ON THE LINEAR TERM CORRECTION FOR NEEDLET/WAVELET NON-GAUSSIANITY ESTIMATORS

Simona Donzelli, Frede K. Hansen, Michele Liguori, Domenico Marinucci, and Sabino Matarrese

1 INAF-Istituto di Astrofisica Spaziale e Fisica Cosmica Milano, Via E. Bassini 15, 20133 Milano, Italy; donzelli@iasf-milano.inaf.it
2 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
3 Centre of Mathematics for Applications, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway
4 Dipartimento di Fisica e Astronomia “G. Galilei,” Università di Padova, Università degli Studi di Padova, Via Marzolo 8, 35131 Padova, Italy
5 INFN, Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy
6 Dipartimento di Matematica, Università di Roma “Tor Vergata,” Via della Ricerca Scientifica 1, I-00133 Roma, Italy

Received 2012 February 7; accepted 2012 May 30; published 2012 July 23

ABSTRACT

We derive the linear correction term for needlet and wavelet estimators of the bispectrum and the nonlinearity parameter \( f_{NL} \) on cosmic microwave background radiation data. We show that on masked WMAP-like data with anisotropic noise, the error bars improve by 10%–20% and almost reach the optimal error bars obtained with the bispectrum estimator also known as “KSW”. In the limit of full-sky and isotropic noise, this term vanishes. We apply needlet and wavelet estimators to the WMAP 7-year data and obtain our best estimate \( f_{NL} = 37.5 \pm 21.8 \) (68% CL).

Key words: cosmic background radiation – cosmology: observations – early universe – methods: data analysis – methods: statistical

1. INTRODUCTION

It is well known that most inflationary models predict the fluctuations in the cosmic microwave background (CMB) to be close to but not exactly Gaussian. Non-Gaussian predictions are strongly model dependent, thus making primordial non-Gaussianity (NG) a powerful tool to discriminate among different early universe scenarios (see, e.g., Bartolo et al. 2004a; Chen 2010; Liguori et al. 2010, and references therein).

In this paper, we will focus on so-called local non-Gaussianity, which can be parameterized in the simple form:

\[
\Phi(x) = \Phi_L(x) + f_{NL}^{local} \left( \Phi_L^2(x) - \langle \Phi_L^2(x) \rangle \right),
\]

where \( \Phi(x) \) is the primordial curvature perturbation field at the end of inflation and \( \Phi_L(x) \) is the Gaussian part of the perturbation. The dimensionless parameter \( f_{NL}^{local} \) describes the amplitude of non-Gaussianity. \(^7\) Local non-Gaussianity is predicted to arise from standard single-field slow-roll inflation (Acquaviva et al. 2003; Maldacena 2003), although at a very tiny level, as well as from multi-field inflationary scenarios, like the curvaton (Mollerach 1990; Enqvist & Sloth 2002; Lyth & Wands 2002; Moroi & Takahashi 2001) or inhomogeneous (pre)reheating models (Dvali et al. 2004; Kolb et al. 2005, 2006). Even alternatives to inflation, such as ekpyrotic and cyclic models (Koyama et al. 2007; Buchbinder et al. 2008), predict a local NG signature.

The expected non-Gaussian amplitude \( f_{NL} \) varies significantly from model to model. For example, standard single-field slow-roll inflation predicts \( f_{NL} \sim 10^{-2} \) at the end of inflation (Acquaviva et al. 2003; Maldacena 2003; and therefore a final value \( \sim \) unity after general relativistic second-order perturbation effects is taken into account; Bartolo et al. 2004b, 2004c). Such a small value is not experimentally detectable and for this reason a detection of a primordial non-Gaussian signal in present and forthcoming CMB data will rule out single-field slow-roll inflation as a viable scenario. Motivated by these considerations, many groups have attempted to measure \( f_{NL} \) using CMB data sets, and Wilkinson Microwave Anisotropy Probe (WMAP) data in particular.

Consistently, with theoretical findings, the most stringent NG bounds have been obtained using estimators of the CMB angular bispectrum (namely, the three-point function of CMB fluctuations in harmonic space). While in the classical approach to \( f_{NL} \) estimation (Komatsu et al. 2005; Creminelli et al. 2006; Yadav & Wandelt 2008; Yadav et al. 2007a) the starting point to build an optimal cubic statistic is a direct multipole expansion of the temperature field, alternative representations can be used as well, like, e.g., the modal bispectrum expansion of Fergusson et al. (2009) or bispectra of wavelet and needlet coefficients. Since all these approaches are just based on expanding the same quantity (the angular bispectrum) in different bases (polynomial modes, wavelets, needlets, etc.), they are also ultimately expected to yield very similar results when applied to data. This is indeed the case. For example, a recent estimate using an optimal bispectrum estimator has been made by Smith et al. (2009). They obtained the smallest error bars on \( f_{NL} \) to date, finding \( f_{NL} = 38 \pm 21 \) on WMAP 5-year data. Consistent results, although with larger error bars, were found by Curto et al. (2009) and Pietrobon et al. (2009), using parts of the bispectrum of Spherical Mexican Hat Wavelets (SMHWs; Martínez-González et al. 2002) and the skewness of needlet coefficients, and by Fergusson et al. (2010), using a modal bispectrum expansion. The most updated optimal result has been found by the WMAP team on the 7-year data with the estimate \( f_{NL} = 32 \pm 21 \) (Komatsu et al. 2011). Wavelets provide again a very similar result of \( f_{NL} = 30 \pm 23 \) (Fisher matrix bound \( \sigma_{f_{NL}} = 22.5 \); Curto et al. 2011c).

At this point one could reasonably ask why it is useful to implement estimators using many different bispectrum representations. After all, in the end they are all expected to produce basically the same output in terms of \( f_{NL} \). The justification is two-fold. First of all, different expansions can provide information beyond \( f_{NL} \) (like mode spectra and full bispectrum reconstruction; Fergusson et al. 2009, 2010). Second of all, different expansions can present important practical advantages, such
as computational rapidity or robustness to a number of contaminants and effects (masking, non-stationarity of the noise, foreground emission, and so on). The WMAP 7-year results quoted above seem indeed to illustrate the latter point well. When we compare the bispectrum and the wavelet results we see that central values and error bars are both very similar. However, to achieve this result the bispectrum estimator needs to include a very important linear correction term. This additional term, originally introduced in Creminelli et al. (2006), subtracts from the measured three-point function a spurious $\mathcal{f}_NL$ contribution due to the breaking of statistical isotropy introduced by masking and non-stationary noise. Without this contribution the error bars of the bispectrum estimator would be much larger than the quoted 21 (a factor at least 4 or 5 larger, as shown, for example, in Figure 4 of Creminelli et al. 2006). It turns out, however, that despite having an error bar only $\sim 10\%$ larger than the bispectrum measurement, the WMAP 7-year wavelet result of $\mathcal{f}_NL = 30.0 \pm 23$ (Fisher matrix bound $\sigma_F = 22.5$; Curto et al. 2011c) was obtained without including any linear correction. It seems then that the wavelet expansion is much less affected by masking and anisotropic noise than the bispectrum estimator. It seems then that the wavelet expansion is much less affected by masking and anisotropic noise than the bispectrum estimator. It seems then that the wavelet expansion is much less affected by masking and anisotropic noise than the bispectrum estimator. It seems then that the wavelet expansion is much less affected by masking and anisotropic noise than the bispectrum estimator.

We will return to a needlet term correction on a needlet bispectrum estimator and show that these results are confirmed by simulations; the procedures are drawn in Section 4. Conclusions are drawn in Section 4.

2. MOTIVATION FOR THE LINEAR TERM CORRECTION

2.1. Some Background Results

2.1.1. Wick Products

We recall first some well-known background facts, to fix notation. Consider Gaussian variables $X_1, X_2, X_3$, such that $\langle X_i \rangle = 0$, $\langle X_i^2 \rangle = \sigma_i^2$, $\langle X_i X_j \rangle = \sigma_{ij}$. The Wick product of the three variables is defined as

$$: X_1, X_2, X_3 := X_1 X_2 X_3 - \sigma_{12} X_3 - \sigma_{13} X_2 - \sigma_{23} X_1. \quad (2)$$

Example 1 For $X_1 = X_2 = X_3 = X$

$$: X, X, X := X^3 - 3\sigma^2 X. \quad (3)$$

For $\sigma^2 = 1$ this is the well-known Hermite polynomial of order 3, $H_3(X) = X^3 - 3X$.

Note that the correction is very large for local NG, but much smaller for other types of NG, hence the reason to consider only local NG in this paper, which is entirely focused on issues related to the linear term.

We refer to the first version of the paper (Curto et al. 2011b) in the pre-print arXiv. The cited paper has been modified after the submission of this work.

For the expected values of Wick products, the following Diagram Formula holds

$$\langle X_{11}, X_{12}, X_{13} \rangle = \langle X_{11} X_{21} \rangle \langle X_{12} X_{22} \rangle \langle X_{13} X_{23} \rangle + 5 \text{permutations}, \quad (3)$$

where the only permutations that are considered are those such that in each pair an element from the first triple $\{X_{11}, X_{12}, X_{13}\}$ is coupled with an element from the second triple $\{X_{21}, X_{22}, X_{23}\}$. It is usually convenient to visualize the elements $\{X_{11}, X_{12}, X_{13}\}, \{X_{21}, X_{22}, X_{23}\}$ as vertices aligned on two different rows, and the pairs as edges connecting two different vertices; the above-mentioned diagram formula is then usually expressed by stating that “flat edges” are ruled out.

Example 2 For the Hermite polynomial we have

$$\langle (H_3(X))^2 \rangle = \langle (X^3 - 3X)^2 \rangle = 6.$$
whence the result is established. In words, Wick polynomials of order 3 are uncorrelated by construction with any other polynomial of smaller order in the same random variables, and hence minimize the variance in the class of cubic polynomials of unit coefficient in the maximal term. This provides the heuristic rationale for the introduction of linear correction terms in standard and wavelet/needlet bispectrum estimators.

2.1.2. Needlets, Mexican Needlets, and SMHW

Let $b(t)$ be a weight function satisfying three conditions, namely,

1. Compact support: $b(t)$ is strictly larger than zero only for $t \in [B^{-1}, B]$, some $B > 1$.
2. Smoothness: $b(t)$ is $C^\infty$.
3. Partition of unity: for all $\ell = 1, 2, \ldots$ we have
   $$\sum_{j=0}^\infty b^2 \left( \frac{\ell}{B^{j}} \right) = 1.$$

Recipes to construct a function $b(t)$ that satisfy these conditions are easy to find and are provided for instance by Marinucci et al. (2008) and Marinucci & Peccati (2011).

Consider now a grid of points $\{\xi_{jk}\}$ on the sphere, e.g., the HEALPix\(^{10}\) centers (Górski et al. 2005); the needlet system is then defined by

$$\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{\ell=B^{-1}}^{B^{-1}} \sum_{m=-\ell}^{\ell} b \left( \frac{\ell}{B^{j}} \right) Y_{\ell m}(x) Y_{\ell m}(\xi_{jk}),$$

with the corresponding needlet coefficients provided by

$$\beta_{jk} = \sqrt{\lambda_{jk}} \int_{S^2} f(x) \psi_{jk}(x) dx$$

$$= \sum_{\ell=B^{-1}}^{B^{-1}} \sum_{m=-\ell}^{\ell} b \left( \frac{\ell}{B^{j}} \right) a_{\ell m} Y_{\ell m}(\xi_{jk}).$$

The coefficients $\lambda_{jk}$ are proportional to the pixel area, see for instance Baldi et al. (2009b) for more details. The needlet idea has been extended by Geller & Mayeli (2009a, 2009b) with the construction of the so-called Mexican needlets (see Scodeller et al. 2011 for numerical analysis and implementation in a cosmological framework). Loosely speaking, the idea is to replace the compactly supported kernel $b(\ell/B^j)$ by a smooth function of the form

$$b \left( \frac{\ell}{B^{j}} \right) = \left( \frac{\ell}{B^{j}} \right)^{2p} \exp \left( -\frac{\ell^2}{B^{2j}} \right),$$

for some integer parameter $p$. Mexican needlets have extremely good localization properties in real space, and for $p = 1$ they provide at high frequencies a good approximation to the so-called SMHW construction. The latter is exploited for instance by Martínez-González et al. (2002) and Curto et al. (2009, 2011a, 2011c), to which we refer for more discussion and definitions. In short, the SMHW coefficients at location $n$ and scale $R$ are provided by

$$w(n, R) = \int_{S^2} f(x) \Psi(x, n; R) dx,$$

where the wavelet filter is defined as

$$\Psi(x, n; R) = \frac{1}{\sqrt{2\pi} N(R)} \left[ 1 + \left( \frac{y}{2} \right)^2 \right] \left[ 2 - \left( \frac{y}{R} \right)^2 \right] e^{-y^2/2R^2};$$

and $N(R) = R \sqrt{1 + R^2/2 + R^4/4}$ is a normalizing constant and $y = 2 \tan \theta/2$ represents the distance between $x$ and $n$, evaluated on the stereographic projection on the tangent plane at $n$; $\theta$ is the corresponding angular distance, evaluated on the spherical surface.

2.2. The Linear Correction Term for the KSW Estimator

For our arguments to follow, and to allow for a proper comparison with the results for needlet/wavelet based estimators below, we shall provide a brief heuristic argument to motivate the need for a linear term correction (Creminelli et al. 2006; Yadav et al. 2007b) in the well-known KSW bispectrum estimator for the $f_{NL}$ parameter. In particular, let us consider any three frequencies $\ell_1, \ell_2, \ell_3$ satisfying standard triangular conditions, and consider the angle-averaged bispectrum

$$B_{\ell_1,\ell_2,\ell_3} = \frac{\sqrt{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}}{4\pi} \left( \ell_1 \ell_2 \ell_3 \right) \sum_{m_1m_2m_3} \sum_{m_1m_2m_3} a_{\ell_1m_1}a_{\ell_2m_2}a_{\ell_3m_3},$$

with $\ell_1, \ell_2, \ell_3$ satisfying standard triangular conditions, and consider the angle-averaged bispectrum

$$B_{\ell_{1},\ell_{2},\ell_{3}} = \int_{S^2} T_{\ell_{1}}(x)T_{\ell_{2}}(x)T_{\ell_{3}}(x) dx,$$

where we have chosen a convenient (albeit non-standard) normalization for the bispectrum to make our argument notionally simpler—these normalizations do not affect by any means the substance of the argument. Now, the bispectrum should be more properly written as

$$B_{\ell_{1},\ell_{2},\ell_{3}} = \int_{S^2} \{ T_{\ell_{1}}(x), T_{\ell_{2}}(x), T_{\ell_{3}}(x) \} dx$$

$$= \int_{S^2} \{ T_{\ell_{1}}(x)T_{\ell_{2}}(x)T_{\ell_{3}}(x) \} dx$$

$$- \int_{S^2} \{ T_{\ell_{1}}(x)T_{\ell_{2}}(x) \} dx$$

$$+ \int_{S^2} \{ T_{\ell_{1}}(x)T_{\ell_{3}}(x) \} dx$$

$$= \Gamma_{\ell_{1},\ell_{2},\ell_{3}}(x),$$

$$\Gamma_{\ell_{1},\ell_{2},\ell_{3}}(x) = \{ T_{\ell_{1}}(x)T_{\ell_{2}}(x), T_{\ell_{1}}(x)T_{\ell_{3}}(x) \} dx,$$

In the presence of full-sky maps with isotropic noise we have that $\Gamma_{\ell_{1},\ell_{2}}(x) \equiv \Gamma_{\ell_{1},\ell_{2}}$, i.e., it would be constant over pixels, whence for instance

$$\int_{S^2} \{ \Gamma_{\ell_{1},\ell_{2}}(x) T_{\ell_{1}}(x) \} dx = \int_{S^2} \{ \Gamma_{\ell_{1},\ell_{2}}(x) T_{\ell_{2}}(x) \} dx$$

$$= \Gamma_{\ell_{1},\ell_{2}} \int_{S^2} \{ T_{\ell_{1}}(x) \} dx = 0,$$

\(^{10}\) http://healpix.jpl.nasa.gov
because
\[
\int S^2 \sum_m a_{lm} Y_m(x) dx = \sum_m a_{lm} \int S^2 Y_m(x) dx \equiv 0
\]  
(16)
(assuming the monopole is zero). On the other hand, in the presence of anisotropic noise and/or masked maps the previous argument cannot hold, whence the linear term does not cancel and the variance of the Wick product (including the Wick product) is systematically smaller than any other cubic statistic; for instance, as before
\[
\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
\[= 6 \int_{(S^2 \setminus M) \times (S^2 \setminus M)} \langle T_j(x) T_j(y) \rangle^3 dxdy
\]
\[+ 9 \int_{(S^2 \setminus M) \times (S^2 \setminus M)} \langle T^2_j(x) \rangle \langle T^2_j(y) \rangle \langle T_j(x) T_j(y) \rangle dxdy.
\]
\[\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
whence the previous expression becomes
\[= 6 \left\{ \frac{2\ell + 1}{4\pi} \right\}^3 \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]
\[\times \int S^2 \times S^2 P^3_\ell(x \cdot y) dxdy
\]
\[+ 9 \left\{ \frac{2\ell + 1}{4\pi} \right\}^2 \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]
\[\times \int S^2 \times S^2 P_\ell(x \cdot y) dxdy
\]
\[= 6 \left\{ \frac{2\ell + 1}{4\pi} \right\}^3 \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]  
where the previous expression becomes
\[\langle T^2_j(x) \rangle = \left\{ \frac{2\ell + 1}{4\pi} \right\} \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]  
whence the previous expression becomes
\[= 6 \left\{ \frac{2\ell + 1}{4\pi} \right\}^3 \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right]^3 \]
\[\times \int S^2 \times S^2 P^3_\ell(x \cdot y) dxdy
\]
\[+ 9 \left\{ \frac{2\ell + 1}{4\pi} \right\} \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right]^2 \]
\[\times \int S^2 \times S^2 P_\ell(x \cdot y) dxdy
\]
\[= 6 \left\{ \frac{2\ell + 1}{4\pi} \right\}^3 \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]  
(17)
\[\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
\[\times \left( \ell \ell 0 0 \right)^2 \]  
\[\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
\[= 6 \left\{ \frac{2\ell + 1}{4\pi} \right\} \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]  
In the previous computations, we have written \( P_\ell \) for Legendre polynomials, \( x \cdot y \) for scalar products, we have used the well-known Wigner’s 3j symbol arising from the (Gaunt) integral of the third power of \( P_\ell \), and we exploited the well-known fact that
\[\int S^2 \times S^2 P_\ell(x \cdot y) dxdy = 0,
\]
\[\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
whence the so-called flat edge terms vanish. This is not so in general, though. Note, however, that
\[\int S^2 \times S^2 \left( \frac{2\ell + 1}{4\pi} \right) \left[ C^\text{CMB}_\ell + C^\text{noise}_\ell \right] \]
\[\int S^2 \times S^2 P_\ell(x \cdot y) dxdy = 0,
\]
\[\text{Var} \left\{ \int S^2 M \right\} T^2_j(x) dx \right\}
\]
so that the flat terms are canceled, even in the presence of anisotropic noise or masked regions.

2.3. Needlet/ Wavelet Non-Gaussianity Estimators

The situation for wavelet- or needlet-like non-Gaussianity estimators is to some extent analogous to the one for the KSW procedure. For instance, in Curto et al. (2011a), Equations (14)–(16), the variance of the following statistic is considered:
\[ \frac{1}{4\pi} \frac{1}{\sigma_1 \sigma_2 \sigma_3} \times \int S^2 \langle w(R_1, n_1) w(R_2, n_1) \rangle d\ell_1,
\]
\[ \frac{1}{4\pi} \frac{1}{\sigma_1 \sigma_2 \sigma_3} \times \int S^2 \langle w(R_1, n_1) w(R_2, n_1) \rangle d\ell_1,
\]
\[ \text{where each } w(R_1, m_1) \text{ is the (random) SMH coefficient at scale } R_1 \text{ and location } n_1 \text{ (Equation (11)); this variance can clearly be written as}
\]
\[ \frac{1}{(4\pi)^3} \frac{1}{\sigma_1 \sigma_2 \sigma_3} \times \int S^2 \times S^2 \langle w(R_1, n_1) w(R_2, n_1) \rangle d\ell_1 \times d\ell_2,
\]
\[ \text{For full-sky maps with isotropic noise, we would have that}
\]
\[ \langle w(R_1, n_1) w(R_2, n_1) \rangle \times \langle w(R_1, n_2) w(R_2, n_2) \rangle \]
\[ \times \langle w(R_1, n_1) w(R_2, n_2) \rangle + 5 \text{ permutations,}
\]
\[ \langle w(R_1, n_1) w(R_2, n_1) \rangle \times \langle w(R_1, n_2) w(R_2, n_2) \rangle + 14 \text{ permutations,}
\]
\[ \text{e.g., the only terms that non-vanish are those where } n_1 \text{ and } n_2 \text{ appear in the same pair. However, as discussed above, in the presence of masked regions and/or anisotropic noise the terms with } n_1, n_2 \text{ do not vanish, and we should rather write}
\]
\[ \langle w(R_1, n_1) w(R_2, n_1) \rangle \times \langle w(R_1, n_2) w(R_2, n_2) \rangle + 14 \text{ permutations,}
\]
\[ \text{i.e., there are 9 further “flat” permutations (those where two terms from the same row are coupled—there are three different way for each row to do this). These missing terms give to the integral a contribution of the form}
\]
\[ \frac{1}{(4\pi)^3} \frac{1}{\sigma_1 \sigma_2 \sigma_3} \int S^2 \times S^2 \langle w(R_1, n_1) w(R_2, n_1) \rangle \]
\[ \times \langle w(R_1, n_2) w(R_2, n_2) \rangle \times \langle w(R_1, n_1) w(R_2, n_2) \rangle d\ell_1 d\ell_2
\]
\[ = \frac{1}{(4\pi)^3} \frac{1}{\sigma_1 \sigma_2 \sigma_3} \int S^2 \times S^2 \langle w(R_1, n_1) w(R_2, n_1) \rangle \]
\[ \times \langle w(R_1, n_2) w(R_2, n_2) \rangle \times \langle w(R_1, n_1) w(R_2, n_2) \rangle d\ell_1 d\ell_2.
\]
\[ \text{Now, for isotropic noise, } \langle w(R_1, n_1) w(R_2, n_1) \rangle \simeq \text{const},
\]
\[ \text{whence the previous quantity is approximately proportional to}
\]
\[ \int S^2 \times S^2 \langle w(R_1, n_1) w(R_2, n_2) \rangle d\ell_1 d\ell_2
\]
\[ = \left( \int S^2 w(R_1, n_1) d\ell_1 \right) \times \left( \int S^2 w(R_2, n_2) d\ell_2 \right) = 0,
\]
\[ \text{because}
\]
\[ \int S^2 w(R_k, n_1) d\ell_1 = 0
\]
\[ \text{in the absence of masked regions. So when the celestial sphere is fully observed, it is equivalent regardless of whether or}
not the extra, “flat” terms with the same indices \( n_1, n_2 \) in the pairs are considered.

As stated earlier, and as for the KSW estimators, these terms are no longer identically zero in the presence of masked regions or anisotropic noise. A linear correction term can therefore be needed; we discuss its derivation in the subsection below.

### 2.3.1. The Linear Correction Term

According to our previous argument, it is straightforward to see how, to decrease the variance of the wavelet cubic statistic, it is enough to change the cubic statistic from

\[
\int_{S^2} w(R_i, n)w(R_j, n)w(R_k, n)dn
\]

to

\[
\int_{S^2} \left( w(R_i, n), w(R_j, n), w(R_k, n) \right) dn,
\]

i.e., subtract a linear term so that we have

\[
\int_{S^2} w(R_i, n)w(R_j, n)w(R_k, n)dn
\]

\[
- \int_{S^2} \left( w(R_i, n)w(R_j, n) \right) w(R_k, n)dn
\]

\[
- \int_{S^2} \left( w(R_i, n)w(R_k, n) \right) w(R_j, n)dn
\]

\[
- \int_{S^2} \left( w(R_k, n)w(R_j, n) \right) w(R_i, n)dn.
\]

A similar situation exists for the needlet bispectrum (Lan & Marinucci 2008; Rudjord et al. 2009), which we can implement as

\[
\int_{S^2} \left( \beta_{jk} \right)_{(i)} = \sum_k \sum_{\ell} \beta_{jk} \beta_{jk} = \sum_k \Gamma_{jk}(k) \beta_{jk} - \sum_k \Gamma_{jk}(k) \beta_{jk}.
\]

where \( \beta_{jk} \) are the usual coefficients for (standard or Mexican) needlets and

\[
\Gamma_{jk}(k) = \left\{ \beta_{jk} \beta_{jk} \right\}.
\]

Of course, under isotropic noise

\[
\Gamma_{jk}(k) = \Gamma_{jk} = \sum_{\ell} \frac{\ell}{B^{1/2}} b \left( \frac{\ell}{B^{1/2}} \right) \frac{2\ell + 1}{4\pi} C_{\ell}.
\]

Note that the linear terms have expected value zero always,

\[
\langle \Gamma_{jk}(k) \beta_{jk} \rangle = \Gamma_{jk}(k) \langle \beta_{jk} \rangle = 0; \tag{35}
\]

however, the observed value of these terms over one realization of the sky is exactly equal to zero only if the sum (or the integral) is taken over the whole sphere and the noise is isotropic (assuming again zero monopole), i.e., for SMHW

\[
\int_{S^2} \left( w(R_i, n)w(R_j, n) \right) w(R_k, n)dn
\]

\[
= \langle w(R_i, n)w(R_j, n) \rangle \int_{S^2} w(R_k, n)dn = 0, \tag{36}
\]

and correspondingly for the needlets

\[
\sum_{k} \Gamma_{jk}(k) \beta_{jk} = \Gamma_{jk} \sum_k \beta_{jk} = 0. \tag{37}
\]

No masked regions and isotropic noise are the assumptions under which the behavior of the needlet bispectrum was investigated by Lan & Marinucci (2008), where it was first introduced in the statistical literature.

It should be moreover noted that in practical situations the contribution of the linear term for wavelet-needlet-like bispectrum estimators will be typically smaller than for KSW. This can be explained as follows:

Consider

\[
\sum_k \Gamma_{jk}(k) \beta_{jk} = \sum_k \Gamma_{jk}(k) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) a_{\ell m} Y_{\ell m}(\xi_{jk})
\]

\[
= \sum_k \Gamma_{jk}(k) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) \int_{S^2} T(x) Y_{\ell m}(x) Y_{\ell m}(\xi_{jk}) dx
\]

\[
\simeq \int_{S^2} \Gamma_{jk}(y) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) T(x) P_{\ell}(x \cdot y) dx dy. \tag{38}
\]

Now it is a consequence of needlet concentration in pixel space that

\[
\int_{S^2} \Gamma_{jk}(y) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) \int_{S^2} T(x) P_{\ell}(x \cdot y) dx dy
\]

\[
= \int_{S^2} T(x) \left[ \int_{S^2} \Gamma_{jk}(y) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy \right] dx
\]

\[
\simeq \int_{S^2} T(x) \left[ \int_{B_r(x)} \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy \right] dx, \tag{39}
\]

where \( N_r(x) \) is a small neighborhood of \( x \); now assuming that noise is approximately constant over \( B_r(x) \), we can write

\[
\int_{B_r(x)} \Gamma_{jk}(y) \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy
\]

\[
\simeq \Gamma_{jk}(x) \left[ \int_{B_r(x)} \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy \right] \simeq 0, \tag{40}
\]

because by localization

\[
\int_{B_r(x)} \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy
\]

\[
\simeq \int_{S^2} \sum_{\ell} b \left( \frac{\ell}{B^{1/2}} \right) P_{\ell}(x \cdot y) dy \simeq 0. \tag{41}
\]

### 2.3.2. The Relationship with Mean Subtraction

We shall now show how, in the presence of nearly isotropic noise, the behavior of the linear term is well approximated by subtracting scale-by-scale the sky average of wavelet or needlet coefficients. Indeed, define

\[
\overline{\beta}_j = \frac{1}{N} \sum_k \beta_{jk}. \tag{42}
\]
then
\[
\frac{1}{N} \sum \frac{1}{k} (\beta_{j,k} - \overline{\beta}_{j,k})(\beta_{k,j} - \overline{\beta}_{k,j})(\beta_{j,k} - \overline{\beta}_{j,k})
\]
\[
= \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \beta_{j,k} - \overline{\beta}_{j,k} \left( \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right)
\]
\[
- \overline{\beta}_{j,k} \left( \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right) - \overline{\beta}_{j,k} \left( \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right)
\]
\[
+ 2 \overline{\beta}_{j,k} \beta_{j,k} \overline{\beta}_{j,k}.
\] (43)

Now for nearly isotropic noise, e.g., if the covariance among coefficients
\[
\Gamma_{j,k}(k) = \langle \beta_{j,k} \beta_{j,k} \rangle \simeq \Gamma_{j,k}
\] (44)
is nearly constant over the sky, then, for high enough j, it is natural to expect that the cross-correlation among wavelet coefficients evaluated on a sky realization will be close to the ensemble average,
\[
\Gamma_{j,k} \simeq \left\{ \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right\}.
\] (45)

Moreover we also expect
\[
\overline{\beta}_{j,k} \left( \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right) \simeq 0,
\] (46)

whence
\[
\overline{\beta}_{j,k} \left( \frac{1}{N} \sum \frac{1}{k} \beta_{j,k} \beta_{j,k} \right) \simeq \Gamma_{j,k,iv} \frac{1}{N} \sum \beta_{j,k}.
\] (47)

and similarly for the permutation terms, whence the linear term will be well approximated by mean subtraction. At smaller frequencies j, this argument will not work, but at these scales noise is likely to be negligible. In general, it should be noted that the approximation will work better in cases where noise anisotropy is not extremely relevant, and less well otherwise. Also, for higher order polynomials or alternative statistics (such as KSW) the equivalence between mean subtraction and linear term correction will no longer hold.

2.3.3. A Toy Counterexample

In Curto et al. (2011b, Section 3.2), it is claimed that the linear term is of order $\int_{\text{sky}}^2$, and hence negligible. This is motivated on the basis of the asymptotic uncorrelation of the wavelet coefficients, implying the validity of the Central Limit theorem (CLT) for the cubic bispectrum statistic.

We do agree on the uncorrelation of the coefficients and the CLT taking place; see for instance Baldi et al. (2009a) and Lan & Marinucci (2008, 2009) for analytic arguments in the needlet and Mexican needlet cases. We fail to see, however, why this should necessarily imply that the resulting linear term should be negligible.

As a toy counterexample, consider a sequence of independent Gaussian random variables $X_i$, with zero mean $\langle X_i \rangle$ and nonconstant (e.g., anisotropic) variance $\langle X_i^2 \rangle = \sigma_i^2$; to fix ideas, we shall take $\sigma_i^2 = 1$ for $i = 1, 3, 5$ (i.e., odd); $\sigma_i^2 = 3$ for $i = 2, 4, \ldots$ (i.e., even). With these assumptions, we try to mimic the behavior of nearly independent wavelet coefficients with anisotropic noise; for simplicity, we neglect the effect of masked regions, but the argument would be analogous. Now consider the statistic
\[
B_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^3,
\]
\[
B_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \overline{X}_n)^3, \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,
\]
\[
B_{3n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i X_i X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^3 - 3\sigma_i^2 X_i),
\]

which correspond, in these circumstances, to the naive cubic statistics/bispectrum (without linear term), the cubic statistic with mean subtraction, and the proper bispectrum with Wick polynomials/linear term subtraction. Because the $X_i$ values are exactly independent, it is trivial to see that the CLT holds, and all three statistics are asymptotically Gaussian. Nevertheless, it is readily seen that

\[
\text{Var} \{ B_{1n} \} = \frac{1}{n} \sum_{i=1}^n \text{Var} \{ X_i^3 \} = \frac{1}{n} \sum_{i=1}^n \langle X_i^6 \rangle
\]
\[
= 15 \sum_{i=1}^n \sigma_i^6 \rightarrow 15 \left\{ \frac{\sigma_i^6 + \sigma_i^6}{2} \right\} = 210,
\]

because $\langle X_i^6 \rangle = 15\sigma_i^6$ by Wick’s theorem/Diagram Formula (3) (which in this case must include the so-called flat edges). On the other hand

\[
\text{Var} \{ B_{3n} \} = \frac{1}{n} \sum_{i=1}^n \text{Var} \{ X_i^3 - 3\sigma_i^2 X_i \}
\]
\[
= \frac{1}{n} \sum_{i=1}^n \{ X_i^6 - 6\sigma_i^2 X_i^4 + 9\sigma_i^4 X_i^2 \}
\]
\[
= \frac{1}{n} \sum_{i=1}^n \{ 15\sigma_i^6 - 18\sigma_i^6 + 9\sigma_i^6 \}
\]
\[
= 15 \sum_{i=1}^n \sigma_i^6 \rightarrow 6 \left\{ \frac{\sigma_i^6 + \sigma_i^6}{2} \right\} = 84.
\]

We see thus that the linear terms are indeed not negligible (Var$\{ 3\sigma_i^2 X_i \} = 9\sigma_i^6$, as compared to Var$\{ X_i^3 \} = 15\sigma_i^6$), and, in view of its negative correlation with the cubic statistics, it induces a major decrease in the variance. Concerning mean subtraction, simple computations show that
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \overline{X}_n)^3
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^3 - \frac{3}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \left( \sum_{j=1}^n X_j \right)
\]
\[
+ \frac{3}{\sqrt{n}} \sum_{i=1}^n X_i X_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^3.
\]

Now the third and fourth terms are easily seen to be converging to zero, from the law of large numbers. For the second summand, switching sums we obtain
\[
\frac{3}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \left( \sum_{j=1}^n X_j \right) = \frac{3}{\sqrt{n}} \sum_{i=1}^n X_i \left( \frac{1}{n} \sum_{j=1}^n X_j^2 \right),
\]
where, again by the law of large numbers we have the convergence (with probability one)
\[
\left( \frac{1}{n} \sum_{j=1}^{n} X_j^2 \right) \rightarrow \frac{\sigma^2_1 + \sigma^2_2}{2}.
\]

Hence, neglecting terms of order \( n^{-1} \) we have
\[
B_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( X_i^3 - 3 \frac{\sigma^2_1 + \sigma^2_2}{2} X_i \right)^3.
\]

Using the Diagram Formula of Equation (3) again, after some manipulations one obtains
\[
\text{Var} \{ B_{2n} \} \rightarrow 9\sigma^4_1 \times \frac{\sigma^4_1 + \sigma^4_2 + 2\sigma^2_1 \sigma^2_2}{8} + 9\sigma^2_2 \times \frac{\sigma^4_1 + \sigma^4_2 + 2\sigma^2_1 \sigma^2_2}{8} - \frac{33\sigma^6_1 + 33\sigma^6_2 - 9\sigma^4_1 \sigma^2_1 - 9\sigma^4_2 \sigma^2_1}{8}.
\]

so that, for \( \sigma^2_1 = 1, \sigma^2_2 = 3 \) we have
\[
\text{Var} \{ B_{2n} \} \rightarrow \frac{33 + 33 \times 27 - 81 - 27}{8} = 102.
\]

The difference between the variance with mean subtraction and the linear term is given by
\[
\text{Var} \{ B_{2n} \} - \text{Var} \{ B_{3n} \} \rightarrow -\frac{9}{8} \sigma^4_1 (\sigma^2_2 - \sigma^2_1) + \frac{9}{8} \sigma^4_2 (\sigma^2_2 - \sigma^2_1) = \frac{9}{8} (\sigma^2_2 - \sigma^2_1)^2 (\sigma^2_2 + \sigma^2_1) = 18.
\]

**Remark 3** Clearly the approximation of the linear term improves when the anisotropy decreases; in fact, the difference \( \text{Var} \{ B_{2n} \} - \text{Var} \{ B_{3n} \} \rightarrow 0 \) as \( \sigma^2_2 - \sigma^2_1 \rightarrow 0 \). For instance, taking \( \sigma^2_1 = 1, \sigma^2_2 = 2 \) we have
\[
\text{Var} \{ B_{1n} \} \rightarrow 15 \left\{ \frac{\sigma^6_1 + \sigma^6_2}{2} \right\} = 135 \frac{2}{2} = 67.5,
\]
\[
\text{Var} \{ B_{3n} \} \rightarrow 6 \left\{ \sigma^6_1 + \sigma^6_2 \right\} = 27,
\]
\[
\text{Var} \{ B_{2n} \} - \text{Var} \{ B_{3n} \} = \frac{9}{8} (\sigma^2_2 - \sigma^2_1)^2 (\sigma^2_2 + \sigma^2_1) = \frac{27}{8},
\]

whence
\[
\text{Var} \{ B_{2n} \} \rightarrow 6 \left\{ \frac{\sigma^6_1 + \sigma^6_2}{2} \right\} = 30.375.
\]

3. APPLICATION TO WMAP DATA

3.1. The Data

In order to test the effect of the linear term correction we applied the needlet and SMHW estimators to the foreground reduced \( V + W \) maps of the WMAP 7-year data. For simplicity we co-added the two frequency band maps with a constant noise weight. We analyzed the maps at HEALPix resolution \( N_{\text{side}} = 512 \) and masking out galactic foregrounds and point sources with the extended temperature analysis mask, known as KQ75 (Gold et al. 2011). Where Gaussian simulations are necessary we used the parameters from the WMAP 7+BAO+H_{0} cosmological data to simulate the CMB sky (Komatsu et al. 2011), then applying the beam and noise properties supplied by the WMAP team.

3.2. The Needlet/Wavelet f_{NL} Estimator with the Linear Term

The \( f_{NL} \) estimator based on the needlet bispectrum has been developed and applied to WMAP data in Rudjord et al. (2009, 2010). We recall the main features of this estimator. The needlet bispectrum without linear term was expressed as
\[
I_{j,k} \equiv \sum_{k} \beta_{j,k} \beta_{j,k} \beta_{j,k} \sigma_{j,k} \sigma_{j,k} \sigma_{j,k},
\]

where \( \beta_{j,k} \) is the needlet coefficient at scale \( j \) and direction, i.e., pixel, \( k \), defined in Equation (9), and \( \sigma_{j,k} \) is the expected standard deviation of \( \beta_{j,k} \). The needlet bispectrum is used to estimate \( f_{NL} \) by a \( \chi^2 \) minimization procedure:
\[
\chi^2(f_{NL}) = \mathbf{d}^T (f_{NL}) \mathbf{C}^{-1} \mathbf{d}(f_{NL}),
\]

where the data vector is
\[
\mathbf{d} = \hat{I}_{j,k} - f_{NL} \langle \hat{I}_{j,k} \rangle.
\]

Here, \( \mathbf{f}_{obs} \) is the bispectrum of the observed data and \( \langle \hat{I}_{j,k} \rangle \) is the average first-order non-Gaussian bispectrum obtained from non-Gaussian simulations. In this analysis, we used simulations with local-type non-Gaussianity generated with the algorithm described in Elsner & Wandelt (2009). The covariance matrix \( \mathbf{C} \) is obtained by means of Monte Carlo simulations:
\[
C_{ij} = \langle d_i d_j \rangle - \langle d_i \rangle \langle d_j \rangle.
\]

Differentiating Equation (49) yields the estimate
\[
f_{NL} = \langle \hat{I}_{j,k} \rangle \mathbf{C}^{-1} \mathbf{f}_{obs} \langle \hat{I}_{j,k} \rangle \mathbf{C}^{-1} \langle \hat{I}_{j,k} \rangle.
\]

Details of the estimation procedure can be found in Rudjord et al. (2009).

We want now to adopt the linear term correction in order to decrease the variance of the estimator (Equation (52)). Following Equation (32), it is straightforward to subtract the linear term from the bispectrum in Equation (48). The needlet bispectrum \( I_{obs, j,k} \) in the data vector (Equation (50)) can be expressed now as
\[
I_{obs, j,k} = \frac{1}{\sigma_{j,k} \sigma_{j,k} \sigma_{j,k}} \sum_{k} \{ \beta_{j,k} \beta_{j,k} \beta_{j,k} \}
\]

where
\[
\Gamma_{j,k}(k) = \langle \beta_{j,k} \beta_{j,k} \rangle.
\]

and similarly for the SMHW where the \( \beta_{j,k} \) are replaced by the \( w(R, n_k) \). We note that the contribution of the linear term to the average \( \langle \hat{I}_{j,k} \rangle \) in the data vector (Equation (50)) is vanishing, as the expected linear term value is always zero (Equation (35)). The terms \( \Gamma_{j,k}(k) \) contain one contribution from the CMB and one from the noise. The former is calculated using Monte Carlo simulations, the latter is calculated analytically.

We implemented slightly different algorithms of the \( f_{NL} \) estimator described above. The cubic statistic in Equations (48)
and (53) has been obtained in three cases, namely, with standard needle, Mexican needle, and SMHW coefficients.

### 3.3. Analysis of the WMAP Data

In order to test the effect of the linear term correction, we applied the \( f_{\text{NL}} \) estimator to the WMAP data described in Section 3.1, both before and after the addition of the linear term. The standard needle coefficients have been obtained with two different bases: \( B = 1.781 \) with scales \( j = 1–13 \) and \( B = 1.34 \) with \( j = 1–22 \). For the Mexican needles we chose \( B = 1.34 \) and \( p = 1 \) with \( j = 3–22 \). In all the cases in addition to the needle scales we considered a “scaled,” i.e., the original map not convolved with the needles. The scales selected for the SMHW are the same scales used in Curto et al. (2011a, 2011c): \( R_0 = 0 \) (the unconvolved map), \( R_1 = 2.9, R_2 = 4.5, R_3 = 6.9, R_4 = 10.6, R_5 = 16.3, R_6 = 24.9, R_7 = 38.3, R_8 = 58.7, R_9 = 90.1, R_{10} = 138.3, R_{11} = 212.3, R_{12} = 325.8, R_{13} = 500, \) and \( R_{14} = 767.3 \) arcmin. The SMHW coefficients have been analyzed in two ways: as they are or after subtracting the scale-by-scale average of the coefficients outside the applied mask. Thereafter we tested the same “mean subtraction” procedure with standard and Mexican needles. All the analyses are performed up to the multipole \( \ell = 1500 \).

With the exception of one case—where we analyzed the full sky with standard needles \( B = 1.781 \)—we always applied the KQ75 mask. For the SMHW analysis it is necessary to properly extend the mask at each scale, because the pixels near the border are affected by the zero values of the cut. Given the similarity with the SMHW, for comparison we extended the mask also for one case with the Mexican needles. Details of the mask extension procedure can be found in McEwen et al. (2005). The fractions of masked sky obtained with the extended masks at each scale are showed in Table 1.

For each analyzed case we simulated a set of 51,000 Gaussian maps with the same beam and noise properties of the WMAP co-added V+W cleaned map. To obtain the covariance matrix in Table 1

| Case                  | Linear Term | Error Bar 1σ | \( \Delta \sigma \) % | NG Sims. |
|-----------------------|-------------|--------------|----------------------|----------|
| SMHW                  |             |              |                      |          |
| \( B = 1.781 \) No    | //          | 18.4         | 5.0                  | 29.2     |
| Full sky Yes          | //          | 18.3         | 0.5                  | 29.2     |
| \( B = 1.781 \) No    | 63.5        | 25.4         | 12.6                 | 30.0     |
| KQ75 Yes              | 41.5        | 22.2         | 12.6                 | 30.0     |
| \( B = 1.34 \) No     | 43.6        | 24.9         | 12.6                 | 30.0     |
| KQ75 Yes              | 39.3        | 22.1         | 11.2                 | 29.5     |
| Mexican needles       |             |              |                      |          |
| \( B = 1.34, p = 1 \) | No          | 37.8         | 25.4                 | 29.5     |
| ext. KQ75 Yes         | Yes         | 26.6         | 12.6                 | 30.3     |
| \( B = 1.34, p = 1 \) | No          | 39.2         | 23.2                 | 29.9     |
| KQ75 Yes              | Yes         | 37.5         | 21.8                 | 6.0      | 29.8     |

### 3.4. Results

The addition of the linear term achieves a decrease in the standard deviation in all the cases. The full-sky analysis shows only a little improvement of the error bar (0.5%), indicating that the mask gives the major contribution to the correction operated by the linear term. Applying the KQ75 mask, all the \( f_{\text{NL}} \) values estimated on the V+W foreground reduced maps are consistent with the WMAP 7-year best estimate of \( f_{\text{NL}} = 32 \pm 21 \) (Komatsu et al. 2011). We began the analysis choosing the same needlet base \( B = 1.781 \) as in Rudjord et al. (2010). With respect to the previous result, \( f_{\text{NL}} = 73 \pm 31 \), we already obtained an 18% smaller error bar analyzing the 7-year data in place of the 5-year release and using more scales. But a further improvement of 12.6% is given by the linear term, demonstrating that the correction is non-negligible. The average \( \langle f_{\text{NL}} \rangle \) over the 700 non-Gaussian simulations shows

![Table 2](image-url)

### Notes.

- A Local \( f_{\text{NL}} \) on co-added V+W foreground reduced maps.
- Standard deviation over 10,200 Gaussian simulations.
- \( \sigma \) (no linear term)−\( \sigma \) (with linear term)\( / \sigma \) (no linear term).
- Average over 700 simulations with input \( f_{\text{NL}} = 30 \).
- Comparison with the “no mean subtracted” \( \sigma \). See the text for further details.
that the estimator is unbiased. We can note also that the $f_{NL}$ value estimated with the correction is closer to the WMAP result than without the linear term. We then chose a smaller needlet base $B = 1.34$, and therefore we constructed the cubic statistic of Equations (48) and (53) with more needlet scales. We indeed note a smaller standard deviation of 24.9 even without the linear correction with respect to $B = 1.781$. The error bar is decreased to 22.1, i.e., another 11.2%, adding the linear term.

Moving to the Mexican needlet case, we started with a conservative analysis applying an extended KQ75 mask as for the SMHW case. Despite the reduced sky coverage (Table 1) we found the same 1σ error bars obtained with standard needlets—$B = 1.781$ (but $f_{NL}$ values closer to WMAP). Motivated by the results of Scodeller et al. (2011), which have shown negligible neighbor bias effects of the mask, we repeated the Mexican needlet analysis with the original KQ75 mask. Looking at the results on the non-Gaussian simulations, we note that the $\{f_{NL}\}$ are indeed unbiased. In this case the linear term correction leads to a 6% improvement of the error bar, sufficient to achieve our best estimate $f_{NL} = 37.5 \pm 21.8$. This standard deviation is very close to the WMAP result with the optimal KSW estimator, where $\sigma = 21$. With the addition of the linear term correction the needlet $f_{NL}$ estimator is almost optimal. We believe that further standard deviation reductions can be achieved with an optimal $V + W$ co-adding and with a different choice of the needlet base $B$ and scales $j$. Anyway these exploitations are behind the primary scope of this work.

Furthermore we consider the case with the SMHW coefficients. The extensions of the KQ75 mask return sky coverage percentages close to Curto et al. (2011a, 2011c). Without subtracting the scale-by-scale coefficient average, the linear term correction leads to a reduction of the error bar from 27.3 to 21.9, equal to 20.1%. With a very similar SMHW analysis on co-added $V + W$ 7-yr cleaned maps, Curto et al. (2011c) found $f_{NL} = 32.5 \pm 23$ (Fisher matrix bound $\sigma = 22.5$), where we here considered the result uncorrected from point-source contribution. Therefore, the linear term addition rather improves their result. However, the analysis of Curto et al. (2011c) is performed after the scale-by-scale mean subtraction. Following the same procedure, we indeed achieved a reduction of the error bar without the linear term correction: from 27.3 to 21.9, equal to 18.3%. The $\sigma$ found subtracting the mean is indeed closer to the one of Curto et al. (2011c). Anyway we note that also in this case the addition of the linear term can still slightly improve the error bar (1.3%), and the $\sigma$ found with this correction—subtracting or not subtracting the mean—is almost identical.

Motivated by the SMHW analysis, we tested the scale-by-scale mean subtraction procedure with both the standard and Mexican needlets. We found that also in the needlet case this procedure well approximates the linear term correction, providing an error bar reduction of 9.6% (i.e., from 24.9 to 22.5) and 4.7% (i.e., from 23.2 to 22.1) for the standard and Mexican needlets, respectively. But again the addition of linear term provides even smaller error bars: 0.8% and 0.4%, respectively. Moreover we obtained the smallest $\sigma$ with the linear term but without subtracting the mean. We indeed got 22.1 against 22.3 for the standard needlets, and 21.8 against 22.0 for the Mexican needlets.

### 4. CONCLUSIONS

In this paper, we have derived for the first time the linear correction term for wavelet/needlet $f_{NL}$ estimators. As expected, under ideal experimental circumstances (isotropic noise, no masked regions) this term turns out to be identically zero. Under a realistic experimental set-up, the term is non-negligible, although smaller than for the KSW estimators; this is due to the localization properties of wavelet/needlet statistics, which soften to some extent the effects of unobserved regions and anisotropic noise. We have also argued that the linear correction term is well approximated by scale-by-scale mean subtraction, thus providing an explanations for recent results from Curto et al. (2011a, 2011b), where numerical estimates on the variance of wavelet estimators were shown to be very close to the KSW bound; in fact, we have shown that mean subtraction can cover approximately 85%–90% of the linear term effect under WMAP-like circumstances. The procedures we advocate are applied to WMAP 7-year $V + W$ foreground cleaned data, confirming that the corrections achieved are non-negligible. Applying the linear term correction we obtained the best estimate $f_{NL} = 37.5 \pm 21.8$. The error bar is very close to the optimal standard deviation $\sigma = 21$ found by the WMAP team (Komatsu et al. 2011).

In view of these results, we argue that wavelet/needlet statistics can provide a statistically sound and computationally convenient technique for non-Gaussianity analysis on CMB data.

F.K.H. acknowledges an OYI grant from the Norwegian Research Council; research by D.M. is supported by the European Research Council grant n.277742 (Pascal). This research has been partially supported by the ASI/INAF Agreement I/072/09/0 for the Planck LFI Activity of Phase E2 and by the PRIN 2009 project “La Ricerca di non-Gaussianit`a Primordiale.” Super computers from NOTUR (The Norwegian metacentre for computational science) have been used in this work. We acknowledge the use of the HEALPix software package (Górski et al. 2005) and the Legacy Archive for Microwave Background Data Analysis (LAMBDA) to retrieve the WMAP data set.

### REFERENCES

Acquaviva, V., Bartolo, N., Matarrese, S., & Riotto, A. 2003, Nucl. Phys. B, 667, 119
Baldi, P., Kerkyacharian, G., Marinucci, D., & Picard, D. 2009a, Ann. Stat., 37, 1150
Baldi, P., Kerkyacharian, G., Marinucci, D., & Picard, D. 2009b, Bernoulli, 15, 438
Bartolo, N., Komatsu, E., Matarrese, S., & Riotto, A. 2004a, Phys. Rept., 402, 103
Bartolo, N., Matarrese, S., & Riotto, A. 2004b, J. Cosmol. Astropart. Phys., JCAP01(2004)003
Bartolo, N., Matarrese, S., & Riotto, A. 2004c, J. High Energy Phys., JHEP04(2004)006
Buchbinder, E. I., Khoury, J., & Ovrut, B. A. 2008, Phys. Rev. Lett., 100, 171302
Chen, X. 2010, Adv. Astron., 2010, 638979 (arXiv:1002.1416)
Creminelli, P., Nicolis, A., Senatore, L., Tegmark, M., & Zaldarriaga, M. 2006, J. Cosmol. Astropart. Phys., JCAP05(2006)004
Curto, A., Martínez-González, E., & Barreiro, R. B. 2011a, MNRAS, 412, 1038
Curto, A., Martínez-González, E., Mukherjee, P., et al. 2009, MNRAS, 393, 615
Curto, A., Martínez-González, E., & Barreiro, B. 2011b, arXiv:1111.3390v1
Curto, A., Martínez-González, E., Barreiro, R. B., & Hobson, M. P. 2011c, MNRAS, 417, 488
Dvali, G., Gruzinov, A., & Zaldarriaga, M. 2004, Phys. Rev. D, 69, 023505
Elsner, F., & Wandelt, B. D. 2000, ApJS, 184, 264
Enqvist, K., & Sloth, M. S. 2002, Nucl. Phys. B, 626, 395
Fergusson, J., Liguori, M., & Shellard, E. P. S. 2009, Phys. Rev. D, 82, 023502
Fergusson, J., Liguori, M., & Shellard, E. P. S. 2010, (arXiv:1006.1642)
Fergusson, J., & Shellard, E. P. S. 2011, (arXiv:1105.2791)
Geller, D., Mayeli, A. 2009a, Math. Z., 262, 895
Geller, D., & Mayeli, A. 2009b, Math. Z., 263, 255
Gold, B., Odegard, N., Weiland, J. L., et al. 2011, ApJS, 192, 15
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
