomas Erben, Maria Cristina Fortuna, Fedor Getman, Benjamin Giblin, Karl Glazebrook, Henk Hoekstra, Shahab Joudaki, Arun Kannawadi, Fabian Köhlinger, Chris Lidman, Lance Miller, Nicola R. Napolitano, David Parkinson, Peter Schneider, HuanYuan Shan, Edwin A. Valentijn, Gijs Verdoes Kleijn, and Christian Wolf. KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *Astron. Astrophys.* , 646:A140, February 2021.

- [13] E. W. Kolb. Dynamics of the Inflationary Era. Technical Report, DE2001-14350; FERMILAB-CONF-99/303-A, November 1999.
- [14] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. Large-scale galaxy bias. *Physics Reports*, 733:1–193, feb 2018.
- [15] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the universe and cosmological perturbation theory. *Phys. Rep.*, 367(1-3):1–248, September 2002.
- [16] Emiliano Sefusatti and Román Scoccimarro. Galaxy bias and halo-occupation numbers from large-scale clustering. *Phys. Rev. D* , 71(6):063001, March 2005.
- [17] Zachary Slepian and Daniel J. Eisenstein. Computing the three-point correlation function of galaxies in $\mathcal{O}(N^2)$ time. *Mon. Not. R. Astron. Soc.* , 454(4):4142–4158, December 2015.
- [18] Edwin Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, March 1929.
- [19] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophys. J.* , 142:419–421, July 1965.
- [20] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, January 1933.
- [21] Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *Astrophys. J.* , 159:379, February 1970.
- [22] J. P. Ostriker, P. J. E. Peebles, and A. Yahil. The Size and Mass of Galaxies, and the Mass of the Universe. *Astrophys. J. Lett.* , 193:L1, October 1974.
- [23] Jaan Einasto, Ants Kaasik, and Enn Saar. Dynamic evidence on massive coronas of galaxies. *Nature*, 250(5464):309–310, July 1974.
- [24] V. C. Rubin, Jr. Ford, W. K., and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *Astrophys. J.* , 238:471–487, June 1980.

- [25] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE Differential Microwave Radiometer First-Year Maps. *Astrophys. J. Lett.* , 396:L1, September 1992.
- [26] C. L. Bennett, A. J. Banday, K. M. Gorski, G. Hinshaw, P. Jackson, P. Keegstra, A. Kogut, G. F. Smoot, D. T. Wilkinson, and E. L. Wright. Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results. *Astrophys. J. Lett.* , 464:L1, June 1996.
- [27] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astron. J.* , 116(3):1009–1038, September 1998.
- [28] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *Astrophys. J.* , 517(2):565–586, June 1999.
- [29] Ya. B. Zel'Dovich. Cosmological Constant and Elementary Particles. *Soviet Journal of Experimental and Theoretical Physics Letters*, 6:316, November 1967.
- [30] Ya B. Zel'dovich. Special Issue: the Cosmological Constant and the Theory of Elementary Particles. *Soviet Physics Uspekhi*, 11(3):381–393, March 1968.
- [31] Bharat Ratra and P. J. E. Peebles. Cosmological consequences of a rolling homogeneous scalar field. *Phys. Rev. D*, 37:3406–3427, Jun 1988.
- [32] Timothy Clifton, Pedro G. Ferreira, Antonio Padilla, and Constantinos Skordis. Modified gravity and cosmology. *Phys. Rep.*, 513(1):1–189, March 2012.

- [33] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM. *Astrophys. J.* , 876(1):85, May 2019.
- [34] Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A. Blazek, Adam S. Bolton, Joel R. Brownstein, Angela Burden, Chia-Hsun Chuang, Johan Comparat, Antonio J. Cuesta, Kyle S. Dawson, Daniel J. Eisenstein, Stephanie Escoffier, Héctor Gil-Marín, Jan Niklas Grieb, Nick Hand, Shirley Ho, Karen Kinemuchi, David Kirkby, Francisco Kitaura, Elena Malanushenko, Viktor Malanushenko, Claudia Maraston, Cameron K. McBride, Robert C. Nichol, Matthew D. Olmstead, Daniel Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, Marcos Pellejero-Ibanez, Will J. Percival, Patrick Petitjean, Francisco Prada, Adrian M. Price-Whelan, Beth A. Reid, Sergio A. Rodríguez-Torres, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Jose Alberto Rubiño-Martín, Shun Saito, Salvador Salazar-Albornoz, Lado Samushia, Ariel G. Sánchez, Siddharth Satpathy, David J. Schlegel, Donald P. Schneider, Claudia G. Scóccola, Hee-Jong Seo, Erin S. Sheldon, Audrey Simmons, Anže Slosar, Michael A. Strauss, Molly E. C. Swanson, Daniel Thomas, Jeremy L. Tinker, Rita Tojeiro, Mariana Vargas Magaña, Jose Alberto Vazquez, Licia Verde, David A. Wake, Yuting Wang, David H. Weinberg, Martin White, W. Michael Wood-Vasey, Christophe Yèche, Idit Zehavi, Zhongxu Zhai, and Gong-Bo Zhao. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. R. Astron. Soc.* , 470(3):2617–2652, September 2017.
- [35] Tilman Tröster, Ariel G. Sánchez, Marika Asgari, Chris Blake, Martín Crocce, Catherine Heymans, Hendrik Hildebrandt, Benjamin Joachimi, Shahab Joudaki, Arun Kannawadi, Chieh-An Lin, and Angus Wright. Cosmology from large-scale structure. Constraining Λ CDM with BOSS. *Astron. Astrophys.* , 633:L10, January 2020.
- [36] M. A. Troxel, N. MacCrann, J. Zuntz, T. F. Eifler, E. Krause, S. Dodelson, D. Gruen, J. Blazek, O. Friedrich, S. Samuroff, J. Prat, L. F. Secco, C. Davis, A. Ferté, J. DeRose, A. Alarcon, A. Amara, E. Baxter, M. R. Becker, G. M. Bernstein, S. L. Bridle, R. Cawthon, C. Chang, A. Choi, J. De Vicente, A. Drlica-Wagner, J. Elvin-Poole, J. Frieman, M. Gatti, W. G. Hartley, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, M. Jarvis, T. Kacprzak, D. Kirk, N. Kokron, C. Krawiec, O. Lahav, A. R. Liddle, J. Peacock, M. M. Rau, A. Refregier, R. P. Rollins, E. Rozo, E. S. Rykoff, C. Sánchez, I. Sevilla-Noarbe, E. Sheldon, A. Stebbins, T. N. Varga, P. Vielzeuf,

- M. Wang, R. H. Wechsler, B. Yanny, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, E. Fernandez, B. Flaugher, P. Fosalba, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. A. Goldstein, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. J. James, T. Jeltema, M. W. G. Johnson, M. D. Johnson, S. Kent, K. Kuehn, S. Kuhlmann, N. Kuropatkin, T. S. Li, M. Lima, H. Lin, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, E. Neilsen, R. C. Nichol, B. Nord, D. Petravick, A. A. Plazas, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, V. Vikram, A. R. Walker, J. Weller, Y. Zhang, and DES Collaboration. Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear. *Phys. Rev. D* , 98(4):043528, August 2018.
- [37] J. Huchra, M. Davis, D. Latham, and J. Tonry. A survey of galaxy redshifts. IV - The data. *Astrophys. J. Suppl.* , 52:89–119, June 1983.
- [38] Will J. Percival, Carlton M. Baugh, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Shaun Cole, Matthew Colless, Chris Collins, Warrick Couch, Gavin Dalton, Roberto De Propris, Simon P. Driver, George Efstathiou, Richard S. Ellis, Carlos S. Frenk, Karl Glazebrook, Carole Jackson, Ofer Lahav, Ian Lewis, Stuart Lumsden, Steve Maddox, Stephen Moody, Peder Norberg, John A. Peacock, Bruce A. Peterson, Will Sutherland, and Keith Taylor. The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *Mon. Not. R. Astron. Soc.* , 327(4):1297–1306, November 2001.
- [39] Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, Francisco J. Castander, Bing Chen, Patrick L. Colestock, A. J. Connolly, J. H. Crocker, István Csabai, Paul C. Czarapata, John Eric Davis, Mamoru Doi, Tom Dombeck, Daniel Eisenstein, Nancy Ellman, Brian R. Elms, Michael L. Evans, Xiaohui Fan, Glenn R. Federwitz, Larry Fiscelli, Scott Friedman, Joshua A. Frieman, Masataka Fukugita, Bruce Gillespie, James E. Gunn, Vijay K. Gurbani, Ernst de Haas, Merle Haldeman, Frederick H. Harris, J. Hayes, Timothy M. Heckman, G. S. Hennessy, Robert B. Hindsley, Scott Holm,

- Donald J. Holmgren, Chi-hao Huang, Charles Hull, Don Husby, Shin-Ichi Ichikawa, Takashi Ichikawa, Željko Ivezić, Stephen Kent, Rita S. J. Kim, E. Kinney, Mark Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, John Korienek, Richard G. Kron, Peter Z. Kunszt, D. Q. Lamb, B. Lee, R. French Leger, Siriluk Limmongkol, Carl Lindenmeyer, Daniel C. Long, Craig Loomis, Jon Loveday, Rich Lucinio, Robert H. Lupton, Bryan MacKinnon, Edward J. Mannery, P. M. Mantsch, Bruce Margon, Peregrine McGehee, Timothy A. McKay, Avery Meiksin, Aronne Merelli, David G. Monet, Jeffrey A. Munn, Vijay K. Narayanan, Thomas Nash, Eric Neilsen, Rich Neswold, Heidi Jo Newberg, R. C. Nichol, Tom Nicinski, Mario Nonino, Norio Okada, Sadanori Okamura, Jeremiah P. Ostriker, Russell Owen, A. George Pauls, John Peoples, R. L. Peterson, Donald Petravick, Jeffrey R. Pier, Adrian Pope, Ruth Pordes, Angela Prosapio, Ron Rechenmacher, Thomas R. Quinn, Gordon T. Richards, Michael W. Richmond, Claudio H. Rivetta, Constance M. Rockosi, Kurt Ruthmansdorfer, Dale Sandford, David J. Schlegel, Donald P. Schneider, Maki Sekiguchi, Gary Sergey, Kazuhiro Shimasaku, Walter A. Siegmund, Stephen Smee, J. Allyn Smith, S. Snedden, R. Stone, Chris Stoughton, Michael A. Strauss, Christopher Stubbs, Mark SubbaRao, Alexander S. Szalay, Istvan Szapudi, Gyula P. Szokoly, Anirudda R. Thakar, Christy Tremonti, Douglas L. Tucker, Alan Uomoto, Dan Vanden Berk, Michael S. Vogeley, Patrick Waddell, Shu-i. Wang, Masaru Watanabe, David H. Weinberg, Brian Yanny, Naoki Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. *Astron. J.* , 120(3):1579–1587, September 2000.
- [40] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligorì, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio,

- F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Noddle, M. Poncet, M. Scodreggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid Definition Study Report. *arXiv e-prints*, October 2011.
- [41] Luca Amendola, Stephen Appleby, David Bacon, Tessa Baker, Marco Baldi, Nicola Bartolo, Alain Blanchard, Camille Bonvin, Stefano Borgani, Enzo Branchini, Clare Burrage, Stefano Camera, Carmelita Carbone, Luciano Casarini, Mark Cropper, Claudia de Rham, Cinzia Di Porto, Anne Ealet, Pedro G. Ferreira, Fabio Finelli, Juan García-Bellido, Tommaso Giannantonio, Luigi Guzzo, Alan Heavens, Lavinia Heisenberg, Catherine Heymans, Henk Hoekstra, Lukas Hollenstein, Rory Holmes, Ole Horst, Knud Jahnke, Thomas D. Kitching, Tomi Koivisto, Martin Kunz, Giuseppe La Vacca, Marisa March, Elisabetta Majerotto, Katarina Markovic, David Marsh, Federico Marulli, Richard Massey, Yannick Mellier, David F. Mota, Nelson J. Nunes, Will Percival, Valeria Pettorino, Cristiano Porciani, Claudia Quercellini, Justin Read, Massimiliano Rinaldi, Domenico Sapone, Roberto Scaramella, Constantinos Skordis, Fergus Simpson, Andy Taylor, Shaun Thomas, Roberto Trotta, Licia Verde, Filippo Vernizzi, Adrian Vollmer, Yun Wang, Jochen Weller, and Tom Zlosnik. Cosmology and Fundamental Physics with the Euclid Satellite. *Living Reviews in Relativity*, 16(1):6, September 2013.

- [42] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balleland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekhazadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrellius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaughner, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Kopysov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L’Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J.

- Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyang Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yèche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*, October 2016.
- [43] Hu Zhan and J. Anthony Tyson. Cosmology with the Large Synoptic Survey Telescope: an overview. *Reports on Progress in Physics*, 81(6):066901, June 2018.
- [44] Tim Eifler, Hironao Miyatake, Elisabeth Krause, Chen Heinrich, Vivian Miranda, Christopher Hirata, Jiachuan Xu, Shoubaneh Hemmati, Melanie Simet, Peter Capak, Ami Choi, Olivier Doré, Cyrille Doux, Xiao Fang, Rebekah Hounsell, Eric Huff, Hung-Jin Huang, Mike Jarvis, Jeffrey Kruk, Dan Masters, Eduardo Rozo, Dan Scolnic, David N. Spergel, Michael Troxel, Anja von der Linden, Yun Wang, David H. Weinberg, Lukas Wenzl, and Hao-Yi Wu. Cosmology with the Roman Space Telescope - multiprobe strategies. *Mon. Not. R. Astron. Soc.*, 507(2):1746–1761, October 2021.

- [45] Olivier Doré, Jamie Bock, Matthew Ashby, Peter Capak, Asantha Cooray, Roland de Putter, Tim Eifler, Nicolas Flagey, Yan Gong, Salman Habib, Katrin Heitmann, Chris Hirata, Woong-Seob Jeong, Raj Katti, Phil Korngut, Elisabeth Krause, Dae-Hee Lee, Daniel Masters, Phil Mauskopf, Gary Melnick, Bertrand Mennesson, Hien Nguyen, Karin Öberg, Anthony Pullen, Alvise Raccanelli, Roger Smith, Yong-Seon Song, Volker Tolls, Steve Unwin, Tejaswi Venumadhav, Marco Viero, Mike Werner, and Mike Zemcov. Cosmology with the SPHEREX All-Sky Spectral Survey. *arXiv e-prints*, page arXiv:1412.4872, December 2014.
- [46] P. J. E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.
- [47] Román Scoccimarro, Hume A. Feldman, J. N. Fry, and Joshua A. Frieman. The Bispectrum of IRAS Redshift Catalogs. *Astrophys. J.* , 546(2):652–664, January 2001.
- [48] Masahiro Takada and Bhuvnesh Jain. The three-point correlation function in cosmology. *Mon. Not. R. Astron. Soc.* , 340(2):580–608, April 2003.
- [49] E. Gaztañaga, P. Norberg, C. M. Baugh, and D. J. Croton. Statistical analysis of galaxy surveys - II. The three-point galaxy correlation function measured from the 2dFGRS. *Mon. Not. R. Astron. Soc.* , 364(2):620–634, December 2005.
- [50] Jun Pan and István Szapudi. The monopole moment of the three-point correlation function of the two-degree Field Galaxy Redshift Survey. *Mon. Not. R. Astron. Soc.* , 362(4):1363–1370, October 2005.
- [51] Felipe A. Marín, Risa H. Wechsler, Joshua A. Frieman, and Robert C. Nichol. Modeling the Galaxy Three-Point Correlation Function. *Astrophys. J.* , 672(2):849–860, January 2008.
- [52] Cameron K. McBride, Andrew J. Connolly, Jeffrey P. Gardner, Ryan Scranton, Jeffrey A. Newman, Román Scoccimarro, Idit Zehavi, and Donald P. Schneider. Three-point Correlation Functions of SDSS Galaxies: Luminosity and Color Dependence in Redshift and Projected Space. *Astrophys. J.* , 726(1):13, January 2011.
- [53] Z. Slepian and D. J. Eisenstein. Modelling the large-scale redshift-space 3-point correlation function of galaxies. *Mon. Not. R. Astron. Soc.* , 469:2059–2076, August 2017.
- [54] Emiliano Sefusatti, Martín Crocce, Sebastián Pueblas, and Román Scoccimarro. Cosmology and the bispectrum. *Phys. Rev. D* , 74(2):023522, July 2006.

- [55] Román Scoccimarro. Cosmological Perturbations: Entering the Nonlinear Regime. *Astrophys. J.* , 487(1):1–17, September 1997.
- [56] Román Scoccimarro, Stéphane Colombi, J. N. Fry, Joshua A. Frieman, Eric Hivon, and Adrian Melott. Nonlinear Evolution of the Bispectrum of Cosmological Perturbations. *Astrophys. J.* , 496(2):586–604, March 1998.
- [57] E. Sefusatti, M. Crocce, and V. Desjacques. The matter bispectrum in N-body simulations with non-Gaussian initial conditions. *Mon. Not. R. Astron. Soc.* , 406(2):1014–1028, August 2010.
- [58] Andrei Lazanu and Michele Liguori. The two and three-loop matter bispectrum in perturbation theories. *Journal of Cosmology and Astro-Particle Physics*, 2018(4):055, April 2018.
- [59] Benjamin Bose and Atsushi Taruya. The one-loop matter bispectrum as a probe of gravity and dark energy. *Journal of Cosmology and Astro-Particle Physics*, 2018(10):019, October 2018.
- [60] Davit Alkhanishvili, Cristiano Porciani, Emiliano Sefusatti, Matteo Biagetti, Andrei Lazanu, Andrea Oddo, and Victoria Yankelevich. The reach of next-to-leading-order perturbation theory for the matter bispectrum. *Mon. Not. R. Astron. Soc.* , 512(4):4961–4981, June 2022.
- [61] Martín Crocce and Román Scoccimarro. Renormalized cosmological perturbation theory. *Phys. Rev. D* , 73(6):063519, March 2006.
- [62] Francis Bernardeau, Martín Crocce, and Román Scoccimarro. Multipoint propagators in cosmological gravitational instability. *Phys. Rev. D* , 78(10):103521, November 2008.
- [63] Francis Bernardeau, Martín Crocce, and Román Scoccimarro. Constructing regularized cosmic propagators. *Phys. Rev. D* , 85(12):123519, June 2012.
- [64] Takahiko Matsubara. Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space. *Phys. Rev. D* , 77(6):063530, March 2008.
- [65] John Joseph M. Carrasco, Simon Foreman, Daniel Green, and Leonardo Senatore. The Effective Field Theory of Large Scale Structures at two loops. *Journal of Cosmology and Astro-Particle Physics*, 2014(7):057, July 2014.

- [66] Mark P. Hertzberg. Effective field theory of dark matter and structure formation: Semianalytical results. *Phys. Rev. D* , 89(4):043521, February 2014.
- [67] Raul E. Angulo, Simon Foreman, Marcel Schmittfull, and Leonardo Senatore. The One-Loop Matter Bispectrum in the Effective Field Theory of Large Scale Structures. *JCAP*, 10:039, 2015.
- [68] Tobias Baldauf, Lorenzo Mercolli, and Matias Zaldarriaga. Effective field theory of large scale structure at two loops: The apparent scale dependence of the speed of sound. *Phys. Rev. D* , 92(12):123007, December 2015.
- [69] Y. P. Jing and G. Boerner. The three-point correlation function of galaxies determined from the 2df galaxy redshift survey. *Astrophys. J.*, 607:140–163, 2004.
- [70] M. Moresco, F. Marulli, L. Moscardini, E. Branchini, A. Cappi, I. Davidzon, B. R. Granett, S. de la Torre, L. Guzzo, U. Abbas, C. Adami, S. Arnouts, J. Bel, M. Bolzonella, D. Bottini, C. Carbone, J. Coupon, O. Cucciati, G. De Lucia, P. Franzetti, A. Fritz, M. Fumana, B. Garilli, O. Ilbert, A. Iovino, J. Krywult, V. Le Brun, O. Le Fèvre, K. Małek, H. J. McCracken, M. Polletta, A. Pollo, M. Scodreggio, L. A. M. Tasca, R. Tojeiro, D. Vergani, and A. Zanichelli. The VIMOS Public Extragalactic Redshift Survey (VIPERS) . Exploring the dependence of the three-point correlation function on stellar mass and luminosity at $0.5 < z < 1.1$. *Astron. Astrophys.* , 604:A133, August 2017.
- [71] Michele Moresco, Alfonso Veropalumbo, Federico Marulli, Lauro Moscardini, and Andrea Cimatti. C^3 : Cluster Clustering Cosmology. II. First Detection of the Baryon Acoustic Oscillations Peak in the Three-point Correlation Function of Galaxy Clusters. *Astrophys. J.* , 919(2):144, October 2021.
- [72] Zachary Slepian et al. Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies. *Mon. Not. R. Astron. Soc.* , 469(2):1738–1751, 2017.
- [73] Naonori S. Sugiyama, Shun Saito, Florian Beutler, and Hee-Jong Seo. Towards a self-consistent analysis of the anisotropic galaxy two- and three-point correlation functions on large scales: application to mock galaxy catalogues. *Mon. Not. R. Astron. Soc.* , 501(2):2862–2896, 2021.
- [74] Alfonso Veropalumbo, Iñigo Sáez Casares, Enzo Branchini, Benjamin R. Granett, Luigi Guzzo, Federico Marulli, Michele Moresco, Lauro Moscardini, Andrea Pezzotta,

- and Sylvain de la Torre. A joint 2- and 3-point clustering analysis of the VIPERS PDR2 catalogue at $z \approx 1$: breaking the degeneracy of cosmological parameters. *Mon. Not. R. Astron. Soc.*, 507(1):1184–1201, October 2021.
- [75] Kevin Pardede, Federico Rizzo, Matteo Biagetti, Emanuele Castorina, Emiliano Seufusatti, and Pierluigi Monaco. Bispectrum-window convolution via Hankel transform. *JCAP*, 10:066, 2022.
- [76] Oliver H. E. Philcox. Cosmology without window functions. II. Cubic estimators for the galaxy bispectrum. *Phys. Rev. D*, 104(12):123529, December 2021.
- [77] Zachary Slepian and Daniel J. Eisenstein. A practical computational method for the anisotropic redshift-space three-point correlation function. *Mon. Not. R. Astron. Soc.*, 478(2):1468–1483, August 2018.
- [78] Naonori S. Sugiyama, Shun Saito, Florian Beutler, and Hee-Jong Seo. A complete FFT-based decomposition formalism for the redshift-space bispectrum. *Mon. Not. R. Astron. Soc.*, 484(1):364–384, March 2019.
- [79] David H. Lyth and Andrew R. Liddle. *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*. Cambridge University Press, 2009.
- [80] Antonio Riotto. Inflation and the theory of cosmological perturbations, 2002.
- [81] D. Gorbunov and Rubakov Valery A. Introduction to the theory of early Universe. Cosmological Perturbations and Inflationary theory. World Scientific, 2011.
- [82] Alexander Eggemeier. Challenges and prospects of probing galaxy clustering with three-point statistics. 2018.
- [83] Valentin Assassi, Daniel Baumann, Daniel Green, and Matias Zaldarriaga. Renormalized halo bias. *Journal of Cosmology and Astroparticle Physics*, 2014(08):056–056, aug 2014.
- [84] Alfonso Veropalumbo. Clustering of clusters as a cosmological probe. 2016.
- [85] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, L. P. L. Colombo, C. Combet, D. Contreras,

- B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, J. Ferguson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, E. Hivon, D. C. Hooper, Z. Huang, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, V. Lindholm, M. López-Cañiego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. D. Meerburg, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, S. Mitra, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, M. Münchmeyer, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, M. Shiraishi, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, S. D. M. White, A. Zacchei, J. P. Zibin, and A. Zonca. Planck 2018 results. X. Constraints on inflation. *Astron. Astrophys.* , 641:A10, September 2020.
- [86] Viviana Acquaviva, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto. Gauge-invariant second-order perturbations and non-gaussianity from inflation. *Nuclear Physics B*, 667(1-2):119–148, sep 2003.
- [87] Juan Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 2003(05):013–013, may 2003.
- [88] Ya. B. Zel’dovich. Gravitational instability: An approximate theory for large density perturbations. *Astron. Astrophys.* , 5:84–89, March 1970.
- [89] T. Buchert, A. L. Melott, and A. G. Weiss. Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models. *Astron. Astrophys.* ,

- 288:349–364, August 1994.
- [90] Paolo Catelan. Lagrangian dynamics in non-flat universes and non-linear gravitational evolution. *Mon. Not. R. Astron. Soc.* , 276(1):115–124, September 1995.
- [91] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz. Perturbative Lagrangian approach to gravitational instability. *Astron. Astrophys.* , 296:575, April 1995.
- [92] Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Phys. Rep.*, 372(1):1–129, December 2002.
- [93] Anatoly Klypin, Francisco Prada, Gustavo Yepes, Steffen Hess, and Stefan Gottlober. Halo Abundance Matching: accuracy and conditions for numerical convergence. *arXiv e-prints*, page arXiv:1310.3740, October 2013.
- [94] James E. Gunn and III Gott, J. Richard. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *Astrophys. J.* , 176:1, August 1972.
- [95] D. Lynden-Bell. Statistical mechanics of violent relaxation in stellar systems. *Mon. Not. R. Astron. Soc.* , 136:101, January 1967.
- [96] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophys. J.* , 187:425–438, February 1974.
- [97] J. A. Peacock and A. F. Heavens. Alternatives to the Press–Schechter cosmological mass function. *Monthly Notices of the Royal Astronomical Society*, 243(1):133–143, 10 1990.
- [98] Ravi K. Sheth and Giuseppe Tormen. An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *Monthly Notices of the Royal Astronomical Society*, 329(1):61–75, 01 2002.
- [99] Ravi K. Sheth, H. J. Mo, and Giuseppe Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 323(1):1–12, 05 2001.
- [100] Titouan Lazeyras and Fabian Schmidt. Beyond LIMD bias: a measurement of the complete set of third-order halo bias parameters. *Journal of Cosmology and Astroparticle Physics*, 2018(09):008–008, sep 2018.

- [101] M. Manera, Ravi K. Sheth, and R. Scoccimarro. Large-scale bias and the inaccuracy of the peak-background split. *Monthly Notices of the Royal Astronomical Society*, 402(1):589–602, dec 2009.
- [102] Andreas A. Berlind and David H. Weinberg. The halo occupation distribution: Toward an empirical determination of the relation between galaxies and mass. *The Astrophysical Journal*, 575(2):587–616, aug 2002.
- [103] Zheng Zheng, Alison L. Coil, and Idit Zehavi. Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering. *The Astrophysical Journal*, 667(2):760–779, oct 2007.
- [104] Peter Coles. Galaxy formation with a local bias. *Monthly Notices of the Royal Astronomical Society*, 262(4):1065–1075, 06 1993.
- [105] Patrick McDonald and Arabindo Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. *Journal of Cosmology and Astroparticle Physics*, 2009(08):020–020, aug 2009.
- [106] Tobias Baldauf, Uroš Seljak, Vincent Desjacques, and Patrick McDonald. Evidence for quadratic tidal tensor bias from the halo bispectrum. *Physical Review D*, 86(8), oct 2012.
- [107] Uroš Seljak. Analytic model for galaxy and dark matter clustering. *Mon. Not. R. Astron. Soc.* , 318(1):203–213, October 2000.
- [108] J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *Mon. Not. R. Astron. Soc.* , 318(4):1144–1156, November 2000.
- [109] Román Scoccimarro, Ravi K. Sheth, Lam Hui, and Bhuvnesh Jain. How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *Astrophys. J.* , 546(1):20–34, January 2001.
- [110] Muntazir Mehdi Abidi and Tobias Baldauf. Cubic halo bias in Eulerian and Lagrangian space. *Journal of Cosmology and Astro-Particle Physics*, 2018(7):029, July 2018.
- [111] Ravi K. Sheth, Kwan Chuen Chan, and Román Scoccimarro. Nonlocal lagrangian bias. *Phys. Rev. D*, 87:083002, Apr 2013.
- [112] F. Zwicky. Clusters of Nebulae. *Publications of the Astronomical Society of the Pacific*, 54(320):185, October 1942.

- [113] D. Nelson Limber. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II. *Astrophys. J.* , 119:655, May 1954.
- [114] L. Katz and Gerard F. W. Mulders. On the Clustering of Nebulae. II. *Astrophys. J.* , 95:565, May 1942.
- [115] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *Mon. Not. R. Astron. Soc.* , 341(4):1311–1332, June 2003.
- [116] Ryuichi Takahashi, Masanori Sato, Takahiro Nishimichi, Atsushi Taruya, and Masamune Oguri. Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *Astrophys. J.* , 761(2):152, December 2012.
- [117] Ryuichi Takahashi, Takahiro Nishimichi, Toshiya Namikawa, Atsushi Taruya, Issha Kayo, Ken Osato, Yosuke Kobayashi, and Masato Shirasaki. Fitting the Nonlinear Matter Bispectrum by the Halofit Approach. *Astrophys. J.* , 895(2):113, June 2020.
- [118] Antony Lewis and Anthony Challinor. CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library, record ascl:1102.026, February 2011.
- [119] Julien Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, page arXiv:1104.2932, April 2011.
- [120] Daniel J. Eisenstein and Wayne Hu. Baryonic Features in the Matter Transfer Function. *Astrophys. J.* , 496(2):605–614, March 1998.
- [121] Diego Blas, Mathias Garny, Mikhail M. Ivanov, and Sergey Sibiryakov. Time-sliced perturbation theory II: baryon acoustic oscillations and infrared resummation. *Journal of Cosmology and Astro-Particle Physics*, 2016(7):028, July 2016.
- [122] Mikhail M. Ivanov and Sergey Sibiryakov. Infrared resummation for biased tracers in redshift space. *Journal of Cosmology and Astro-Particle Physics*, 2018(7):053, July 2018.
- [123] Hee-Jong Seo, Ethan R. Siegel, Daniel J. Eisenstein, and Martin White. Nonlinear Structure Formation and the Acoustic Scale. *Astrophys. J.* , 686(1):13–24, October 2008.

- [124] Ariel G. Sánchez, Román Scoccimarro, Martín Crocce, Jan Niklas Grieb, Salvador Salazar-Albornoz, Claudio Dalla Vecchia, Martha Lippich, Florian Beutler, Joel R. Brownstein, Chia-Hsun Chuang, Daniel J. Eisenstein, Francisco-Shu Kitaura, Matthew D. Olmstead, Will J. Percival, Francisco Prada, Sergio Rodríguez-Torres, Ashley J. Ross, Lado Samushia, Hee-Jong Seo, Jeremy Tinker, Rita Tojeiro, Mariana Vargas-Magaña, Yuting Wang, and Gong-Bo Zhao. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges. *Mon. Not. R. Astron. Soc.* , 464(2):1640–1658, January 2017.
- [125] Kenneth G. Wilson. The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47(4):773–840, October 1975.
- [126] Alexander Eggemeier, Román Scoccimarro, and Robert E. Smith. Bias loop corrections to the galaxy bispectrum. *Phys. Rev. D* , 99(12):123514, June 2019.
- [127] Mathias Garny, Dominik Laxhuber, and Roman Scoccimarro. Perturbation theory with dispersion and higher cumulants: non-linear regime. October 2022.
- [128] Alexander Eggemeier, Román Scoccimarro, Robert E. Smith, Martin Crocce, Andrea Pezzotta, and Ariel G. Sánchez. Testing one-loop galaxy bias: Joint analysis of power spectrum and bispectrum. *Phys. Rev. D* , 103(12):123550, June 2021.
- [129] Rupert A. C. Croft and Enrique Gaztanaga. Reconstruction of cosmological density and velocity fields in the Lagrangian Zel’dovich approximation. *Mon. Not. R. Astron. Soc.* , 285(4):793–805, March 1997.
- [130] Adi Nusser and Enzo Branchini. On the least action principle in cosmology. *Mon. Not. R. Astron. Soc.* , 313(3):587–595, April 2000.
- [131] Enzo Branchini, Amiram Eldar, and Adi Nusser. Peculiar velocity reconstruction with the fast action method: tests on mock redshift surveys. *Mon. Not. R. Astron. Soc.* , 335(1):53–72, September 2002.
- [132] F. S. Kitaura and T. A. Enßlin. Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization. *Mon. Not. R. Astron. Soc.* , 389(2):497–544, September 2008.
- [133] Francisco-Shu Kitaura, Raul E. Angulo, Yehuda Hoffman, and Stefan Gottlöber. Estimating cosmic velocity fields from density fields and tidal tensors. *Mon. Not. R. Astron. Soc.* , 425(4):2422–2435, October 2012.

- [134] Florian Beutler, Chris Blake, Jun Koda, Felipe A. Marín, Hee-Jong Seo, Antonio J. Cuesta, and Donald P. Schneider. The BOSS-WiggleZ overlap region - I. Baryon acoustic oscillations. *Mon. Not. R. Astron. Soc.* , 455(3):3230–3248, January 2016.
- [135] E. Sarpa, C. Schimd, E. Branchini, and S. Matarrese. BAO reconstruction: a swift numerical action method for massive spectroscopic surveys. *Mon. Not. R. Astron. Soc.* , 484(3):3818–3830, April 2019.
- [136] A. J. S. Hamilton. Uncorrelated modes of the nonlinear power spectrum. *Mon. Not. R. Astron. Soc.* , 312:257–284, 2000.
- [137] Zachary Slepian and Daniel J. Eisenstein. On the signature of the baryon-dark matter relative velocity in the two- and three-point galaxy correlation functions. *Mon. Not. R. Astron. Soc.* , 448(1):9–26, March 2015.
- [138] James D. Talman. Numerical Fourier and Bessel Transforms in Logarithmic Variables. *Journal of Computational Physics*, 29(1):35–48, October 1978.
- [139] Valentin Assassi, Marko Simonović, and Matias Zaldarriaga. Efficient evaluation of angular power spectra and bispectra. *Journal of Cosmology and Astro-Particle Physics*, 2017(11):054, November 2017.
- [140] Henry S. Grasshorn Gebhardt and Donghui Jeong. Fast and accurate computation of projected two-point functions. *Phys. Rev. D* , 97(2):023504, January 2018.
- [141] Nils Schöneberg, Marko Simonović, Julien Lesgourgues, and Matias Zaldarriaga. Beyond the traditional line-of-sight approach of cosmological angular statistics. *Journal of Cosmology and Astro-Particle Physics*, 2018(10):047, October 2018.
- [142] Zachary Slepian, Yin Li, Marcel Schmittfull, and Zvonimir Vlah. Rotation method for accelerating multiple-spherical Bessel function integrals against a numerical source function. November 2019.
- [143] Marcel Schmittfull and Zvonimir Vlah. FFT-PT: Reducing the two-loop large-scale structure power spectrum to low-dimensional radial integrals. *Phys. Rev. D*, 94(10):103530, 2016.
- [144] Joseph E. McEwen, Xiao Fang, Christopher M. Hirata, and Jonathan A. Blazek. FAST-PT: a novel algorithm to calculate convolution integrals in cosmological perturbation theory. *JCAP*, 09:015, 2016.

- [145] Marcel Schmittfull, Zvonimir Vlah, and Patrick McDonald. Fast large scale structure perturbation theory using one-dimensional fast Fourier transforms. *Phys. Rev. D*, 93(10):103528, 2016.
- [146] Marko Simonović, Tobias Baldauf, Matias Zaldarriaga, John Joseph Carrasco, and Juna A. Kollmeier. Cosmological perturbation theory using the FFTLog: formalism and connection to QFT loop integrals. *JCAP*, 04:030, 2018.
- [147] Obinna Umeh. Optimal computation of anisotropic galaxy three point correlation function multipoles using 2DFFTLOG formalism. *Journal of Cosmology and Astro-Particle Physics*, 2021(5):035, May 2021.
- [148] Xiao Fang, Tim Eifler, and Elisabeth Krause. 2D-FFTLog: Efficient computation of real space covariance matrices for galaxy clustering and weak lensing. *Mon. Not. R. Astron. Soc.*, 497(3):2699–2714, 2020.
- [149] Joseph E. McEwen, Xiao Fang, Christopher M. Hirata, and Jonathan A. Blazek. FAST-PT: a novel algorithm to calculate convolution integrals in cosmological perturbation theory. *JCAP*, 09:015, 2016.
- [150] M. Guidi, A. Veropalumbo, E. Branchini, A. Eggemeier, and C. Carbone. Modelling the next-to-leading order matter three-point correlation function using FFTLog. *arXiv e-prints*, page arXiv:2212.07382, December 2022.
- [151] A. Veropalumbo, A. Binetti, E. Branchini, M. Moresco, P. Monaco, A. Oddo, A. G. Sánchez, and E. Sefusatti. The halo 3-point correlation function: a methodological analysis. *Journal of Cosmology and Astro-Particle Physics*, 2022(9):033, September 2022.
- [152] Ross O’Connell, Daniel Eisenstein, Mariana Vargas, Shirley Ho, and Nikhil Padmanabhan. Large covariance matrices: smooth models from the two-point correlation function. *Mon. Not. R. Astron. Soc.*, 462(3):2681–2694, November 2016.
- [153] István Szapudi and Alexander S. Szalay. A New Class of Estimators for the N-point Correlations. *arXiv e-prints*, pages astro-ph/9704241, April 1997.
- [154] István Szapudi. Cosmic Statistics of Statistics: N-point Correlations. In James N. Fry, J. Robert Buchler, and Henry Kandrup, editors, *The Onset of Nonlinearity in Cosmology*, volume 927, page 94, January 2001.

- [155] Jan Niklas Grieb, Ariel G. Sánchez, Salvador Salazar-Albornoz, and Claudio Dalla Vecchia. Gaussian covariance matrices for anisotropic galaxy clustering measurements. *Mon. Not. R. Astron. Soc.* , 457(2):1577–1592, April 2016.
- [156] Carmelita Carbone, Margarita Petkova, and Klaus Dolag. DEMNUni: ISW, Rees-Sciama, and weak-lensing in the presence of massive neutrinos. *Journal of Cosmology and Astro-Particle Physics*, 2016(7):034, July 2016.
- [157] Emanuele Castorina, Carmelita Carbone, Julien Bel, Emiliano Sefusatti, and Klaus Dolag. DEMNUni: the clustering of large-scale structures in the presence of massive neutrinos. *Journal of Cosmology and Astro-Particle Physics*, 2015(7):043–043, July 2015.
- [158] M. Zennaro, J. Bel, J. Dossett, C. Carbone, and L. Guzzo. Cosmological constraints from galaxy clustering in the presence of massive neutrinos. *Mon. Not. R. Astron. Soc.* , 477(1):491–506, June 2018.
- [159] Rossana Ruggeri, Emanuele Castorina, Carmelita Carbone, and Emiliano Sefusatti. DEMNUni: massive neutrinos and the bispectrum of large scale structures. *Journal of Cosmology and Astro-Particle Physics*, 2018(3):003, March 2018.
- [160] J. Bel, A. Pezzotta, C. Carbone, E. Sefusatti, and L. Guzzo. Accurate fitting functions for peculiar velocity spectra in standard and massive-neutrino cosmologies. *Astron. Astrophys.* , 622:A109, February 2019.
- [161] G. Parimbelli, S. Anselmi, M. Viel, C. Carbone, F. Villaescusa-Navarro, P. S. Corasaniti, Y. Rasera, R. Sheth, G. D. Starkman, and I. Zehavi. The effects of massive neutrinos on the linear point of the correlation function. *Journal of Cosmology and Astro-Particle Physics*, 2021(1):009, January 2021.
- [162] G. Parimbelli, C. Carbone, J. Bel, B. Bose, M. Calabrese, E. Carella, and M. Zennaro. DEMNUni: comparing nonlinear power spectra prescriptions in the presence of massive neutrinos and dynamical dark energy. *Journal of Cosmology and Astro-Particle Physics*, 2022(11):041, November 2022.
- [163] E. Carella, C. Carbone, M. Zennaro, G. Girelli, M. Bolzonella, F. Marulli, L. Pozzetti, M. Calabrese, and L. Guzzo. DEMNUni: The galaxy-halo connection in the presence of dynamical dark energy and massive neutrinos. *In prep.*

- [164] M. Roncarelli, C. Carbone, and L. Moscardini. The effect of massive neutrinos on the Sunyaev-Zel'dovich and X-ray observables of galaxy clusters. *Mon. Not. R. Astron. Soc.* , 447(2):1761–1773, February 2015.
- [165] Giulio Fabbian, Matteo Calabrese, and Carmelita Carbone. CMB weak-lensing beyond the Born approximation: a numerical approach. *Journal of Cosmology and Astro-Particle Physics*, 2018(2):050, February 2018.
- [166] Christina D. Kreisch, Alice Pisani, Carmelita Carbone, Jia Liu, Adam J. Hawken, Elena Massara, David N. Spergel, and Benjamin D. Wandelt. Massive neutrinos leave fingerprints on cosmic voids. *Mon. Not. R. Astron. Soc.* , 488(3):4413–4426, September 2019.
- [167] Nico Schuster, Nico Hamaus, Alice Pisani, Carmelita Carbone, Christina D. Kreisch, Giorgia Pollina, and Jochen Weller. The bias of cosmic voids in the presence of massive neutrinos. *Journal of Cosmology and Astro-Particle Physics*, 2019(12):055, December 2019.
- [168] Giovanni Verza, Alice Pisani, Carmelita Carbone, Nico Hamaus, and Luigi Guzzo. The void size function in dynamical dark energy cosmologies. *Journal of Cosmology and Astro-Particle Physics*, 2019(12):040, December 2019.
- [169] Giovanni Verza, Carmelita Carbone, and Alessandro Renzi. The Halo Bias inside Cosmic Voids. *Astrophys. J. Lett.* , 940(1):L16, November 2022.
- [170] P. Vielzeuf, M. Calabrese, C. Carbone, G. Fabbian, and C. Baccigalupi. DEMNUni: The imprint of massive neutrinos on the cross-correlation between cosmic voids and CMB lensing. *In prep.*
- [171] Viviana Cuozzo, Carmelita Carbone, Matteo Calabrese, and Elisabetta Carella. DEMNUni: CMB-galaxy cross-correlation in the presence of massive neutrinos. *In prep.*
- [172] M. Zennaro, J. Bel, F. Villaescusa-Navarro, C. Carbone, E. Sefusatti, and L. Guzzo. Initial conditions for accurate N-body simulations of massive neutrino cosmologies. *Mon. Not. R. Astron. Soc.* , 466(3):3244–3258, April 2017.
- [173] E. Keihänen, H. Kurki-Suonio, V. Lindholm, A. Viitanen, A. S. Suur-Uski, V. Allevalo, E. Branchini, F. Marulli, P. Norberg, D. Tavagnacco, S. de la Torre, J. Valiviita, M. Viel, J. Bel, M. Frailis, and A. G. Sánchez. Estimating the galaxy two-point correlation function using a split random catalog. *Astron. Astrophys.* , 631:A73, November 2019.

- [174] Michele Moresco, Alfonso Veropalumbo, Federico Marulli, Lauro Moscardini, and Andrea Cimatti. C3: Cluster Clustering Cosmology. ii. First Detection of the Baryon Acoustic Oscillations Peak in the Three-point Correlation Function of Galaxy Clusters. *Astrophys. J.*, 919(2):144, 2021.
- [175] Douglas Potter, Joachim Stadel, and Romain Teyssier. PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1):2, May 2017.
- [176] Martin Kerscher, István Szapudi, and Alexander S. Szalay. A Comparison of Estimators for the Two-Point Correlation Function. *Astrophys. J. Lett.* , 535(1):L13–L16, May 2000.
- [177] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, January 2010.
- [178] Volker Springel. The cosmological simulation code GADGET-2. *Mon. Not. R. Astron. Soc.* , 364(4):1105–1134, December 2005.
- [179] Ariel G. Sánchez, Eyal A. Kazin, Florian Beutler, Chia-Hsun Chuang, Antonio J. Cuesta, Daniel J. Eisenstein, Marc Manera, Francesco Montesano, Robert C. Nichol, Nikhil Padmanabhan, Will Percival, Francisco Prada, Ashley J. Ross, David J. Schlegel, Jeremy Tinker, Rita Tojeiro, David H. Weinberg, Xiaoying Xu, J. Brinkmann, Joel R. Brownstein, Donald P. Schneider, and Daniel Thomas. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological constraints from the full shape of the clustering wedges. *Mon. Not. R. Astron. Soc.* , 433(2):1202–1222, August 2013.
- [180] Héctor Gil-Marín, Jorge Noreña, Licia Verde, Will J. Percival, Christian Wagner, Marc Manera, and Donald P. Schneider. The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity. *Mon. Not. R. Astron. Soc.* , 451(1):539–580, July 2015.
- [181] Héctor Gil-Marín, Licia Verde, Jorge Noreña, Antonio J. Cuesta, Lado Samushia, Will J. Percival, Christian Wagner, Marc Manera, and Donald P. Schneider. The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - II. Cosmological interpretation. *Mon. Not. R. Astron. Soc.* , 452(2):1914–1921, September 2015.

- [182] Héctor Gil-Marín, Will J. Percival, Licia Verde, Joel R. Brownstein, Chia-Hsun Chuang, Francisco-Shu Kitaura, Sergio A. Rodríguez-Torres, and Matthew D. Olmstead. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies. *Mon. Not. R. Astron. Soc.* , 465(2):1757–1788, February 2017.
- [183] David W. Pearson and Lado Samushia. A Detection of the Baryon Acoustic Oscillation features in the SDSS BOSS DR12 Galaxy Bispectrum. *Mon. Not. R. Astron. Soc.* , 478(4):4500–4512, August 2018.
- [184] Oliver H. E. Philcox and Daniel J. Eisenstein. Estimating covariance matrices for two- and three-point correlation function moments in Arbitrary Survey Geometries. *Mon. Not. R. Astron. Soc.* , 490(4):5931–5951, December 2019.
- [185] X. Xu, M. White, N. Padmanabhan, D. J. Eisenstein, J. Eckel, K. Mehta, M. Metchnik, P. Pinto, and H. J. Seo. A New Statistic for Analyzing Baryon Acoustic Oscillations. *Astrophys. J.* , 718(2):1224–1234, August 2010.
- [186] Román Scoccimarro, Matias Zaldarriaga, and Lam Hui. Power Spectrum Correlations Induced by Nonlinear Clustering. *Astrophys. J.* , 527(1):1–15, December 1999.
- [187] Xiao Fang, Jonathan A. Blazek, Joseph E. McEwen, and Christopher M. Hirata. FAST-PT II: an algorithm to calculate convolution integrals of general tensor quantities in cosmological perturbation theory. *JCAP*, 02:030, 2017.
- [188] Zachary Slepian. On decoupling the integrals of cosmological perturbation theory. *Mon. Not. R. Astron. Soc.* , 507(1):1337–1360, 2021.
- [189] J. Barriga and Enrique Gaztanaga. The 3-point function in the large scale structure. 1. the weakly nonlinear regime in n-body simulations. *Mon. Not. R. Astron. Soc.* , 333:443–453, 2002.
- [190] P. J. E. Peebles and E. J. Groth. Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog. *Astrophys. J.* , 196:1–11, February 1975.
- [191] Daniel Baumann, Alberto Nicolis, Leonardo Senatore, and Matias Zaldarriaga. Cosmological non-linearities as an effective fluid. *Journal of Cosmology and Astro-Particle Physics*, 2012(7):051, July 2012.
- [192] Tobias Baldauf. Advanced Cosmology, Statistics, Non-Gaussianity and Non-Linearity.

- [193] Ryuichi Takahashi, Takahiro Nishimichi, Toshiya Namikawa, Atsushi Taruya, Issha Kayo, Ken Osato, Yosuke Kobayashi, and Masato Shirasaki. Fitting the Nonlinear Matter Bispectrum by the Halofit Approach. *Astrophys. J.* , 895(2):113, June 2020.
- [194] Enrico Pajer and Drian van der Woude. Divergence of perturbation theory in large scale structures. *Journal of Cosmology and Astro-Particle Physics*, 2018(5):039, May 2018.
- [195] Volker Springel. The cosmological simulation code GADGET-2. *Mon. Not. R. Astron. Soc.* , 364(4):1105–1134, December 2005.
- [196] Matteo Viel, Martin G. Haehnelt, and Volker Springel. The effect of neutrinos on the matter distribution as probed by the intergalactic medium. *Journal of Cosmology and Astro-Particle Physics*, 2010(6):015, June 2010.
- [197] M. H. Goroff, B. Grinstein, S. J. Rey, and M. B. Wise. Coupling of modes of cosmological mass density fluctuations. *Astrophys. J.* , 311:6–14, December 1986.
- [198] F. Marulli, A. Veropalumbo, and M. Moresco. CosmoBolognaLib: C++ libraries for cosmological calculations. *Astronomy and Computing*, 14:35–42, January 2016.
- [199] Daniel J. Eisenstein, Hee-Jong Seo, and Martin White. On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter. *Astrophys. J.* , 664(2):660–674, August 2007.
- [200] Wendy L. Freedman, Barry F. Madore, Brad K. Gibson, Laura Ferrarese, Daniel D. Kelson, Shoko Sakai, Jeremy R. Mould, Jr. Robert C. Kennicutt, Holland C. Ford, John A. Graham, John P. Huchra, Shaun M. G. Hughes, Garth D. Illingworth, Lucas M. Macri, and Peter B. Stetson. Final results from theiubble space telescope/ikey project to measure the hubble constant. *The Astrophysical Journal*, 553(1):47–72, may 2001.
- [201] Robert J. Nemiroff and Bijunath Patla. Adventures in friedmann cosmology: A detailed expansion of the cosmological friedmann equations. *American Journal of Physics*, 76(3):265–276, mar 2008.
- [202] C. Wagner, F. Schmidt, C. T. Chiang, and E. Komatsu. Separate universe simulations. *Mon. Not. R. Astron. Soc.* , 448:L11–L15, March 2015.
- [203] Fabian Schmidt, Donghui Jeong, and Vincent Desjacques. Peak-background split, renormalization, and galaxy clustering. *Phys. Rev. D* , 88(2):023515, July 2013.

- [204] Patrick McDonald and Arabindo Roy. Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS. *Journal of Cosmology and Astro-Particle Physics*, 2009(8):020, August 2009.
- [205] Tobias Baldauf, Uroš Seljak, Vincent Desjacques, and Patrick McDonald. Evidence for quadratic tidal tensor bias from the halo bispectrum. *Phys. Rev. D*, 86(8):083540, October 2012.
- [206] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [207] T. Hahn. CUBA—a library for multidimensional numerical integration. *Computer Physics Communications*, 168(2):78–95, June 2005.
- [208] Andrea Pezzotta, Martin Crocce, Alexander Eggemeier, Ariel G. Sánchez, and Román Scoccimarro. Testing one-loop galaxy bias: Cosmological constraints from the power spectrum. *Phys. Rev. D*, 104:043531, Aug 2021.
- [209] Alexander Eggemeier, Román Scoccimarro, Martin Crocce, Andrea Pezzotta, and Ariel G. Sánchez. Testing one-loop galaxy bias: Power spectrum. *Phys. Rev. D*, 102:103530, Nov 2020.
- [210] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762(2):109, January 2013.
- [211] Martha Lippich, Ariel G. Sánchez, Manuel Colavincenzo, Emiliano Sefusatti, Pierluigi Monaco, Linda Blot, Martin Crocce, Marcelo A. Alvarez, Aniket Agrawal, Santiago Avila, Andrés Balaguera-Antolínez, Richard Bond, Sandrine Codis, Claudio Dalla Vecchia, Antonio Dorta, Pablo Fosalba, Albert Izard, Francisco-Shu Kitaura, Marcos Pellejero-Ibanez, George Stein, Mohammadjavad Vakili, and Gustavo Yepes. Comparing approximate methods for mock catalogues and covariance matrices - I. Correlation function. *Mon. Not. R. Astron. Soc.*, 482(2):1786–1806, January 2019.

- [212] Andrea Oddo, Emiliano Sefusatti, Cristiano Porciani, Pierluigi Monaco, and Ariel G. Sánchez. Toward a robust inference method for the galaxy bispectrum: likelihood function and model selection. *Journal of Cosmology and Astro-Particle Physics*, 2020(3):056, March 2020.
- [213] J. P. Ostriker and P. J. E. Peebles. A Numerical Study of the Stability of Flattened Galaxies: or, can Cold Galaxies Survive? *Astrophys. J.* , 186:467–480, December 1973.
- [214] H. Hildebrandt, M. Viola, C. Heymans, S. Joudaki, K. Kuijken, C. Blake, T. Erben, B. Joachimi, D. Klaes, L. Miller, C. B. Morrison, R. Nakajima, G. Verdoes Kleijn, A. Amon, A. Choi, G. Covone, J. T. A. de Jong, A. Dvornik, I. Fenech Conti, A. Grado, J. Harnois-Déraps, R. Herbonnet, H. Hoekstra, F. Köhlinger, J. McFarland, A. Mead, J. Merten, N. Napolitano, J. A. Peacock, M. Radovich, P. Schneider, P. Simon, E. A. Valentijn, J. L. van den Busch, E. van Uitert, and L. Van Waerbeke. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Mon. Not. R. Astron. Soc.* , 465(2):1454–1498, February 2017.
- [215] L Verde, J A Peacock, R Jimenez, M Kamionkowski, and S Matarrese. The 2df galaxy redshift survey: The bias of galaxies and the density of the universe. *Mon. Not. Roy. Astron. Soc.*, 335(1):432, 2002.
- [216] R. G. Mann, J. A. Peacock, and A. F. Heavens. Eulerian bias and the galaxy density field. *Mon. Not. R. Astron. Soc.* , 293(3):209–221, January 1998.