MNRAS **434,** 796–805 (2013)
Advance Access publication 2013 July 5
doi:10.1093/mnras/stt1072

Exploring local $f_{\rm NL}$ estimators based on the binned bispectrum

B. Casaponsa,[1,2]★ R. B. Barreiro,[1] E. Martínez-González,[1] A. Curto,[1,3] M. Bridges[3] and M. P. Hobson[3]

[1]*Instituto de Física de Cantabria, CSIC-Universidad de Cantabria, Avda. de los Castros s/n, 39005 Santander, Spain*
[2]*Dpto. de Física Moderna, Universidad de Cantabria, Avda. los Castros s/n, 39005 Santander, Spain*
[3]*Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge CB3 0HE, UK*

Accepted 2013 June 12. Received 2013 June 5; in original form 2013 April 30

ABSTRACT

We explore different estimators of the local non-linear coupling parameter, $f_{\rm NL}$, based on the binned bispectrum presented in Bucher et al. Using simulations of *Wilkinson Microwave Anisotropy Probe* (*WMAP*)-7-year data, we compare the performance of a regression neural network with a χ^2-minimization and study the dependence of the results on the presence of the linear term in the analysis and on the use of inpainting for masked regions. Both methods obtain similar results and are robust to the use of inpainting, but the neural network estimator converges considerably faster. We also examine the performance of a simplified χ^2 estimator that assumes a diagonal matrix and has the linear term subtracted, which considerably reduces the computational time; in this case inpainting is found to be crucial. The estimators are also applied to real *WMAP*-7-year data, yielding constraints at 95 per cent confidence level of $-3 < f_{\rm NL} < 83$.

Key words: methods: data analysis – methods: statistical – cosmic background radiation.

1 INTRODUCTION

Cosmic microwave background (CMB) fluctuations naturally arise in inflationary models. Discriminating between different models is a difficult task, but can be addressed by observing very faint non-Gaussian signals in the high-order correlation functions of the CMB temperature anisotropies. A popular approach is to search for the local form of non-Gaussianity, where the initial curvature Gaussian perturbations are expanded up to the second order as

$$\Phi = \Phi_{\rm g} + f_{\rm NL}\left[\Phi_{\rm g}^2 - \left\langle \Phi_{\rm g}^2 \right\rangle\right]$$

(for more details see e.g. Babich, Creminelli & Zaldarriaga 2004; Bartolo et al. 2004).

Wilkinson Microwave Anisotropy Probe (*WMAP*) constraints on the amplitude of the local form of non-Gaussianity have been able to rule out exotic models such as ghost inflation (Arkani-Hamed et al. 2004). New data sets, such as the recent release from *Planck* satellite (Planck Collaboration 2013), significantly reduce the uncertainties on local $f_{\rm NL}$, ruling out some ekpyrotic models and imposing strong constraints on multifield inflationary models. In fact, for single-field inflation, $f_{\rm NL}$ (hereafter $f_{\rm NL}$ is the local form) should be of the order of the spectral index (Creminelli & Zaldarriaga 2004), given the consistency relation derived in Maldacena (2003). Recent papers show that this relation does not hold for non-vacuum initial states (Agullo & Parker 2011; Ganc 2011) and non-constant superhorizon modes (Chen et al. 2013), but the vast majority of single-field models should be ruled out by a detection of a larger $f_{\rm NL}$ value.

This type of primordial non-Gaussianity may be detected using higher order correlation functions. The simplest of these is third order, which is equivalent to the bispectrum in spherical harmonic space. The first derivation of the optimal estimator, in the sense of an unbiased estimator that saturates the Cramer–Rao inequality, is given in Babich (2005), assuming an isotropic field. Working with real data, however, is usually more complicated. In particular, CMB maps have anisotropic noise due to the scanning strategy and masked regions, both of which break the isotropy assumption for these theoretical estimators. The masked regions are particularly difficult to treat, as they introduce correlations among the Fourier modes, which are otherwise expected to be independent. Creminelli et al. (2006) applied the optimal estimator to real data, showing that the presence of a term proportional to the $a_{\ell m}$ is required to account for such anisotropies. In that paper the constraints are computed using an approximation to avoid numerical difficulties. Finally, this estimator was successfully applied in its complete form to *WMAP* data by Smith, Senatore & Zaldarriaga (2009), Komatsu et al. (2011) and Bennett et al. (2012) for 5th, 7th and 9th year, respectively.

New imaging reconstruction techniques have recently been used to pre-process CMB maps by smoothing the contours of the masked

★E-mail: casaponsa@ifca.unican.es

regions. A simple approach is to apodize the mask by introducing a smooth function in the pixels surrounding the masked regions. Another approach is to fill the masked regions with a pseudo-signal, which is termed inpainting. Several techniques have been proposed in the literature for inpainting, which is a very delicate process since the signal can be distorted (Bajkova 2005; Abrial et al. 2008; Starck, Fadili & Rassat 2013).

Consequently, primordial non-Gaussianity analyses can be computationally demanding, and new techniques should therefore be investigated to overcome the computational cost of large matrix estimations and inversions. Here we investigate the utility of a neural network to obtain the necessary weights in the $f_{\rm NL}$ estimator and compare it with the direct approach via χ^2 minimization. Over the last 20 yr, artificial intelligence techniques have been use in a number of areas of astrophysical analysis: morphological galaxy determination, photoredshift estimations and classification of different objects are examples of successful applications of neural networks (Storrie-Lombardi et al. 1992; Firth, Lahav & Somerville 2003; Vanzella et al. 2004; Carballo et al. 2008). In particular, for cosmological analysis, they have recently been used to reduce the computational time of cosmological parameter estimation from observations of the CMB power spectrum (Auld et al. 2007; Auld, Bridges & Hobson 2008). Also in CMB analysis, Casaponsa et al. (2011b) used neural networks to define a new non-Gaussianity estimator and showed that networks are a valuable tool for bypassing the inversion of ill-conditioned matrices, and to avoid covariance matrix estimation in a χ^2 analysis.

The aim of the present work is to continue our earlier study of the power of the neural networks in the statistical analysis associated with CMB non-Gaussianity. To this end, this paper is focused on the study of different tools, in order to identify the most robust and efficient estimator when dealing with real data. We compare three different approaches to estimate $f_{\rm NL}$, based on the binned bispectrum. The first estimator is obtained by minimizing a χ^2 of the binned bispectrum components. A second approach is based on the optimal estimator, without taking into account the correlations among the binned bispectrum components, which for an isotropic field would be the same as the former. And the third method uses the weights of a regression neural network. From these approaches we construct different estimators to account for the effects of preprocessing the data with inpainting and the presence of a linear term.

The paper is organized as follows. An overview of the type of neural network employed and the training procedure is given in Section 2. In Section 3 we describe the binned bispectrum. The definition of the estimators is presented in Section 4 followed by an explanation of the main details of the implementation in Section 5. The results are presented in Section 6, and finally the conclusions are summarized in Section 7.

2 NEURAL NETWORKS

Artificial neural networks (ANN) are a methodology for computing, based on massive parallelism and redundancy, which are features also found in animal brains. They consist of a number of interconnected nodes each of which processes information and passes it to other nodes in the network. Well-designed networks are able to 'learn' from a set of training data and to make predictions when presented with new, possibly incomplete, data. These algorithms have been successfully applied in several areas, in particular, we note the following applications in cosmology: Baccigalupi et al. (2000), Firth et al. (2003), Ball et al. (2004), Auld et al. (2007, 2008), Casaponsa et al. (2011b) and Nørgaard-Nielsen (2012).

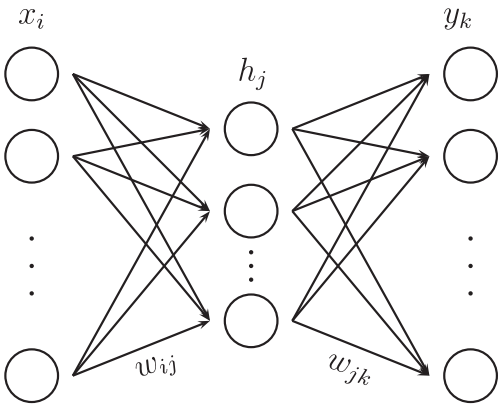

Figure 1. Schematic diagram of a three-layer feed-forward neural network.

The basic building block of an ANN is the *neuron* or *node*. Information is passed as inputs to the neuron, which processes them and produces an output. The output is typically a simple mathematical function of the inputs. The power of the ANN comes from assembling many neurons into a network. The network is able to model very complex behaviour from input to output. We use a three-layer feed-forward network consisting of a layer of input neurons, a layer of 'hidden' neurons and a layer of output neurons. Fig. 1 shows a schematic design of such a network.

The outputs of the hidden layer and the output layer are related to their inputs as follows:

$$\text{hidden layer: } h_j = g^{(1)}\left(f_j^{(1)}\right);\ f_j^{(1)} = \sum_i w_{ji}^{(1)} x_i + \theta_j^{(1)}, \tag{1}$$

$$\text{output layer: } y_k = g^{(2)}\left(f_k^{(2)}\right);\ f_k^{(2)} = \sum_j w_{kj}^{(2)} h_j + \theta_k^{(2)} \tag{2}$$

for each hidden node j and each output node k. The index i runs over all input nodes. The functions $g^{(1)}$ and $g^{(2)}$ are called activation functions. The non-linear nature of $g^{(1)}$ is a key ingredient in constructing a viable and practically useful network. This non-linear function must be bounded, smooth and monotonic; we use $g^{(1)}(x) = \tanh x$. For $g^{(2)}$ we simply use $g^{(2)}(x) = x$. The layout and number of nodes are collectively termed the *architecture* of the network. For a basic introduction to artificial neural networks the reader is directed to Mackay (2003) and Golden (1996).

For a given architecture, the weights $\boldsymbol{w}$ and biases $\boldsymbol{\theta}$ define the operation of the network and are the quantities we wish to determine by some *training* algorithm. Basically, the training process is an iterative algorithm that optimizes a given objective function that quantifies the accuracy of the network outputs. We denote $\boldsymbol{w}$ and $\boldsymbol{\theta}$ collectively by the network parameters $\boldsymbol{a}$. As these parameters vary during training, a very wide range of non-linear mappings between inputs and outputs is possible. In fact, according to a 'universal approximation theorem' (Leshno 1993), a standard three-layer feed-forward network can approximate any continuous function to *any* degree of accuracy with appropriately chosen activation functions and a sufficient number of hidden nodes.

In our previous application of ANN to the estimation of $f_{\rm NL}$, a classification neural network was used (Casaponsa et al. 2011b). Here, we instead use a *regression* network, which we find to be as useful as the classification approach, and also allows a more direct comparison with the χ^2 minimization procedure. Additionally, using a regression network has the advantage of reducing the network parameter space, making the training faster.

In Casaponsa et al. (2011b), we used neural networks for which the inputs were third-order moments of two wavelet decompositions of the CMB map: the Healpix wavelet (HW; Casaponsa et al. 2011a) and the spherical Mexican hat wavelet (SMHW; Curto, Martínez-González & Barreiro 2009, 2011). We found the resulting $f_{\rm NL}$ estimator had the same accuracy as the standard one based on χ^2-minimization, but was much faster to evaluate. Here, the inputs to our neural networks are the estimator for the bispectrum proposed by Bucher, van Tent & Carvalho (2010), defined in a number of bins in l-space, which reduces the dimension of the problem by a factor of 10^5. Our aim is to learn a mapping from the binned bispectrum components of the (possibly) non-Gaussian CMB (assembled into an input feature vector $\boldsymbol{x}$) to the corresponding $f_{\rm NL}$ of the map; this is discussed in more detail below.

A suitable objective function for this problem is

$$\mathcal{L}(\boldsymbol{a}) = \frac{1}{2}\sum_n \sum_k \left[t_k^{(n)} - y_k^{(n)}(\boldsymbol{x}^{(n)}, \boldsymbol{a})\right]^2, \tag{3}$$

where the index n runs over the training data set $\mathcal{D} = \{\boldsymbol{x}^{(n)}, \boldsymbol{t}^{(n)}\}$, in which the target vector $\boldsymbol{t}^{(n)}$ for the network outputs are the $f_{\rm NL}$ values, as explained in the next section. One then wishes to find network parameters $\boldsymbol{a}$ that minimize this objective function as the training progresses. This is, however, a highly non-linear, multimodal function in many dimensions whose optimization poses a non-trivial problem. We perform this optimization using the MemSys package (Gull & Skilling 1999). This algorithm considers the parameters $\boldsymbol{a}$ to have prior probabilities proportional to $\mathrm{e}^{\alpha S(\boldsymbol{a})}$, where $S(\boldsymbol{a})$ is the positive–negative entropy functional (Hobson & Lasenby 1998), and α is a hyperparameter of the prior that sets the scale on which variations in $\boldsymbol{a}$ are expected. The value of α is chosen to maximize its marginal posterior probability, value of which is inversely proportional to the standard deviation of the prior. Thus for a given α, the log-posterior probability is proportional to $\mathcal{L}(\boldsymbol{a}) + \alpha S(\boldsymbol{a})$. For each chosen α there is a solution $\hat{\boldsymbol{a}}$ that maximizes the posterior. As α varies, the set of solutions $\hat{\boldsymbol{a}}$ is called the *maximum-entropy trajectory*. We wish to find the solution for which $\mathcal{L}$ is minimized which occurs at the end of the trajectory where $\alpha = 0$. For practical purposes we start at a large value of α and iterate downwards until α is sufficiently small so that the posterior is dominated by the $\mathcal{L}$ term. MemSys performs this algorithm using conjugate gradient descent at each step to converge to the maximum-entropy trajectory. The required matrix of second derivatives of $\mathcal{L}$ is approximated using vector routines only, thus circumventing the need for $O(N^3)$ operations required for exact calculations. The application of MemSys to the problem of network training allows for the fast efficient training of relatively large network structures on large data sets that would otherwise be difficult to perform in a reasonable time. Moreover the MemSys package also computes the Bayesian evidence for the model (i.e. network) under consideration (see e.g. Jaynes & Bretthorst 2003, for a review), which provides a powerful model selection tool. In principle, values of the evidence computed for each possible architecture of the network (and training data) provide a mechanism to select the most appropriate architecture, which is simply the one that maximizes the evidence.

3 BINNED BISPECTRUM

Several approaches to bispectrum analyses have been proposed in the literature, such as the KSW (Komatsu, Spergel & Wandelt 2005), Skew-Cls (Munshi & Heavens 2010), wavelets (Curto et al. 2009, 2011; Casaponsa et al. 2011a) or needlets (Pietrobon et al. 2009; Donzelli et al. 2012) among others. Furthermore, Bucher et al. (2010) and Fergusson & Shellard (2011) presented bispectrum estimators that reduce the dimensionality of the problem without losing significant information. In particular, we use the bispectrum estimator defined in Bucher et al. (2010). The proposed method consists of joining the bispectrum components in bins, significantly reducing the computational time, but maintaining the quality of the estimator of $f_{\rm NL}$. Bucher et al. (2010) show that this is the case for ideal maps, with isotropic noise and small symmetric masks. The binned bispectrum is also applied to *Planck* data in Planck Collaboration (2013) to constrain primordial non-Gaussianity. Here we study with more detail its applications to realistic data, for which we used simulations with *WMAP*-7-year characteristics.

As a starting point, the angle-averaged reduced bispectrum is defined by

$$b_{l_1 l_2 l_3} = \int T_{\ell_1} T_{\ell_2} T_{\ell_3}\, \mathrm{d}\Omega, \tag{4}$$

where $T_\ell(\boldsymbol{n}) = \sum_m a_{\ell m} Y(\boldsymbol{n})$. The binned reduced bispectrum is then

$$b_{\rm abc} = \sum_{\ell_1 \in I_{\rm a}} \sum_{\ell_2 \in I_{\rm b}} \sum_{\ell_3 \in I_{\rm c}} b_{\ell_1 \ell_2 \ell_3}, \tag{5}$$

where I_n are bins in ℓ. This definition of the reduced bispectrum, differing from the standard one by the factor $I^2_{\ell_1 \ell_2 \ell_3}$ (for details see Komatsu 2002; Bucher et al. 2010), is convenient since one can write $b_{\rm abc}$ in terms of $T_{\rm a}$, $T_{\rm b}$ and $T_{\rm c}$ which are the binned maps:

$$T_n = \sum_{\ell_i \in I_n} T_{\ell_i}. \tag{6}$$

The advantage of constructing maps in ℓ-bins is that the number of transformations to spherical harmonic space is significantly reduced. Then, the resulting bispectrum estimator is faster to construct than the one based on the KSW estimator (Komatsu et al. 2005) or the SMHW (Curto et al. 2011).

4 $f_{\rm NL}$ ESTIMATORS

The optimal estimator for $f_{\rm NL}$, in the sense of an unbiased estimator that saturates the Cramer–Rao inequality, is obtained by performing an Edgeworth expansion of the probability distribution of the $a_{\ell m}$ for weakly non-Gaussian data (Babich 2005; Creminelli et al. 2006; Smith et al. 2009). This estimator is found to have a cubic term and a linear term in $a_{\ell m}$. The latter term plays an important role under realistic conditions, where anisotropic instrumental noise and/or a mask are present.

The form of this estimator can also be understood using the properties of the Wick product. As demonstrated in Donzelli et al. (2012), Marinucci & Peccati (2011) and Peccati & Taqqu (2011), the Wick product of a cubic variable, which is given by

$$:x_1, x_2, x_3: = x_1 x_2 x_3 - x_1 \langle x_2 x_3 \rangle - x_2 \langle x_1 x_3 \rangle - x_3 \langle x_1 x_2 \rangle, \tag{7}$$

has a smaller variance than the cubic variable itself, while not affecting the mean value as long as the variables x_i are Gaussian and have a mean value of zero. Then, if we replace each cubic term in an estimator by its Wick product, it will yield an estimator with lower

variance. Following this reasoning, the binned bispectrum defined in Section 3 can be replaced by its Wick product

$$:T_{I_a}T_{I_b}T_{I_c}: = T_{I_a}T_{I_b}T_{I_c} - \left\langle T_{I_a}T_{I_b}\right\rangle T_{I_c} - \left\langle T_{I_b}T_{I_c}\right\rangle T_{I_a} - \left\langle T_{I_a}T_{I_c}\right\rangle T_{I_b}. \quad (8)$$

Note that $T_i = T_i(x)$, since there is a dependence on the pixel for anisotropic maps.

Donzelli et al. (2012) have proved that for the case of wavelet and needlet coefficients, the linear term is basically equivalent to removing the mean value of the coefficients. In order to see if this is the case for the binned bispectrum, we explore the option of substituting $T'_n = T_n - \langle T_n\rangle$, where $\langle T_n\rangle$ is computed with the unmasked pixels. This would be less costly than estimating the correlation matrix $\langle T_{\rm a}T_{\rm b}\rangle$ required for the computation of the linear term.

In the following subsections, we describe three methods for choosing the weights that are used to construct the final $f_{\rm NL}$ estimator. In each case, estimators are constructed with and without the linear term contribution to explore its importance. Also, the performance of these estimators is tested on inpainted and non-inpainted maps, with the methodology explained in Section 5.3. In all cases the original mask M is applied again at the final stage when computing the binned bispectrum components:

$$b_{\rm abc} = \sum_{i=1}^{N_{\rm pix}} \frac{M_i(T_{{\rm a},i}T_{{\rm b},i}T_{{\rm c},i})}{4\pi N_{\rm pix}}, \quad (9)$$

where $N_{\rm pix} = \sum_i M_i$. The efficiency achieved by the estimators will be compared to that defined by the Cramer–Rao inequality. The Cramer–Rao bound states that the minimum variance for any unbiased estimator is given by the inverse of the Fisher matrix information. A useful reference value in the case of partial sky coverage is obtained from the full sky estimator corrected by the fraction of the available sky. Therefore, the minimum variance for $f_{\rm NL}$ is estimated to be

$$\sigma^2_{\rm fh} = \left[f_{\rm sky}\sum_{\ell_1\le\ell_2\le\ell_3}\frac{\left(\langle B_{\ell_1\ell_2\ell_3}\rangle^1\right)^2}{\Delta C_{\ell_1}C_{\ell_2}C_{\ell_3}}\right]^{-1}, \quad (10)$$

where Δ takes values 1, 2 or 6 when all ℓs are different, two are equal or all are the same and $f_{\rm sky}$ is the fraction of the sky available. For (10) to be used for a realistic case, the power spectrum must include the noise and the beam contribution. The beam also needs to be included in the bispectrum part. We have used *WMAP*-7-year characteristics, in particular the average of the two channels of 61 and 94 GHz (V and W) and the extended mask KQ75. In terms of the reduced bispectrum defined in Section 3, the angular average bispectrum $B_{\ell_1\ell_2\ell_3}$ is

$$B_{\ell_1\ell_2\ell_3} = \sqrt{\frac{4\pi}{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}} \times \begin{pmatrix}\ell_1 & \ell_2 & \ell_3\\ 0 & 0 & 0\end{pmatrix}^{-1} b_{\ell_1\ell_2\ell_2}. \quad (11)$$

4.1 Approximated maximum likelihood estimator (AMLE)

The standard approach in this type of analysis is to use the fact that the third-order moments are nearly Gaussian, and therefore the maximum-likelihood estimator is obtained approximately by the minimization of a χ^2 given by

$$\chi^2 = \sum_{\rm abc,def}\left(b_{\rm abc} - f_{\rm NL}\langle b_{\rm abc}\rangle^1\right)C^{-1}_{\rm abc,def}\left(b_{\rm def} - f_{\rm NL}\langle b_{\rm def}\rangle^1\right), \quad (12)$$

where $\langle b_{\rm def}\rangle^1$ is the expected value for $f_{\rm NL} = 1$ and $C^{-1}_{\rm abc,def} = \langle b_{\rm abc}\rangle\langle b_{\rm def}\rangle - \langle b_{\rm abc}b_{\rm def}\rangle$. From the previous equation is straightforward to show that the $f_{\rm NL}$ estimator for an observed map is

$$f_{\rm NL} = \sum_{\rm abc,def}\frac{\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}b^{\rm obs}_{\rm def}}{\sum_{\rm abc,def}\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}\langle b_{\rm def}\rangle^1}. \quad (13)$$

In order to include the linear term correction, $T_{\rm a}T_{\rm b}T_{\rm c}$ should be substituted by its Wick product (8), wherever it appears. The mean value of the linear term is zero, and thus it vanishes in the term of the estimator related to the model, whereas it needs to be included in the covariance matrix. Thus, the corresponding estimator is

$$f_{\rm NL} = \sum_{\rm abc,def}\frac{\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}}{\sum_{\rm abc,def}\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}\langle b_{\rm def}\rangle^1} \times\left(\frac{1}{4\pi N_{\rm pix}}\sum_i^{N_{\rm pix}} T_{{\rm d},i}T_{{\rm e},i}T^{\rm obs}_{{\rm f},i} - \langle T_{{\rm d},i}T_{{\rm e},i}\rangle T^{\rm obs}_{{\rm f},i} - \langle T_{{\rm d},i}T_{{\rm f},i}\rangle T^{\rm obs}_{{\rm e},i} - \langle T_{{\rm e},i}T_{{\rm f},i}\rangle T^{\rm obs}_{{\rm d},i}\right), \quad (14)$$

where $\langle b_{\rm abc}\rangle^1$ is estimated using the regression coefficient of a linear fit to the mean values of 1000 non-Gaussian simulations with different $f_{\rm NL}$ values. In particular, this is a conservative number that ensures that the mean values have converged.[1] For C^{-1} we assume that it is independent of $f_{\rm NL}$, which is a good approximation in the limit of weak non-Gaussianity, and it is thus estimated with Gaussian simulations ($\sim$25 000 are necessary for convergence issues). The term $\langle T_{\rm a}T_{\rm b}\rangle$ is estimated with 1000 Gaussian simulations.

4.2 Approximated maximum likelihood estimator with diagonal covariance matrix (AMLED)

The estimator proposed by Bucher et al. (2010) used the approximation of assuming a diagonal covariance matrix. In this case, the estimator simplifies significantly, since the covariance matrix does not need to be estimated or inverted, and one obtains

$$f_{\rm NL} = \sum_{\rm abc}\frac{\langle b_{\rm abc}\rangle^1 b^{\rm obs}_{\rm abc}/{\rm var}(b_{\rm abc})}{\sum_{\rm def}(\langle b_{\rm def}\rangle^1)^2/{\rm var}(b_{\rm def})}, \quad (15)$$

where ${\rm var}(b_{\rm abc})$ is the variance of the binned bispectrum components, which is computed with simulations. Besides its computational efficiency, another advantage of this estimator is that can be obtained analytically (see Bucher et al. 2010, for details).

Strictly speaking, this estimator is optimal only for a full-sky CMB experiment with isotropic noise (although it has been shown to work well also in presence of a reduced symmetric mask). Under realistic conditions, a linear term of a similar form to that used

[1] To check that convergence is reached with N simulations, we simply test that two independent sets of $N/2$ realizations give consistent results.

above needs to be added, such that

$$f_{\rm NL} = \sum_{\rm abc} \frac{\langle b_{\rm abc}\rangle^1/{\rm var}(b_{\rm abc})}{\sum_{\rm def}(\langle b_{\rm def}\rangle^1)^2/{\rm var}(b_{\rm def})} \left(\frac{1}{4\pi N_{\rm pix}} \sum_i^{N_{\rm pix}} T_{{\rm a},i} T_{{\rm b},i} T^{\rm obs}_{{\rm c},i} - \langle T_{{\rm a},i} T_{{\rm b},i}\rangle T_{{\rm c},i} - \langle T_{{\rm a},i} T_{{\rm c},i}\rangle T_{{\rm b},i} - \langle T_{{\rm b},i} T_{{\rm c},i}\rangle T_{{\rm a},i} \right). \quad (16)$$

As with the previous estimator, 1000 simulations were used for the model estimation and another 1000 to obtain var($b_{\rm abc}$). This implies a reduction by a factor >10 in the number of simulations required with respect to the AMLE, as the convergence in the estimation of the variance is reached with much smaller number of realizations than that of the covariance matrix.

4.3 Neural network estimator (NNE)

The architecture of our three-layer neural network is defined by three parameters: the number of input, output and hidden nodes. The first two are determined by the problem at hand; in this case the dimension of the input vector depends on the number of bins chosen and there is a single output.

Although the MemSys algorithm provides routines to determine the optimal value of the number of hidden nodes using the Bayesian evidence (Gull & Skilling 1999), in this application $n_{\rm hid}$ is determined empirically by measuring the accuracy of the trained networks on an independent testing set. In this application, we have found that in fact the optimal architecture contains no hidden nodes, resulting in what is effectively a linear mapping between input and output. This is not surprising, since we are effectively 'asking' the network to learn the mean value and dispersion of the binned bispectrum components for each $f_{\rm NL}$; since the expectation value is linearly dependent on the $f_{\rm NL}$, this network architecture trivially satisfies this requirement. Indeed, networks of this sort provide a simple way of obtaining the (pseudo)inverse of any matrix.

Then, for zero hidden nodes, the single network output is just a linear function of the inputs. Once the network parameters $(\boldsymbol{w}, \theta)$ are found during the training process, the estimator for $f_{\rm NL}$ is thus given by

$$f_{\rm NL} = \sum_{\rm abc} w_{\rm abc} b_{\rm abc} + \theta. \quad (17)$$

As with the previous estimators the network is also trained including the linear term, in which case

$$f_{\rm NL} = \sum_{\rm abc} w_{\rm abc} \left(\frac{1}{4\pi N_{\rm pix}} \sum_i^{N_{\rm pix}} T_{{\rm a},i} T_{{\rm b},i} T_{{\rm c},i} - \langle T_{{\rm a},i} T_{{\rm b},i}\rangle T_{{\rm c},i} - \langle T_{{\rm a},i} T_{{\rm c},i}\rangle T_{{\rm b},i} - \langle T_{{\rm b},i} T_{{\rm c},i}\rangle T_{{\rm a},i} \right) + \theta. \quad (18)$$

Comparing with the AMLE estimator, we can see that it is equivalent to a neural network with parameters

$$w_{\rm def} \mapsto \sum_{\rm abc} \frac{\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}}{\sum_{\rm abc,def}\langle b_{\rm abc}\rangle^1 C^{-1}_{\rm abc,def}\langle b_{\rm def}\rangle^1}, \quad (19)$$

$$\theta \mapsto 0. \quad (20)$$

If this were the optimal linear combination to estimate $f_{\rm NL}$, the neural network would find the same result as the AMLE but avoiding all the expensive calculations required in the direct computation of this estimator (provided that we have chosen a linear combination for the NNE). Conversely, if that combination were not optimal, the network should be able to find different, more optimal, weights. For instance, for the AMLE to be optimal, the considered statistics should follow a Gaussian distribution, whereas the NNE does not make any assumptions about the intrinsic distribution of the inputs. Therefore, the neural network is expected to perform better when working with non-Gaussian statistics. In addition, the neural network does not require to assume that the covariance matrix is independent of $f_{\rm NL}$. Even if this approximation works well for the current application, it may not always be the case, which would significantly complicate the calculation of the AMLE. In such cases the NNE would represent a clear advantage over the χ^2 minimization. Finally, we would also like to point out that, although for the current application a linear combination was found to be the best choice for the NNE, in a general case, this estimator is not restricted to a linear combination of the inputs, which can be useful in other problems.

5 IMPLEMENTATION

In this section the non-Gaussian simulations used for the analyses as well as some technical details required for the implementation of the estimators are described.

5.1 Non-Gaussian simulations

Two different sets of non-Gaussian realizations are used. A set generated with the map-making method proposed in Fergusson, Liguori & Shellard (2010) and described also in Curto et al. (2011), and a set of publicly available realizations[2] generated by Elsner & Wandelt (2009). In the first method, the non-Gaussian part of the map ($a^{\rm NG}_{\ell m}$) is taken directly from the theoretical bispectrum, while the second algorithm starts from the primordial curvature fluctuations and is therefore more precise.

The two different sets are used for the following reasons. Having a large number of independent realizations is necessary to train the network, as well as to test its performance with the number of training data. Since the first set is faster to produce, 30 000 independent realizations were generated as in Curto et al. (2011). In the analysis with the SMHW of Curto et al. (2011), they found the dispersion on $f_{\rm NL}$ to be slightly larger than using the simulations of set 2. In Curto et al. (2011), constraints on $f_{\rm NL}$ are obtained with both sets finding a discrepancy of 5 per cent. We find similar deviations for the binned bispectrum. This is observed if the average bispectrum at the numerator in (10) is computed with simulations with both sets. Then, as the model of set 1 is given by an approximation, the minimum dispersion of the parameter obtained with realizations is slightly larger than using the analytical dispersion in equation (10). Conversely, using realizations of set 2 we find a closer value to the analytically computed lower bound.

Hence, after proving that the NNE converges with few thousand realizations for the best performing form of the estimator, the second set is used for the final results. This is convenient to be able to compare our results with the Fisher dispersion of (10), and with the ones obtained with the optimal estimator (Komatsu et al. 2011), where simulations equivalent to the ones of set 2 are used.

The Gaussian and non-Gaussian harmonic coefficients of the CMB realizations, $a^{\rm NG}_{\rm lm}$ and $a^{\rm G}_{\rm lm}$, either generated from set 1 or set 2,

[2] http://planck.mpa-garching.mpg.de/cmb/fnl-simulations/

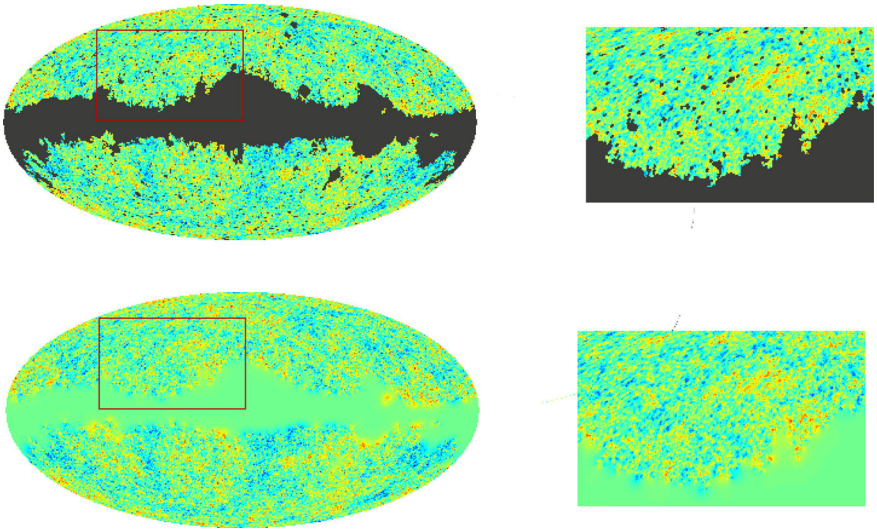

Figure 2. Inpainting effect shown in the masked *WMAP*-7-year map. On the top the initial temperature map with the mask in dark grey and an amplified region are presented and on the bottom, the same map and region are given after inpainting.

are combined to obtain the non-Gaussian realization with different values of $f_{\rm NL}$:

$$a_{\rm lm} = a_{\rm lm}^{\rm G} + f_{\rm NL} a_{\rm lm}^{\rm NG}. \quad (21)$$

Noise-weighted $V + W$ band *WMAP*-7-year realizations were then constructed as explained in Curto et al. (2009) and Casaponsa et al. (2011a), and the KQ75 mask was then applied, which covers roughly 29 per cent of the sky.

5.2 Binning scheme

One is free to choose the number and size of the bins in ℓ-space for the binned bispectrum. Bucher et al. (2010) found that for $\ell_{\rm max} = 2000$ and 64 bins the results obtained were 99.3 per cent of the optimal value. For an application to *WMAP*, one has $\ell_{\rm max} = 1024$, so the corresponding number of bins is 32. We have tested the performance of the estimators with different number of bins and find that for $n_{\rm bin} = 28$ the results have converged. Therefore, the following results use this number of bins, which also provides a modest saving in computation with respect to 32 bins. Conversely to the exhaustive choosing of the binning scheme done in Bucher et al. (2010) estimator, here we simply use logarithmic bins. The logarithmic scale is chosen by imposing the condition that all bins have at least one ℓ.

The binned bispectrum components are computed from combinations of three binned maps $T_{\rm a}T_{\rm b}T_{\rm c} = \sum_{\ell_1 \in I_{\rm a}} \sum_{\ell_2 \in I_{\rm b}} \sum_{\ell_3 \in I_{\rm c}} T_{\ell_1} T_{\ell_2} T_{\ell_3}$. It can be noticed that some of the combinations $\ell_1\ell_2\ell_3$ might not satisfy the triangle condition ($\ell_3 - \ell_2 \leq \ell_1 \leq \ell_2 + \ell_3$). To avoid as far as possible those non-contributing combinations, we discard the binned bispectrum components where all the contained ℓ combinations do not meet the triangle condition. For that reason the components used are the ones that hold the following condition:

$$\ell_{I_{\rm c}}^{\rm min} - \ell_{I_{\rm b}}^{\rm max} \leq \ell_{I_{\rm a}}^{\rm max} \leq \ell_{I_{\rm c}}^{\rm max} + \ell_{I_{\rm b}}^{\rm max},$$

where $\ell_{I_n}^{\rm min}$ and $\ell_{I_n}^{\rm max}$ are the minimum and maximum value of ℓ of the bin I_n. Then, the binned bispectrum for $n_{\rm bin} = 28$ consists of 1077 components, whereas the full bispectrum would have $\sim 10^8$ components.

5.3 Inpainting

Several inpainting methods have been developed for general imaging reconstruction (see e.g. the review by Bertalmio et al. 2000). The goal of these methods is to restore missing or damaged regions of an image to recover the original signal as far as possible. For CMB map reconstruction, the ideal inpainting method would lead to a restored map preserving the statistical properties of the unmasked map.

Different approaches have been used to reduce the discontinuities generated by the mask edges in CMB maps, since they introduce undesirable correlations among the binned bispectrum components. As the intention here is to reduce this impact, rather than reconstruct the full map, we use a simple iterative process that averages over the direct neighbours of the masked pixels, and is based on the work of Oliveira et al. (2001).

One begins with the map $T(\boldsymbol{x})$ and the binary mask $M(\boldsymbol{x})$. Then each pixel of the masked map $T' = TM$ with value zero is substituted by the average of its immediate neighbours, whether masked or not, using the HEALPix subroutine NEIGHBOURS. The process is repeated 1000 times, leaving the masked point sources completely inpainted and smoothing the edges of the galactic mask. The results of this process are illustrated in Fig. 2. We find that, in this case, the technique is more effective than simply using an apodized mask.

5.4 Neural network training process

To train our $f_{\rm NL}$ network we provide it with an ensemble of training data $\mathcal{D} = \{\boldsymbol{x}^{(n)}, t^{(n)}\}$. The nth input vector $\boldsymbol{x}^{(n)}$ contains the binned bispectrum components, explained in Section 3, of the ith simulated CMB map. The output target is the corresponding $f_{\rm NL}$ value of the ith CMB simulation. Thus, for $n_{\rm bin} = 28$ the input vector has 1077 components, and the target vector $t^{(n)}$ for the network consists of only one component. From the training set, 20 per cent of the realizations are reserved for the validation process.

The network weights are computed during the training procedure, which in this case requires only a few seconds. The performance of the network is validated during the training process using an independent set of testing data.

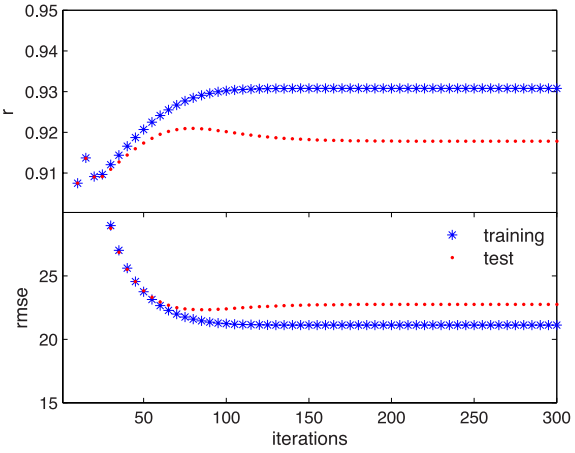

Figure 3. In the top panel the Pearson correlation coefficient between true $f_{\rm NL}$ value and the network estimator $\hat{f}_{\rm NL}$ for case 3 of Table 2 versus the number of iterations. Bottom panel is for the root mean squared error of $f_{\rm NL}$ at each iteration. Asterisks denote training data and dots denote validation data.

Fig. 3 illustrates the training evolution for the regression network with $n_{\rm hid} = 0$ and $n_{\rm data} = 10\,000$. In the top panel we plot the correlation coefficient between the target and the network outputs on the training set and the test set. We see that a divergence occurs around 60 iterations of the MemSys optimizer due to overfitting. The same behaviour is confirmed if the root mean squared error is studied (bottom panel). The network parameters use to construct our final network estimator in (17) and (18) are the ones that give a maximum value of the correlation coefficient and a minimum of the root mean squared error in the validation data set.

It is worth noting that for training the neural network, we need to choose a certain range of $|f_{\rm NL}|$ to generate the required simulations. We find that $[-220\ 220]$ is a safe interval for training the network, without significantly biasing the results for $|f_{\rm NL}|$ up to 30.

6 RESULTS

As a preliminary check, we applied the three estimators to Gaussian full-sky maps without noise, finding very similar results in all cases (see Table 1). In this ideal case, the AMLE should in principle coincide exactly with the AMLED, but because of the lack of correlations among the binned bispectrum components the AMLED seems to be more efficient. This is probably due to numerical uncertainties that arise in the covariance matrix estimation. The neural network is found to be nearly as efficient as the AMLED. It is also worth noticing that the estimators do not present a significant bias.

Table 1. Results for noiseless full-sky maps of set 1. The first column is for the estimator used, second column indicates the expected dispersion for $\ell_{\rm max} = 1024$ and in the last two columns the dispersion and mean value found for 1000 Gaussian maps are shown.

Estimator	$\sigma_{\rm fh}$	$\sigma_{\rm g}$	$\langle f_{\rm NL}\rangle^{\rm Gauss}$
AMLED		9.7	-0.2
AMLE	9.7	10.3	-0.3
NN		9.8	-0.2

An important difference between the estimators is the total number of realizations required to converge, which is directly related to the computational efficiency. In particular, the generation of a simulation at $l_{\rm max} = 1024$ takes around 1 min and the obtention of the binned bispectrum components takes approximately 3 min of CPU time. For the AMLED, a few hundred realizations are sufficient to estimate the variance of the binned bispectrum. For the AMLE estimator, however, it is necessary to estimate the covariance matrix, which requires at least 25 000 Gaussian simulations. For the NNE, a few thousand realizations are required for the training process to converge. Nonetheless, it is worth noting that the number of training realizations required by the NNE does vary with the case being studied. For example, for inpainted maps where neither the linear term is taken into account nor the mean is subtracted (case 1 of Table 2), the NNE needs 10 000 independent simulations to converge.

In applying the three estimators to realistic simulations, based on *WMAP*-7-year data, larger differences are observed in the results; these are summarized in Table 2. We find that the AMLED estimator reaches values close to the expected dispersion if and only if the linear term is subtracted and inpainting is performed. Actually, if the estimator is applied to non-inpainted maps, the dispersion worsens by a $\sim$60 per cent. Of course, in the absence of the linear term, the estimator becomes highly suboptimal, giving errors of 300 per cent. This is not the case for the other two estimators. We notice that the full covariance matrix χ^2 estimator and the neural network give similar results if instead of taking into account the linear term, the mean value of the intermediate maps is subtracted, as is the case for wavelets and needlets (Donzelli et al. 2012). This is observed in both inpainted and non-inpainted maps, comparing cases 2 and 3 and 4 and 5, respectively (see Table 2). Indeed, these estimators appear more robust, since the improvement due to the inpainting is small. In particular, comparing cases 2 and 4, the NNE estimator without inpainting increases the dispersion only by 5 per cent and for the full χ^2 estimator by $\sim$10 per cent, while for the AMLED the results are much worse. Although similar results are found with the AMLE and the NNE estimators, one important difference is the number of simulations required to construct them. As commented before, 25 000 Gaussian realizations were used to estimate the covariance matrix in AMLE. As shown in top panel of Fig. 4, the NNE requires dramatically fewer training realizations and also has the advantage that the average value of the binned bispectrum at $f_{\rm NL} = 1$ does not need to be estimated. In the same figure, bottom panel, we plot the bias found for the $f_{\rm NL}$ estimates for 1000 Gaussian realizations for the three estimators with the number of simulations used. One sees that the AMLE requires more realizations than the other two estimators to produce unbiased results.

All these results indicate that the neural network is a viable short cut to obtaining the necessary weights to construct the AMLE estimator. In Fig. 5 the weights found for the neural network are compared to those of the AMLE. Note that the weights of both estimators are very similar, validating the relation stated in (19). The contribution of the network parameter θ is negligible for all cases.

In terms of computational demand, the most efficient estimators are the NNE and the AMLED, with the number of simulations required at least 10 times smaller than for the AMLE. Note that for the AMLED we have used realizations to estimate the average of the bispectrum at $f_{\rm NL} = 1$, therefore, the final number of realizations employed is similar to the ones used for training the NNE.

Table 2. Comparison of results depending on the estimator. The columns are the characteristic of the estimator, if an inpainting of the simulations is made, if the linear term is added and if the mean was subtracted on the binned intermediate maps. Next columns are $\sigma(f_{NL})$ and $\langle f_{NL}\rangle$ for 1000 Gaussian simulations. Finally, the relative error related to the minimum expected dispersion is shown in the last column.

Case	Inpainting	Linear term	Mean subs.	Estimator	σ_g	$\langle f_{NL}\rangle^{Gauss}$	$(\sigma_{fh}-\sigma_g)/\sigma_{fh}$(percent)
1	Yes	No	No	AMLED	107	3	300
				AMLE	32.7	−1	45
				NN	29.7	−0.3	32
2	Yes	Yes	No	AMLED	**22**	0.7	0.9
				AMLE	**23**	0.7	3.5
				NN	**22**	0.7	0.4
3	Yes	No	Yes	AMLED	31.5	0.7	40
				AMLE	24.0	0.7	6.7
				NN	23.1	0.5	2.7
4	No	Yes	No	AMLED	35.9	−0.3	60
				AMLE	24.3	0.1	9.3
				NN	23.6	0.6	4.8
5	No	No	Yes	AMLED	37.0	1.5	64
				AMLE	24.6	−0.4	8.0
				NN	23.6	0.4	4.8

For all three estimators, the best results are obtained when the map is inpainted and the linear term is subtracted (see case 3 of Table 2, indicated in bold face). For this optimal case, we compute $\langle b_{abc}\rangle^1$ with 1000 simulations of set 2 (Elsner & Wandelt 2009), to compare it with the expected dispersion for a *WMAP*-7-year characteristics, computed as in (10). The neural network is now trained with this set of $a_{\ell m}$. As we have seen, the NNE typically requires 2500 independent training realizations to converge. Since only 1000 are available, we therefore generated 10 000 simulations using the same set of $a_{\ell m}$ rotating them and adding different noise contributions. This procedure was used in Casaponsa et al. (2011b) and was found to be useful when only a small number of realizations is available.

In Table 3 the final results for all of the estimators are shown. The values for *WMAP*-7-year data are without point sources correction, which is given in the last column of the same table. The unresolved point sources contribution to f_{NL} is obtained using the same procedure as in Curto et al. (2009) and Casaponsa et al. (2011a). As expected, by looking at the preliminary results, the tightest constraints are given by the NNE and AMLED estimators. For comparison, the *WMAP*-7-year map f_{NL} estimate with the optimal estimator obtained by Komatsu et al. (2011) is 42, without the point sources correction. Note that the closest value is given by the NNE. The constraints for f_{NL} with the point source contribution taken into account at 95 per cent confidence level are $-3 < f_{NL} < 83$ to be compare with $-2 < f_{NL} < 82$ given by the optimal estimator.

7 CONCLUSIONS

We have trained a regression network with the binned bispectrum components of non-Gaussian realizations in order to obtain constraints on the local non-linear coupling parameter f_{NL}. We have compared the results with those obtained with a maximum likelihood estimator, using either a diagonal or a full covariance matrix. We also studied the effect of the addition of the linear term, mean subtraction and the use of inpainting.

We find that the three estimators become close to optimal if the linear term is subtracted and inpainting is performed. We find that the linear term is absolutely necessary if a diagonal covariance matrix is used. However, its effect is very small if the full covariance matrix or the neural network is used and the mean is subtracted from the binned maps, as found for wavelets and needlets in Donzelli et al. (2012) and Curto, Martínez-González & Barreiro (2012). In that sense, the choice of the estimator depends on the difficulty of computing the linear term. Although the best results for all estimators are obtained when inpainted maps are used, the largest effect of this technique is seen in the AMLED estimator, with the other two being less affected by the presence of a mask. Thus, the most robust tools are the AMLE and the NNE estimators, with the NNE displaying a clear computational advantage, since the covariance matrix does not need to be estimated or inverted; this reduces significantly the number of simulations required. Another advantage of the neural network estimator arises from the fact that for χ^2 minimization the dependence of the covariance matrix on f_{NL} makes a full solution computationally hard, if not unfeasible, for certain problems. Conversely, the NNE bypasses such calculations, thereby simplifying the analysis.

We conclude that the most efficient tools are the neural network regression estimator and the AMLED estimator. The latter would be the choice if a small set of non-Gaussian simulations is available ($\sim$1000), or analytical models are preferred. However, the AMLED depends on a specific pre-processing of the data. Neural networks give almost optimal results, without the use of inpainting, thereby avoiding the need to alter the data.

Finally, the constraints for *WMAP*-7-year data, with the unresolved point sources contribution included, at 95 per cent confidence level would be $-3 < f_{NL} < 83$. These results are compatible with $f_{NL} = 0$, as found in Komatsu et al. (2011), Curto et al. (2011) and Bennett et al. (2012). Note that we have used foreground reduced maps, and the foregrounds have not been marginalized over in this analysis.

We note that neural networks would be a useful method to estimate jointly other forms of non-Gaussianity, such as those where

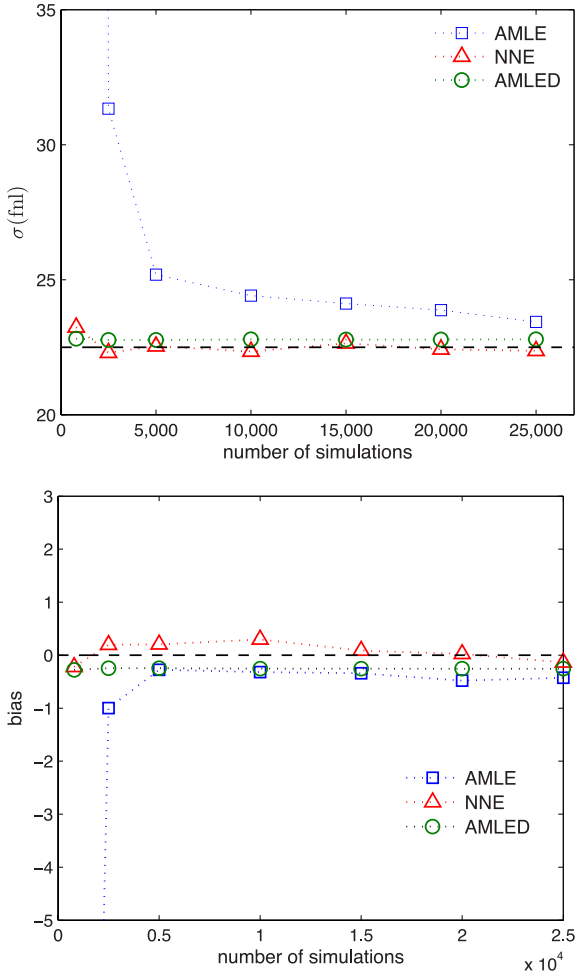

Figure 4. Comparison of the efficiency (top) and bias (bottom) of the three estimators with respect to the number of simulations used to construct the estimator. For reference, the optimal values for the dispersion and bias (dashed black line) are also shown. Note that for the NNE, the simulations are used for the training process, whereas for the AMLE they are employed to estimate the covariance matrix. For the AMLED, they correspond to the number of simulations used to obtain the diagonal elements of the covariance matrix.

the number of outputs were set to a number of different $f_{\rm NL}$ shapes (e.g. local, equilateral, orthogonal), but this is left for future work.

ACKNOWLEDGEMENTS

The authors thank Bartjan van Tent for useful discussions. We acknowledge partial financial support from the Spanish Ministerio de Ciencia e Innovación project AYA2007-68058-C03-02 and AYA2010-21766-C03-01. We also acknowledge partial financial support from the Spanish Minsterio de Economía y Competitividad AYA2010-21766-C03-01 and Consolider- Ingenio 2010 CSD2010-00064 projects. BC thanks the Spanish Ministerio de Ciencia e Innovación for a pre-doctoral fellowship. The authors acknowledge the computer resources, technical expertise and assistance provided by the Spanish Supercomputing Network (RES) node at Universidad de Cantabria. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), part of the High Energy Astrophysics Science Archive Center (HEASARC). HEASARC/LAMBDA is a service of the Astrophysics Science Division at the NASA Goddard Space Flight Center. The HEALPix package was used throughout the data analysis (Górski et al. 2005).

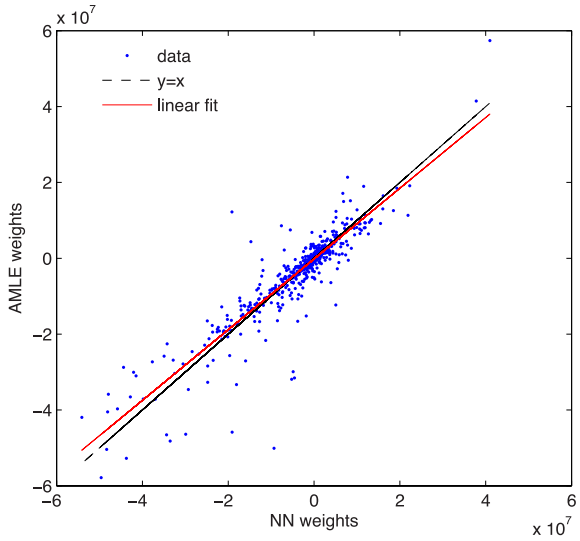

Figure 5. Weights for the AMLE estimator involving the covariance matrix and the model versus the NN weights obtained after the training process. This comparison is made when both estimators have converged presenting a linear fit slope and intercept of $a = 91, b = 2 \times 10^5$.

Table 3. Results for inpainted Gaussian realizations. Model estimated and neural network trained using Elsner & Wandelt simulations (set 2). The columns from left to right are the estimator used, the Fisher σ computed from equation 10, the dispersion and mean value of $\hat{f}_{\rm NL}$ for 1000 Gaussian simulations. Followed by the $f_{\rm NL}$ value found for *WMAP*-7-year data and the contribution expected by the unresolved point sources ($\Delta f_{\rm NL}$).

Estimator	$\sigma_{\rm fh}$	$\sigma_{\rm g}$	$\langle f_{\rm NL}\rangle^{\rm Gauss}$	$f_{\rm NL}^{\rm map}$	$\Delta f_{\rm NL}$
AMLED		21.7	−0.2	33.4	3 ± 2
AMLE	21.3	22.4	−0.1	39.8	3 ± 2
NN		21.4	0.5	44.2	4 ± 2

REFERENCES

Abrial P., Moudden Y., Starck J.-L., Fadili J., Delabrouille J., Nguyen M. K., 2008, Stat. Methodol., 5, 289
Agullo I., Parker L., 2011, Phys. Rev. D, 83, 063526
Arkani-Hamed N., Creminelli P., Mukohyama S., Zaldarriaga M., 2004, J. Cosmol. Astropart. Phys., 4, 1
Auld T., Bridges M., Hobson M. P., Gull S. F., 2007, MNRAS, 376, L11
Auld T., Bridges M., Hobson M. P., 2008, MNRAS, 387, 1575
Babich D., 2005, Phys. Rev. D, 72, 043003
Babich D., Creminelli P., Zaldarriaga M., 2004, J. Cosmol. Astropart. Phys., 8, 9
Baccigalupi C. et al., 2000, MNRAS, 318, 769
Bajkova A. T., 2005, Astron. Rep., 49, 947
Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, MNRAS, 348, 1038
Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, Phys. Rep., 402, 103
Bennett C. L. et al., 2012, preprint (arXiv:1212.5225)

Bertalmio M., Caselles V., Masnou S., Sapiro G., 2000, available at: http://math.univ-lyon1.fr/~masnou/fichiers/publications/survey.pdf
Bucher M., van Tent B., Carvalho C. S., 2010, MNRAS, 407, 2193
Carballo R., González-Serrano J. I., Benn C. R., Jiménez-Luján F., 2008, MNRAS, 391, 369
Casaponsa B., Barreiro R. B., Curto A., Martínez-González E., Vielva P., 2011a, MNRAS, 411, 2019
Casaponsa B., Bridges M., Curto A., Barreiro R. B., Hobson M. P., Martínez-González E., 2011b, MNRAS, 416, 457
Chen X., Firouzjahi H., Namjoo M. H., Sasaki M., 2013, Europhys. Lett., 102, 59001
Creminelli P., Zaldarriaga M., 2004, J. Cosmol. Astropart. Phys., 10, 6
Creminelli P., Nicolis A., Senatore L., Tegmark M., Zaldarriaga M., 2006, J. Cosmol. Astropart. Phys., 5, 4
Curto A., Martínez-González E., Barreiro R. B., 2009, ApJ, 706, 399
Curto A., Martínez-González E., Barreiro R. B., 2011, MNRAS, 412, 1038
Curto A., Martínez-González E., Barreiro R. B., 2012, MNRAS, 426, 1361
Donzelli S., Hansen F. K., Liguori M., Marinucci D., Matarrese S., 2012, ApJ, 755, 19
Elsner F., Wandelt B. D., 2009, ApJS, 184, 264
Fergusson J. R., Shellard E. P. S., 2011, preprint (arXiv:1105.2791)
Fergusson J. R., Liguori M., Shellard E. P. S., 2010, Phys. Rev. D, 82, 023502
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Ganc J., 2011, Phys. Rev. D, 84, 063514
Golden R. M., 1996, Mathematical Methods for Neural Network Analysis and Design. MIT Press, Cambridge, MA
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gull S. F., Skilling J., 1999, Quantified Maximum Entropy: Mem-Sys 5 Users Manual. Maximum Entropy Data Consultants Ltd, Royston
Hobson M. P., Lasenby A. N., 1998, MNRAS, 298, 905
Jaynes E. T., Bretthorst G. L., 2003, Probability Theory and the Logic of Science. Cambridge Univ. Press, Cambridge
Komatsu E., 2002, PhD thesis, Tohoku University, preprint (astro-ph/0206039)
Komatsu E., Spergel D. N., Wandelt B. D., 2005, ApJ, 634, 14
Komatsu E. et al., 2011, ApJS, 192, 18
Leshno M. S. S., 1993, Neural Networks, 6, 861
Mackay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambirdge
Maldacena J., 2003, J. High Energy Phys., 5, 13
Marinucci D., Peccati G., 2011, Random Fields on the Sphere: Applications. Cambridge Univ. Press, Cambridge
Munshi D., Heavens A., 2010, MNRAS, 401, 2406
Nørgaard-Nielsen H. U., 2012, Ap&SS, 340, 161
Oliveira M., Bowen R., McKenna R., Chang Y., 2001, in Proc. Int. Conf. Visualization, Imaging and Image Processing. ACTA Press
Peccati G., Taqqu M. S., 2011, Wiener Chaos: Moments, Cumulants and Diagrams. Springer-Verlag, Milan
Pietrobon D., Cabella P., Balbi A., de Gasperis G., Vittorio N., 2009, MNRAS, 396, 1682
Planck Collaboration, 2013, preprint (arXiv:1303.5084)
Smith K. M., Senatore L., Zaldarriaga M., 2009, J. Cosmol. Astropart. Phys., 9, 6
Starck J.-L., Fadili M. J., Rassat A., 2013, A&A, 550, A15
Storrie-Lombardi M. C., Lahav O., Sodre L., Jr, Storrie-Lombardi L. J., 1992, MNRAS, 259, 8P
Vanzella E. et al., 2004, A&A, 423, 761

This paper has been typeset from a TEX/LATEX file prepared by the author.