

Bispectral analysis of nonlinear interaction, predictability and stochastic modelling with application to ENSO

By CARLOS A. L. PIRES^{1*}, and ABDEL HANNACHI², ¹*Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal;* ²*Department of Meteorology, Stockholm University, Stockholm, Sweden*

(Manuscript Received 26 August 2020; in final form 16 December 2020)

ABSTRACT

Non-Gaussianity and nonlinearity have been shown to be ubiquitous characteristics of El Niño Southern Oscillation (ENSO) with implication on predictability, modelling, and assessment of extremes. These topics are investigated through the analysis of third-order statistics of El Niño 3.4 index in the period 1870–2018, namely bicovariance and bispectrum. Likewise, the spectral decomposition of variance, the bispectrum provides a spectral decomposition of skewness. Positive and negative bispectral contributions identify modes contributing respectively to La Niñas and El Niños, mostly in the period range 2–6 years. The ENSO bispectrum also shows statistically significant features associated with nonlinearity. The analysis of bicovariance reveals a nonlinear correlation between the Boreal Spring and following Winter, coming from an asymmetry of the persistence of El Niño, contributing hence to a reduction of Spring Predictability Barrier. The positive skewness and main features of the ENSO bicovariance and bispectrum are shown to be well reproduced by fitting a bilinear stochastic model. This model shows improved forecasts, with respect to benchmark linear models, especially of the amplitude of extreme El Niños. This study is relevant, particularly in a changing climate, to better characterize and predict ENSO extremes coming from non-Gaussianity and nonlinearity.

Keywords: ENSO spring predictability barrier, bispectrum, El Niño skewness, nonlinear predictability, bilinear models

1. Introduction

Earth's weather and climate vary on a wide range of spatio-temporal scales. Whereas external orbital variations are believed to be the dominant driving force for macroclimate (on millennial time scales), weather, macro-weather (Lovejoy, 2018) and climate variations (on shorter time scales) are mainly the result of complex nonlinear interactions between very many degrees-of-freedom (Sura and Hannachi, 2015) and also due to many climate sub-components with different time scales. The atmospheric (and climate) system is an excellent example of high-dimensional and highly complex dynamical system. One outstanding and ubiquitous feature of the large scale (and low frequency) atmospheric (and climate) variability is non-Gaussianity (Franzke et al., 2007; Proistosescu et al., 2016; Pires and Hannachi, 2017; Hannachi and Iqbal,

2019). For instance, Sura and Sardeshmukh (2008) show that sea surface temperature (SST) has non-Gaussian probability distribution function (PDF) with particular tail extrema. Many processes, e.g. subgrid scales, large-scale teleconnections and nonlinearity can lead to various kinds of uncertainties, which can affect the accuracy of our understanding, such as forecasts. Stochastic modelling can help overcome some of the previous problems and improve accuracies (Berner et al., 2017).

Although non-Gaussianity and nonlinearity can be considered as two distinct aspects of weather, macro-weather and climate (e.g., Sura and Sardeshmukh, 2008; Sura and Hannachi, 2015), observations do suggest that these are two inter-related characteristic features of the system (e.g., Woollings et al., 2010; Hannachi et al., 2017; and references therein). Some systems can exhibit weak (deterministic) nonlinearity, but with strong non-Gaussian statistics, which may be explained by a multiplicative or

*Corresponding author. e-mail: clpires@fc.ul.pt

state-dependent noise as detailed, e.g. in Sura and Sardeshmukh (2008), Sardeshmukh and Sura 2009, see also Hannachi et al. (2017) for a review. In contrast, non-Gaussianity can also result from systems with strong nonlinearity and additive noise, see e.g., Hannachi et al. (2017) for a review and further references. Notwithstanding this link between nonlinearity and non-Gaussianity, it is sometimes helpful to disentangle the contributions of nonlinearity and stochastic noise to the system PDF. There is a general consensus that understanding non-Gaussian statistics of weather and climate is important for a number of reasons, not least for weather and climate prediction, planning and risk assessment. Extreme events, for instance, which are important in planning and risk assessment, depend closely on the structure of the non-Gaussian PDF. Various sources exist, which contribute to the non-Gaussianity. Sura and Hannachi (2015) provide a detailed account of the different sources and mechanisms contributing to the observed non-Gaussianity of the atmospheric large-scale and low-frequency variability. They discussed, in particular, nonlinear regime dynamics, multiplicative noise, cross-frequency coupling, jet stream meandering and nonlinear boundary layer drags. They investigated specifically non-Gaussianity in geopotential heights, jet stream latitude and SST fields, with particular focus on skewness and kurtosis. Skewness and kurtosis are simple measures of non-Gaussianity, in the time domain, that does not directly reflect frequencies. A convenient way to examine nonlinearity and/or non-Gaussianity is to use high (e.g., third and fourth) order statistics in the frequency domain, namely, e.g. bispectrum and trispectrum (e.g. Brillinger and Rosenblatt, 1967; Nikias and Raghuveer, 1987). Spectral analysis of stationary time series is well rooted in the study of time series $x(t)$, $t = 0, 1, 2, \dots$ (supposed to be zero-mean and unit variance for simplicity) from weather and climate and other fields. The duality between the time and spectral domains allows to investigate the distribution of, e.g. variance and skewness, as a function of spectral bins. For example, the duality between the second-order cumulant, or autocovariance function $\gamma_x(\cdot)$ and the power spectrum $\Gamma_x(f)$ reveals that the integral of the latter is precisely the variance of the time series, i.e.,

$$E(x^2) = \gamma_x(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_x(f) df.$$

This means, in particular, that the power spectrum can be viewed as a decomposition of the variance by frequency bins. Likewise, the duality between the bispectrum $\Gamma_{3,x}(f_1, f_2)$ and the third-order moment or bico-variance $\gamma_x(\tau_1, \tau_2) \equiv E[x(t)x(t+\tau_1)x(t+\tau_2)]$ means that the skewness is the integral of the bispectrum,

$$E(x^3) = \gamma_x(0, 0) = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \text{Re}[\Gamma_{3,x}(f_1, f_2)] df_1 df_2.$$

This also implies that the bispectrum can be viewed as a decomposition of the skewness by bins of frequency pairs. This can be used to identify nonlinearly interacting pairs of spectral bins contributing to the skewness which are often attributed to phase locking between components at frequency triplets $f_1, f_2, f_1 + f_2$ producing triadic resonance (Pires and Perdigão, 2015).

Bispectral analysis has been used in signal processing to study nonlinearity detection and bicoherency in a number of different fields including econometry (Ashley et al., 1986; Rusticelli et al., 2008), acoustics (Richardson and Hodgkiss, 1994), and Earth Sciences (Müller, 1987; Biswas et al., 1995; Hocke and Kämpfer, 2008).

Here we focus on the bispectral analysis of El Niño Southern Oscillation (ENSO), the main atmosphere-ocean interannual mode (Neelin et al., 1998; Wang et al., 2016). ENSO aspects, like nonlinearity and complexity (Timmermann, 2003; Frauen et al., 2014; Berner et al., 2018; Bianucci et al., 2018), non-Gaussianity (Burgers and Stephenson, 1999; Chunzai, 2018; Boucharel et al., 2009; Hannachi et al., 2017) and also its modelling (Kondrashov et al., 2005) have been extensively studied, both from a physical and signal-processing perspective. In particular, some ENSO bispectral analysis was performed by Timmermann et al. (2001, 2018), to fit a low-order nonlinear dynamical model of El Niño index and by Schulte et al. (2020) to compute the cross-bi-coherency and synchronization with the Indian Monsoon.

The present study aims to perform a systematic and thorough analysis of third-order statistics, both in the time and spectral domains to infer and improve the understanding of ENSO non-Gaussianity through skewness, nonlinearity (e.g. Hinich, 1982; Cox, 1991), and nonlinear predictability on time scales ranging from seasons to years. We will emphasize, in particular, the role of nonlinear lagged correlation with the intra-annual time scale in overcoming of the Spring predictability barrier of ENSO (Duan and Wei, 2013). The bispectrum of El Niño index and its statistical significance are thus computed, looking for the most relevant wave-triad contributing to the skewness, which are associated with ENSO extremes by a phase synchronization, see e.g. Jajcay et al. (2018). Then, we check how simple stochastic models driven by lagged correlated additive multiplicative (CAM) noise (Monahan, 2020) can produce the main bispectral features of ENSO and El Niño/La-Niña asymmetry (Martinez-Villalobos et al., 2019).

The manuscript starts with data description with a synthesis of exploratory statistics (Section 2). Then the

autocorrelation function and spectrum (Section 3) are shown. In Section 4, we present the tests of nonlinearity and non-Gaussianity based on the bico-variance function. Section 5 details the bispectrum properties and its estimation and statistical significance applied to ENSO. Section 6 develops a minimal bilinear stochastic model fitting data. A summary and conclusion are given in the last section. Most of the technical details and symbols are put in appendices. The details presented in the main text and appendices, some of them conventional, make the paper convenient particularly for didactic purposes and ready to apply to other time-series beyond ENSO.

2. Data

The raw data used are monthly anomalies (with respect to the 1981–2010 period) of El Niño 3.4 sea surface temperature (SST) Index, obtained by area averaged SST in the geographical region 5S–5N by 170–120 W, taken in the 119-year period 1870–2018 and extracted from the periodically updated website https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data (Rayner et al., 2003). Raw data exhibit a positive trend of $0.19^\circ\text{C}/\text{century}$, associated to oceanic global warming. It reveals also a decadal-scale variability (observed in the 50 yr running means) in phase with the Pacific Decadal Oscillation (PDO), also known as a long-lived El Niño pattern (Zhang et al., 1997). After the removal of the linear trend, we get a detrended zero-centred time-series x_{cent} of the index, with a standard deviation $\sigma(x_{cent}) = 0.77^\circ\text{C}$. The skewness is $sk(x_{cent}) = 0.44$, associated to the prevalence of El Niños to La Niñas (Burgers and Stephenson, 1999) and an excess kurtosis ($ekurt$) of 0.46. The monthly dependence of statistics is particularly striking in the skewness and excess kurtosis, ranging from a positively skewed leptokurtic probability distribution function (PDF) in NH winter ($sk = 1.81$, $ekurt = 2.93$ in January), favouring extreme El Niños, up to a platykurtic, nearly unskewed PDF in the NH Spring ($sk = -0.01$, $ekurt = -0.65$ in May). Spring has weaker extreme El Niños compared to Winter, and the extrapolation of the Spring’s El Niño state to the next Winter is rather unskilful i.e. the Spring predictability barrier (Duan and Wei, 2013). The intra-seasonal variability is quite small, compared to the inter-seasonal and interannual variability. Short-range persistence is quite high leading to 1- and 2-month lag-correlations of 0.92 and 0.85. Three-monthly seasonal averages (JFM, AMJ, JAS, OND) are constructed with an overall std $\sigma = 0.74^\circ\text{C}$, $sk = 0.46$ and $ekurt = 0.41$, peaking again during the NH or Boreal Winter’s ($sk = 1.04$, $ekurt = 1.35$ in JFM) (Stuecker et al., 2013; Stein et al., 2014). The statistics are quite similar to those obtained with more commonly used trimesters (e.g. for DJF, we get $sk = 1.07$, $ekurt = 1.68$. The remaining

three seasons are much less skewed with $sk \in [0.25, 0.28]$ and a rather small $ekurt$ except for the platykurtic behaviour in Boreal Spring (AMJ), with $ekurt = -0.51$. All the analysis that follows is performed on the standardized time series, hereafter denoted as $x(t)$, with sample size $N = 596$ trimesters (trm), being shown in Fig. 1.

3. Autocorrelation and spectrum

3.1. Autocorrelation function

We start by evaluating second-order lagged moments of the time series. The autocorrelation function $C_x(\tau) \equiv \frac{1}{N-\tau} \sum_{t=0}^{N-\tau-1} x(t)x(t+\tau)$ is shown in Fig. 2. It has oscillatory behaviour with typical ENSO timescales (3–4 years). In order to test the null hypothesis H_0 of a vanishing autocorrelation, taking into account the serial correlation,

we use an approximate effective sample size $N_{eff} \approx \frac{N[1-C_x(1)]}{[1+C_x(1)]} \approx 66$ (Wilks, 2011). Then, H_0 is rejected if

$$|C_x(\tau)|^2 > \left[q_{\frac{\alpha}{2}}(N_{eff}) \right]^2 / \left\{ \left[q_{\frac{\alpha}{2}}(N_{eff}) \right]^2 + N_{eff} - 2 \right\} \quad (\text{von}$$

Storch and Zwiers, 1999) where $q_{\frac{\alpha}{2}}(N_{eff})$ is the $\frac{\alpha}{2}$ -th quantile of a t -Student PDF with N_{eff} degrees of freedom. From Fig. 2, some local extremes and significant correlations appear also at decadal lags.

3.2. The spectrum and its estimation

A regularly sampled stationary time-series $x(t)$ can be decomposed using discrete Fourier transform (DFT) as: $X_x(f) = \sum_{t=0}^{N-1} x(t) \exp(-2\pi ift)$, for every frequency f in cycles per sampling period and can be reconstructed through the inverse Fourier transform (FT) as $x(t) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} X_x\left(\frac{k}{N}\right) \exp\left(2\pi i \frac{k}{N} t\right)$ (for N even). The sample spectrum or periodogram is the DFT of the autocovariance function:

$$S_x(f) = \frac{1}{N} |X_x(f)|^2 = \sum_{\tau=-(N-1)}^{N-1} C_x(\tau) \exp(-2\pi if\tau) \geq 0 \quad (1)$$

whereas the asymptotic spectrum is given by:

$$\Gamma_x(f) = \lim_{N \rightarrow \infty} E[S_x(f)] \quad (2)$$

where $E(\cdot)$ is the expectation operator over realizations of the stochastic process. According to the Wiener–Khinchine theorem (2) is the FT of the autocovariance function i.e.:

$$\gamma_x(\tau) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_x(f) \exp(2\pi if\tau) df. \quad (3)$$

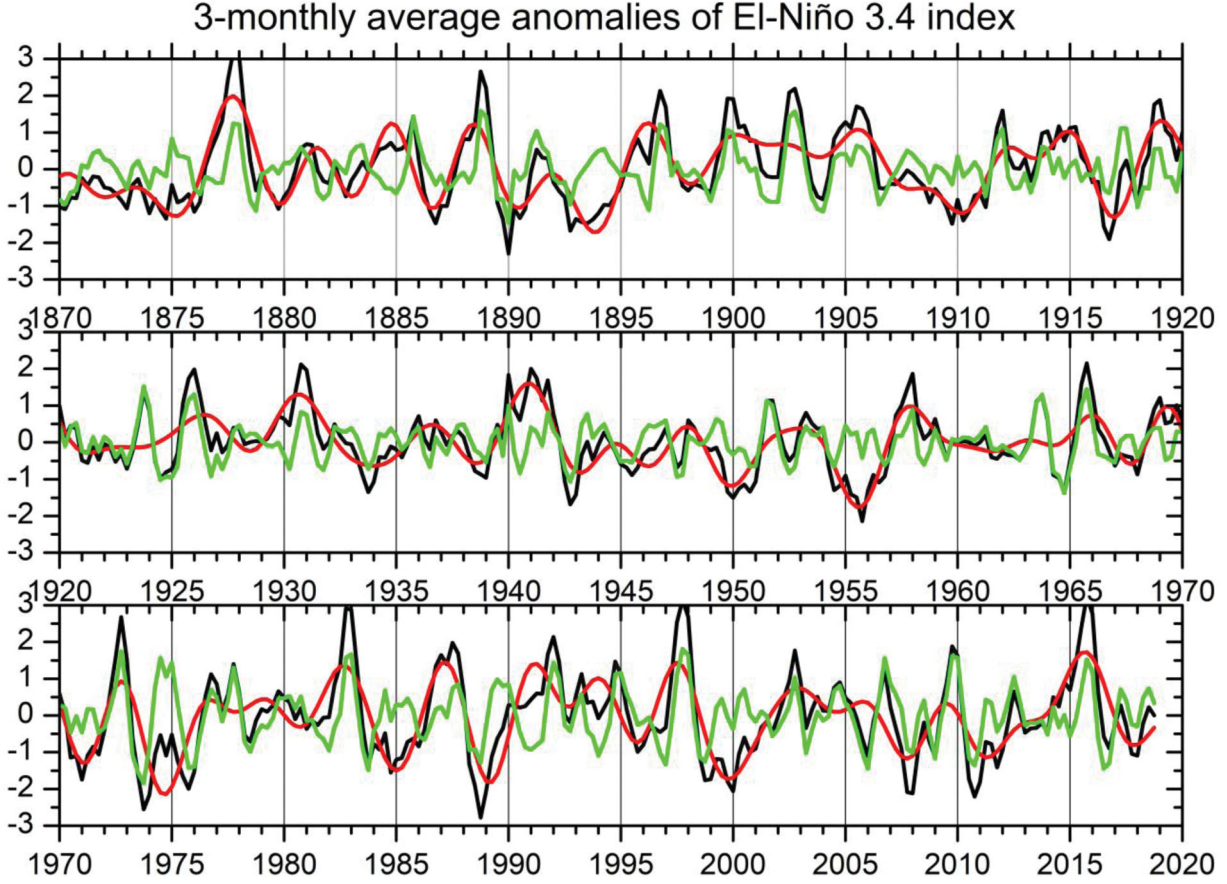


Fig. 1. Detrended and standardized time series in the period 1870–2018, of the three-monthly average anomalies (with respect to the annual cycle) of El Niño 3.4 index (black) and its low-pass (red) and high-pass (green) components with a cutoff frequency of 0.08 cycles per trimester roughly corresponding, respectively, to inter- and intra-triennial timescales variability.

Equation (3) yields, in particular, the variance of the time series

$$\gamma_x(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_x(f) df \approx \frac{1}{N} \left[S_x(0) + 2 \sum_{k=1}^{\frac{N}{2}} S_x\left(\frac{k}{N}\right) \right], \quad (4)$$

where $\Gamma_x(f)df$ in (4) provides the contribution from the spectrum to the time series variance within the frequency interval $[f, f + df]$.

For a stationary process, the DFTs $X_x(f)$, estimated at different frequencies are nearly uncorrelated (von Storch and Zwiers, 1999), and the periodogram (1), is known to be inconsistent, and hence a smoothed estimator is often used:

$$\hat{S}_x(f) = \sum_{\tau=-(N-1)}^{N-1} C_x(\tau) \lambda\left(\frac{\tau}{M}\right) \exp(-2\pi i f \tau) \quad (5)$$

where $\lambda(\tau) \equiv \hat{\lambda}\left(\frac{\tau}{M}\right)$ where $\hat{\lambda}(\cdot)$ is a symmetric standardized positive lag window function (Priestley, 1981), characterized by a standardized bandwidth $b_1 \equiv \int_{-\infty}^{\infty} \hat{\lambda}(u)^2 du$. The

window scale length M is chosen, based on a trade-off between spectral resolution (b_1/M), low values of the variance ($\text{var}[\hat{S}_x(f)] \sim \left[\frac{\Gamma_x(f)^2 M}{Nb_1} \right]$) and bias ($E[\hat{S}_x(f)] - \Gamma_x(f) \sim \frac{-\hat{\lambda}''(0)\Gamma_x''(f)}{4\pi^2 M^2}$) of (5) (Jenkins and Watts, 1968), where double apostrophe represents second derivatives. The 90% confidence interval of $\Gamma_x(f)$ is given by $\left[\frac{\hat{S}_x(f)\nu}{q_{95\%}(\chi_\nu^2)}, \frac{\hat{S}_x(f)\nu}{q_{5\%}(\chi_\nu^2)} \right]$, where $\nu = \frac{2Nb_1}{M}$ is the number of degrees of freedom of the χ_ν^2 .

We have also computed, the maximum entropy estimator of the spectrum by fitting an autoregressive (AR) model of generic order p , via the Yule–Walker equations. The order is chosen by minimizing the Akaike Information Criterion, AIC (Akaike, 1974). The goal of computing the theoretical maximum entropy spectrum is to build a null hypothesis spectrum $\Gamma_{x,0}(f)$. Therefore, in order to objectively check if the spectral peaks of (1) and (5) are truly distinguishable and significantly larger than those of $\Gamma_{x,0}(f)$, at a significance level α , the estimated

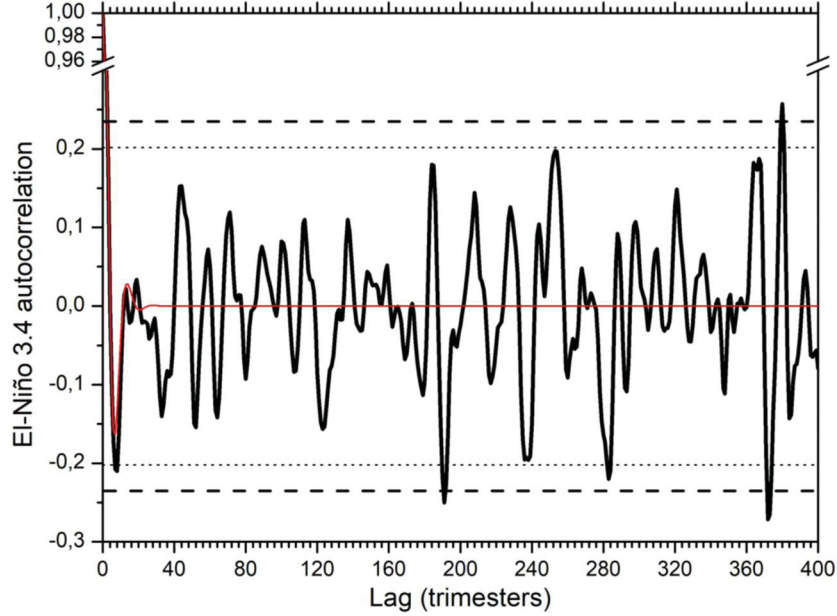


Fig. 2. Autocorrelation function of El Niño 3.4 index (solid black) along with the 95% (dashed black) and 90% (dotted black) confidence interval. The autocorrelation function of the AR(5) model fitting data is also shown (solid red).

spectra (1,5) have to be compared to the threshold $\frac{\Gamma_{x,0}(f)q_{1-\alpha}(\chi^2_\nu)}{\nu}$ (Wilks, 2011) with $\nu = \frac{2Nb_1}{M}$ for (5) and $\nu = 1$ for (1) corresponding to $M = N$, and a standardized bandwidth equal to the Nyquist frequency.

3.3. El Niño index spectrum

3.3.1. Periodogram and smoothed spectrum. The periodogram (1) of the time series is shown in Fig. 3. There is evidence of sharp and well separated peaks near the frequencies 0.021, 0.045, 0.071, 0.087, 0.105, and 0.168 cycles per trimester (cpt), corresponding (in the same order) to periods of 12.2 years, linked to decadal variability (Sun and Yu, 2009; Kravtsov, 2012), 5.62 years (Kim, 2002), and 3.53, 2.87, 2.38 and 1.49 years in the range of the ‘Quasi-quadrennial and quasi-biennial variability’ (Jiang et al., 1995). These peaks are also found by Deser et al. (2010) that are responsible for variations in the ENSO frequency, intensity, propagation, and predictability (e.g., An and Wang, 2000; Fedorov and Philander, 2000; Timmermann, 2003; An and Jin, 2004; Yeh and Kirtman, 2004; Wang, 2018).

For the smoothed spectrum (5), we chose the Bartlett–Priestley window-lag function: $\hat{\lambda}(u) = \frac{3}{(\pi u)^2} \left[\frac{\sin(\pi u)}{\pi u} - \cos(\pi u) \right] (u \neq 0); \hat{\lambda}(0) = 1$ (Priestley, 1981) for which $b_1 = 0.855$ and $\hat{\lambda}''(0) = -1.97$. In order to resolve the main spectral peaks, the bandwidth $\frac{b_1}{M}$ must

be smaller than the minimum difference between consecutive leading frequencies, i.e. $\frac{b_1}{M} < 0.016$ cpt, hence $M > \frac{b_1}{0.016} \sim 52$ trimesters. To further increase spectral resolution, we show in Fig. 3 the estimator (5) using $M = 80$ (20 years), corresponding to $\nu \sim 13$. In this case, the standard deviation of the filtered estimator is about 39% of the true spectrum. The 90% confidence interval for $\Gamma_x(f)$ is $[0.52\hat{S}_x(f), 2.62\hat{S}_x(f)]$ which is quite large, a consequence of the high window length M . The local maxima of the filtered estimator lie close to the high spikes. Moreover, the periodogram lies well within the 95% confidence interval of the true spectrum (see thin black lines of Fig. 3).

3.3.2. Maximum entropy spectrum. An AR(5) model (see the AIC in Table 1) is fitted here:

$$\tilde{x}(t) = 1.18\tilde{x}(t-1) - 0.46\tilde{x}(t-2) - 0.09\tilde{x}(t-3) + 0.17\tilde{x}(t-4) - 0.12\tilde{x}(t-5) + 0.524w(t), \quad (6)$$

where $w(t)$ is a standard Gaussian white noise. The autocorrelation function of the AR(5) model (Fig. 2) closely captures the empirical one for lags up to 12–20 trm but misses low-frequency oscillations. The corresponding (theoretical) maximum entropy spectrum is shown in Fig. 3 showing a spectral bump in the range (3–5 years period) and with no other spectral peaks. Figure 3 also shows the thresholds of significance (at $\alpha = 10\%$) for the periodogram and for the smoothed spectrum.

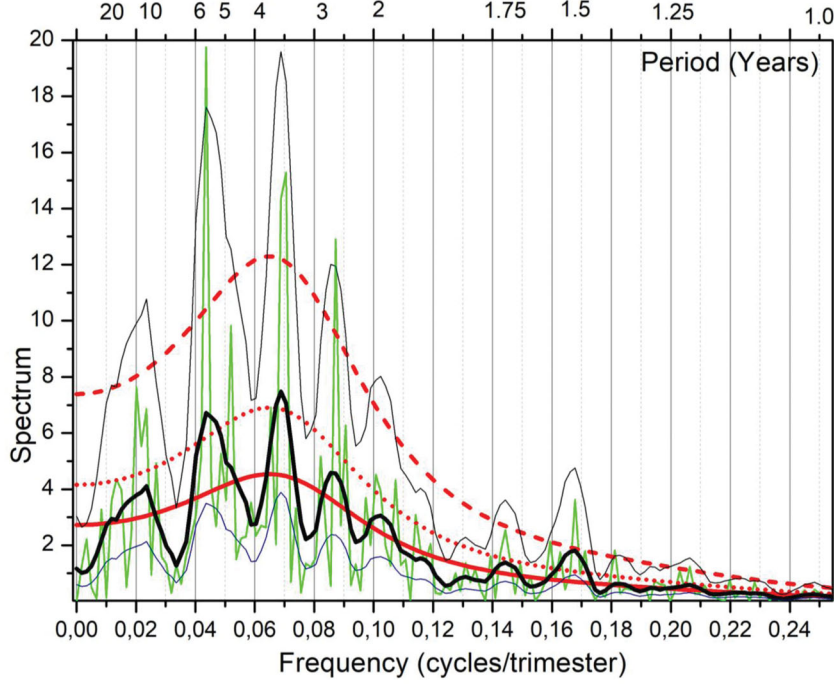


Fig. 3. Empirical periodogram (thin green); Smoothed spectrum (thick black) and corresponding 95% confidence interval (thin black); theoretical AR(5) spectrum (red), and 10% significance level for the periodogram (red dashed) and for the smoothed spectrum (red dotted). Units are $(^{\circ}\text{C})^2/\text{cpt}$. The top axis presents periods in years.

Table 1. Values of the Akaike Information Criterion $AIC(p) = N \ln(2\pi e \sigma_{w,p}^2) + 2p$, where $\sigma_{w,p}^2$ is the residual noise variance of an $AR(p)$ model fitting.

p	0	1	2	3	4	5	6
$AIC(p)$	1691.4	1085.2	939.9	941.9	943.4	936.6	937.6

The peaks with periods 5.62, 3.53, and 1.49 years are significant, both in the periodogram (1) and the smoothed spectrum (5) whereas the peak 2.87 years is significant in the periodogram (1) but only marginally significant in the smoothed spectrum (5).

4. Bicovariance, non-Gaussianity and nonlinearity

4.1. General properties

The bicovariance function $\gamma_x(\tau_1, \tau_2)$ is a generalization of the autocovariance function, given by third-order cumulants between lagged values of $x(t)$, which for a zero mean stationary process writes as $\gamma_x(\tau_1, \tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)]$ and is estimated here, for $\tau_1, \tau_2 \geq 0$ as:

$$C_x(\tau_1, \tau_2) = \frac{1}{N - \max(\tau_1, \tau_2)} \sum_{t=0}^{N - \max(\tau_1, \tau_2) - 1} x(t)x(t+\tau_1)x(t+\tau_2). \quad (7)$$

Stationarity implies time invariance of lagged statistics leading to several identities reflecting the symmetry of the bicovariance (Rao and Gabr, 1984), namely:

$$\begin{aligned} \gamma_x(\tau_1, \tau_2) &= \gamma_x(\tau_2, \tau_1) = \gamma_x(-\tau_1, \tau_2 - \tau_1) = \gamma_x(-\tau_2, \tau_1 - \tau_2) \\ &= \gamma_x(\tau_2 - \tau_1, -\tau_1) = \\ &= \gamma_x(\tau_1 - \tau_2, -\tau_2) \end{aligned} \quad (8)$$

leading to the partition into 6 sectors of the plane (τ_1, τ_2) as illustrated in Fig. 4a. We stress here that the origin of a non-vanishing bicovariance is the non-Gaussianity of the process $x(t)$. In fact, for a Gaussian, (necessarily linear) process, the bicovariance vanishes because cumulants of order 3 (the equivalent to lagged skewness) or higher vanish. For finite time series, Gaussianity is rejected if the absolute value of skewness is large enough.

For a nonvanishing third-order cumulant, the bicorrelation provides a measure of predictability (for lags $\Delta \geq 0$), coming from nonlinear correlation:

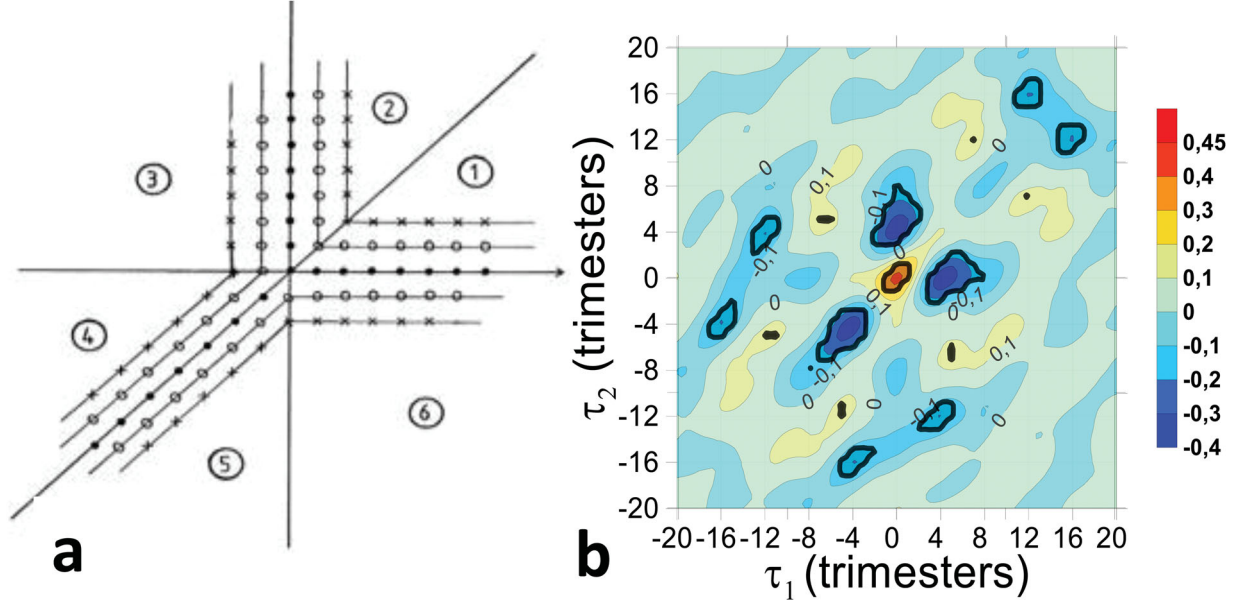


Fig. 4. Symmetries (a) of the bivariate function in the delay plan (adapted from Rao and Gabr, 1984) and bicorrelation of El Niño 3.4 index (b). Thick black contours show the 5% significance level.

$$\text{cor}[x(t + \Delta), x(t)x(t - \tau_1)] = \frac{C_x(\tau_1, \tau_1 + \Delta)}{\sqrt{\text{var}[x(t)x(t + \tau_1)]}}. \quad (9)$$

Local maxima and minima of bicorrelation can thus provide sources of predictability due to non-Gaussianity and/or nonlinearity. However, we must note that part of the nonlinear correlation (9), in particular at $\tau_1 = 0$, is due to skewness, and hence to eliminate that contribution, we must consider the nonlinear component or residual $x_{nl}(t) = x(t)^2 - [sk_x x(t) + 1]$, of the predictor $x(t)^2$ after removing the linear regression with $x(t)$, where sk_x is a regression constant equal to the skewness of x .

Another useful advantage of bicorrelation is the Cox (1991) test of nonlinearity, that we will apply bellow, also based on skewness. More general tests exist, which analyse nonlinear predictability originating from a set of past values (Granger and Anderson, 1978). Under the null hypothesis of linearity, the test $T_{Cox}(\Delta) \equiv \text{cor}[x(t + \Delta), x_{nl}(t)]^2$ must vanish for all lags. Threshold values of T_{Cox} under an inexistant predictability hypothesis are obtained by randomly shuffling $x(t)$ and $x_{nl}(t)$ (10,000 times) and then computing the 95% quantile (denoted as $T_{Cox-95\%}$) of the sorted values of $T_{Cox}(0)$. Nonlinearity of El Niño is thus accepted if $T_{Cox}(\Delta) > T_{Cox-95\%}$ at a 5% significance level.

4.2. Results for El Niño index

4.2.1. Bicorrelation. The bicorrelation (Fig. 4b), exhibits fluctuations of the order of 12–20 trimesters similar to

the autocorrelation (Fig. 2). The maximum value of the bicorrelation coincides with the skewness: $C_x(0, 0) = 0.46$.

In order to test the departure of the bicorrelation from the Gaussian hypothesis, we have generated 10,000, N -sized simulations with the AR(5) and computed uncertainties. The 90% (95%) quantiles of $|C_x(0, 0)|$ are equal to 0.24 (0.29), which are both below the observed skewness 0.46, hence the null hypothesis of Gaussianity is rejected at the 5% significance level. For $\tau_1 \sim \tau_2$ (diagonals of the bivariate graph) or $(\tau_1 \sim 0 \text{ and/or } \tau_2 \sim 0)$, the 95% (90%) quantiles are ~ 0.14 (0.16) and 0.10 (0.12) elsewhere. The rejection regions rejecting of the null hypothesis (at the 5% significance level) appear within thick black contours in Fig. 4b.

From inspection of Fig. 4b, we verify that the deepest bivariate minimum $C_x(0, 4) = -0.40$, is significant at the 5% significance level, corresponding to a nonlinear correlation (Eq. (9)) of $\text{cor}[x(t + 4), x(t)^2] = -0.24$ with $\tau_1 = 0, \Delta = 4 \text{ trm}$. This implies that an extreme El Niño or La Niña (high $x(t)^2$ than average) favours the occurrence of La Niña occurrence four trimesters later (negative $x(t + 4)$), whereas mild conditions favour El Niño. Another local bicorrelation minimum $C_x(8, 8) = -0.15$ (10% significance level) implies that La Niña event favours a strong El Niño or La Niña 2 years later (8 trm).

4.2.2. Linear and nonlinear predictability. As regards El Niño predictability skill, Fig. 5a shows the linear and nonlinear correlations: $\text{cor}[x(t + \Delta), x(t)]$ and

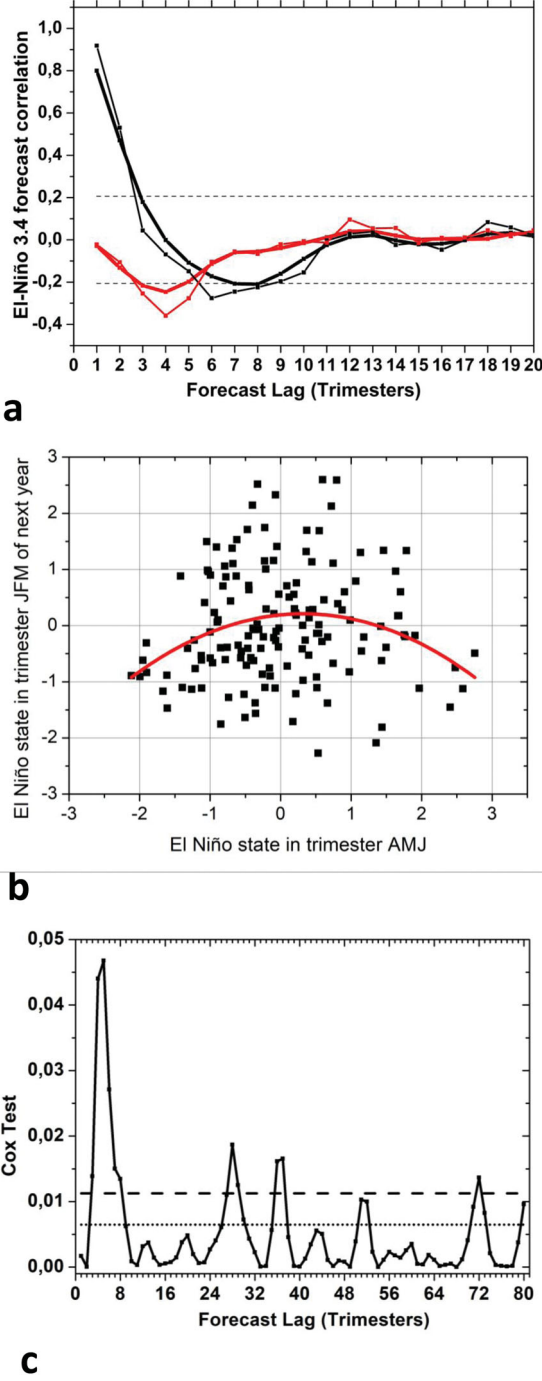


Fig. 5. (a) Linear (thick black) and nonlinear (thick red) correlations ($cor[x(t+\Delta), x(t)]$ and $cor[x(t+\Delta), x_{nl}(t)]$ respectively) of El Niño 3.4 index. The same correlations restricted to forecast trimester JFM (black thin and red thin respectively). Horizontal dashed lines show the 10% significance level interval $\left[-\frac{1.64}{\sqrt{N_{eff}-3}}, \frac{1.64}{\sqrt{N_{eff}-3}}\right]$. (b) AMJ versus following year JFM El Niño and best quadratic fitting. (c) Cox test $T_{Cox}(\Delta)$ (thick) and the 95% (dotted) and 99% (dashed) confidence level threshold of nonrejection of the linearity hypothesis.

$cor[x(t+\Delta), x_{nl}(t)]$, with $x_{nl}(t)$ defined in Section 4.1 and for forecast lags Δ up to 20 trm. The linear correlations of El Niño 3.4 are significant at 10% level for lags $\Delta \leq 3$ trm (Fig. 5a thick black line) whereas the nonlinear correlations show significant values for lags $3 \text{ trm} \leq \Delta \leq 5 \text{ trm}$ (Fig. 5a thick red line). Those correlations are also evaluated for the stronger El Niño season i.e., the JFM trimester ($t+\Delta$ in JFM) (Fig. 5a thin black line for linear and thin red line for nonlinear correlations respectively). We note here the presence of the Spring predictability barrier of El Niño (Duan and Wei, 2013) i.e., the barely weak linear extrapolation (i.e. persistence) of El Niño index from current Spring (AMJ) to the next Winter (JFM), as shown by the small value (~ 0.05) of $cor[x(t+\Delta), x(t) \mid t+\Delta = JFM]$ where $\mid$ means 'conditioned to' and $\Delta = 3$. However, the Spring barrier is reduced if we include the nonlinear term in the forecast. In fact, the nonlinear correlation evaluated at JFM (Fig. 5a thin red line) for forecast lags $\Delta = 3-5$ trm (-0.25 to -0.4) is statistically significant, e.g. a negative value of $cor[x(t+3), x(t)^2 \mid t = AMJ, t+3 = JFM]$ between $x(t)$ in Spring and $x(t+3)$ in next Winter (see Fig. 5b).

This is corroborated by the conditional expectations of the Winter signal conditioned to the previous Spring signal. First, we have $E[x(t+3) \mid x(t) < -1] = -0.38$ (see left sector of Fig. 5b). We argue that strong trade-winds (or persistent eastern wind bursts in the Eastern Pacific), favouring strong La Niña Spring conditions tend to persist over some trimesters eventually reaching the next Winter season (e.g. JFM of years 1989, 1974, and 2011).

On the other hand, we have $E[x(t+3) \mid -1 < x(t) < 1] = 0.12$ (central sector of Fig. 5b) corresponding to (Spring) near climatological conditions in the Eastern Pacific. Here, the tendency is to favour a Winter with El Niño predominance in agreement with Boreal Winter phase-locking. In particular, strong Winter El Niños (e.g. 1983, 1998, and 2016) were preceded by quite mild Spring conditions. Finally, from $E[x(t+3) \mid x(t) > 1] = -0.19$, strong El Niño Spring conditions (associated with westerly winds anomalies) tend to reverse in the next trimesters. This suggests that the nonlinear predictability is a consequence of the asymmetric persistence of El Niño signal and Pacific trade winds in Spring, as a function of their intensity and phase. A possible mechanism for this is seasonal growth rate dependence on the ENSO regime and feedbacks controlling SST (Yishuai et al., 2020), and the seasonal dependence of easterly wind bursts from Spring to Autumn (Fan et al., 2019).

Finally, we compute the nonlinear forecast score $T_{Cox}(\Delta)$ for lags up to 80 trm (Fig. 5c). Clearly, the nonlinearity is especially significant at certain lags (4–8, 28, 36, 52, 72, 80 trm) where $T_{Cox}(\Delta)$ is even larger than the

quantile $T_{\text{CoX}-99\%}$ of nonrejection of the linearity hypothesis. Those lags are related to phase synchronization between Fourier frequencies, namely those with periods τ_1 and $\tau_2 = 2\tau_1$. Lags are generally close to multiples of the half-period of the shorter oscillation, i.e. $\Delta = \frac{n\tau_1}{2}$, $n \in \mathbb{N}$. For instance, from Fig. 3, the Fourier spectral peaks at periods $\tau_1 = \frac{1}{0.087} = 11.5 \text{ trm}$ and $\tau_2 = \frac{1}{0.043} \approx 2\tau_1$ justify the peak of $T_{\text{CoX}}(\Delta)$ at lag $\Delta = \frac{5\tau_1}{2} \approx 29 \text{ trm}$. This frequency relationship is known as quadratic phase coupling (Biswas et al., 1995) and examples in relation to decadal variability of ENSO are given in Timmermann (2003).

5. Bispectrum

5.1. General properties

5.1.1. Bispectrum background. The bicovariance of El Niño time series (Fig. 4b) exhibits certain features and periodicities in the lag-time domain. Therefore, the two-dimensional Fourier transform of the bicovariance, i.e. bispectrum (Brillinger and Rosenblatt, 1967), can provide a dual complementary information about the most relevant spectral interactions contributing to the bicovariance and skewness of the time series making easier the physical interpretation of such interactions.

The bispectrum is the two-dimensional version of polyspectra (Brillinger, 1965), providing relevant information on non-Gaussian processes. The bispectrum $\Gamma_{3,x}(f_1, f_2)$ is given by:

$$\Gamma_{3,x}(f_1, f_2) \equiv \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \gamma_x(\tau_1, \tau_2) \exp[-2\pi i(f_1 \tau_1 + f_2 \tau_2)], \quad (10)$$

where its real and imaginary parts are discrete Fourier transforms (DFT) of the symmetric and antisymmetric parts of the bicovariance, respectively, with the last one vanishing if the underlying stochastic process is reversible (Weiss, 1975).

For the simplest case of a purely random noise $w(t)$, the bicovariance is a spike at the origin i.e. $\gamma_w(\tau_1, \tau_2) = E(w^3) \delta(\tau_1) \delta(\tau_2)$ where $\delta(\cdot)$ is the Kronecker delta, yielding a flat, constant and real bispectrum $\Gamma_{3,w}(f_1, f_2) = E(w^3)$.

5.1.2. Bispectrum and spectral components. The asymptotic bispectrum of a N -sized time series writes in terms of DFTs $X_x(\cdot)$ of the time series, taken at triplets of frequencies (multiples of the minimum frequency $\frac{1}{N}$) (Hinich, 1982) at f_1, f_2 and $f_3 = \left[\left(f_1 + f_2 + \frac{1}{2} \right) \bmod(1) - \frac{1}{2} \right]$ as:

$$\Gamma_{3,x}(f_1, f_2) = \lim_{N \rightarrow \infty} \frac{1}{N} E[X_x(f_1) X_x(f_2) X_x(f_3)^*] \quad (11)$$

where $(*)$ stands for conjugate complex. Eq. (11) shows that the bispectrum comes from the interaction between Fourier components at three frequencies f_1, f_2 and $f_3 = f_1 + f_2 < 1/2$ in the area outside of the aliasing region, i.e. lower than the Nyquist frequency. When $f_1 + f_2 > 1/2$, then $f_3 = 1 - (f_1 + f_2)$ becomes an aliased frequency (Hinich and Wolinsky, 1988).

Equation (11) can still be interpreted in terms of the amplitude and phase of the DFT in polar form, i.e. $X_x(f) \equiv A_x(f) e^{i\Theta_x(f)}$. By denoting $A_x(f_1) A_x(f_2) A_x(f_3) \equiv A_{x,123}$, and $e^{i[\Theta_x(f_1) + \Theta_x(f_2) - \Theta_x(f_3)]} \equiv e^{i\Theta_{x,123}}$, and applying the product expectation decomposition to (11), we get (Kovach et al., 2018):

$$\Gamma_{3,x}(f_1, f_2) = \lim_{N \rightarrow \infty} \frac{1}{N} E[A_{x,123}] E[e^{i\Theta_{x,123}}] + \lim_{N \rightarrow \infty} \frac{1}{N} \text{cov}(A_{x,123}, e^{i\Theta_{x,123}}), \quad (12)$$

where the first r.h.s. term of (12) depends on phase synchronization of the three Fourier components and the second term is a covariance between amplitudes and phases, vanishing for linear processes. In fact, the Volterra representation of a linear process writes as a convolution: $x(t) = \sum_k \alpha(k) w(t-k)$, where $w(t)$ is a purely random noise. Its DFT is $X_x(f) = X_\alpha(f) X_w(f) = X_\alpha(f) A_w(f) e^{i\Theta_w(f)}$ where $X_\alpha(f), X_w(f)$ are DFTs, respectively of the sequence $\alpha(\cdot)$ and of the noise. The independence between $A_w(f)$ and $\Theta_w(f)$ yields $\Gamma_{3,x}(f_1, f_2) = X_\alpha(f_1) X_\alpha(f_2) X_\alpha(f_3)^* E(w^3)$, restricted to the synchronization term (Nikias and Raghuveer, 1987), where $E(w^3) = \lim_{N \rightarrow \infty} \frac{1}{N} E[A_{w,123}] E[e^{i\Theta_{w,123}}]$, showing hence the intrinsic nonlinear origin of the covariance term of (12).

5.1.3. Properties of the bispectrum. Like the spectrum, the bispectrum of a real signal satisfies $\Gamma_{3,x}(-f_1, -f_2) = \Gamma_{3,x}(f_1, f_2)^*$, and (11) leads to 5 identities of symmetry, namely:

$$\begin{aligned} \Gamma_{3,x}(f_1, f_2) &= \Gamma_{3,x}(f_2, f_1) = \Gamma_{3,x}(f_1, -f_1 - f_2) = \Gamma_{3,x}(-f_1 - f_2, f_2) \\ &= \Gamma_{3,x}(-f_1 - f_2, f_1) = \Gamma_{3,x}(f_2, -f_1 - f_2). \end{aligned} \quad (13)$$

This allows for a partition of the bispectrum domain (BD) into 12 polygonal regions in which the bispectrum can be reproduced from the Principal Domain (PD): the triangle of vertices (0,0), (0,1/2) and (1/3,1/3) (shown by triangle 1 in Fig. 6a). The periodicities are evident in the theoretical bispectrum in Fig. 6b (real part) and Fig. 6c (imaginary part) of a non-Gaussian AR(5) model of El Niño (presented later in Section 5.3.1).

The reconstruction of bicovariance is obtained through the inverse FT of (10):

$$\gamma_x(\tau_1, \tau_2) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \Gamma_{3,x}(f_1, f_2) \exp[2\pi i(f_1 \tau_1 + f_2 \tau_2)] df_1 df_2, \quad (14)$$

which leads to the skewness decomposition in terms of positive or negative contributions given by the real parts along the PD:

$$E(x^3) = \gamma_x(0, 0) = 12 \iint_{PD} \text{Re}[\Gamma_{3,x}(f_1, f_2)] df_1 df_2. \quad (15)$$

It is important to note here that, like the power spectrum (4), (15) implies that the element of the bispectrum $\text{Re}[\Gamma_{3,x}(f_1, f_2)] df_1 df_2$ provides the contribution to the skewness $E(x^3)$ from the bi-spectral bin $[f_1, f_1 + df_1] \times [f_2, f_2 + df_2]$.

Finally, since the square correlation is a predictability measure coming from third-order moments (see Eq. (9)), its overall sum can be distributed over the bispectral domain through the Parseval relationship:

$$\begin{aligned} \sum_{\tau_1} \sum_{\tau_2} [\gamma_x(\tau_1, \tau_2)]^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} |\Gamma_{3,x}(f_1, f_2)|^2 df_1 df_2 \\ &= 12 \iint_{PD} |\Gamma_{3,x}(f_1, f_2)|^2 df_1 df_2. \end{aligned} \quad (16)$$

5.2. Bispectrum estimation

The estimation of bispectrum has been addressed by many authors (e.g. Brillinger and Rosenblatt, 1967; Raghuvver and Nikias, 1986; Nikias and Raghuvver, 1987). The empirical bispectrum or biperiodogram of a finite sample of length N is the two-dimensional DFT of the sampled bicovariance, which can also be expressed in terms of DFTs of the signal (see Section 3.1):

$$\begin{aligned} S_{3,x}(f_1, f_2) &\equiv \sum_{\tau_1=-(N-1)}^{N-1} \sum_{\tau_2=-(N-1)}^N C_x(\tau_1, \tau_2) \\ &\times \exp[-2\pi i(f_1 \tau_1 + f_2 \tau_2)] = \frac{1}{N} X_x(f_1) X_x(f_2) X_x(f_3)^* \end{aligned} \quad (17)$$

where $f_3 = (f_1 + f_2 + \frac{1}{2}) \bmod(1) - \frac{1}{2}$. The inverse FT of Eq. (17) yields the empirical bicovariance:

$$\begin{aligned} C_x(\tau_1, \tau_2) &= \frac{1}{N^2} \sum_{k_1=-\frac{N}{2}}^{\frac{N}{2}} \sum_{k_2=-\frac{N}{2}}^{\frac{N}{2}} S_{3,x}\left(\frac{k_1}{N}, \frac{k_2}{N}\right) \\ &\times \exp\left[2\pi i\left(\frac{k_1}{N} \tau_1 + \frac{k_2}{N} \tau_2\right)\right] \end{aligned} \quad (18)$$

Like the periodogram (1), the estimator (17) is not consistent, which can be overcome by (i) smoothing the sample bispectrum (Hinich, 1982) or dividing the sample into

pieces and then averaging and smoothing bispectra (Lii and Rosenblatt, 1982); (ii) using multi-tapers (Birkelund and Hanssen, 1999, 2000) or (iii) smoothing the bicovariance function (indirect-method) (Rao and Gabr, 1984), which we use here.

The smoothed bispectrum is:

$$\begin{aligned} \hat{S}_{3,x}(f_1, f_2) &= \sum_{\tau_1=-N}^N \sum_{\tau_2=-N}^N C_x(\tau_1, \tau_2) \Lambda(\tau_1, \tau_2) \\ &\times \exp[-2\pi i(f_1 \tau_1 + f_2 \tau_2)] \end{aligned} \quad (19)$$

where the 2D-lag window $\Lambda(\cdot)$ satisfies similar properties of symmetry as the bicovariance (8) and is taken to be $\Lambda(\tau_1, \tau_2) = \hat{\lambda}(\frac{\tau_1}{M_2}) \hat{\lambda}(\frac{\tau_2}{M_2}) \hat{\lambda}(\frac{\tau_1}{M_2} - \frac{\tau_2}{M_2})$ where $\hat{\lambda}(\cdot)$ is the lag window function and M_2 is the window length. The equivalent to bandwidth is given by $\frac{b_2}{M_2^2}$ where $b_2 = 1 / \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{\lambda}(u) \hat{\lambda}(v) \hat{\lambda}(u-v)]^2 du dv = 1.19$ (Appendix A).

5.3. Bispectrum estimation of El Niño index

5.3.1. Null hypothesis bispectrum. To reproduce the observed skewness, we can construct an AR process driven by a non-Gaussian noise as a null hypothesis H_0 .

The model we wish to fit is like model (6) $\tilde{x}(t) = \sum_{k=1}^p a(k) \tilde{x}(t-k) + \sigma_w w(t)$, but with a non-Gaussian $w(t)$ white noise, $\sigma_w^2 = \frac{E(x^2)}{\int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^{-2} df}$ and $A(f) = 1 -$

$\sum_{k=1}^p a(k) e^{-2\pi i k f}$. The bispectrum of such a linear process (Nikias and Raghuvver, 1987) is:

$$\Gamma_{3,\tilde{x}}(f_1, f_2) = \frac{\sigma_w^3 E(w^3)}{A(f_1) A(f_2) A(f_1 + f_2)^*}, \quad (20)$$

with:

$$E(w^3) = \frac{E(x^3)}{\sigma_w^3 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} [A(f_1) A(f_2) A(f_3)^*]^{-1} df_1 df_2} \quad (21)$$

By using the coefficients of the AR(5) model (6), we get the approximations $\sigma_w^2 = 0.275$ and $E(w^3) = 0.895$, hence $\sigma_w^3 E(w^3) = 0.129$. This null H_0 is hereafter designated NGAR(5).

Figure 6 shows (Fig. 6b) and imaginary (Fig. 6c) parts of the bispectrum (20) of the NGAR(5) model over the global bifrequency domain. The real part is mostly positive whereas the imaginary part is formed by dipolar structures. Both parts reflect the symmetry shown in (13) and exhibit a positive band of maximum absolute values in the region near $f_1 + f_2 \sim 0.09$ cpt and $f_1, f_2 \in [0.02, 0.08]$ cpt, coinciding with the frequency range where the power of the AR(5) model (6) exceeds 3 (°C)²/cpt as seen in Fig. 3.

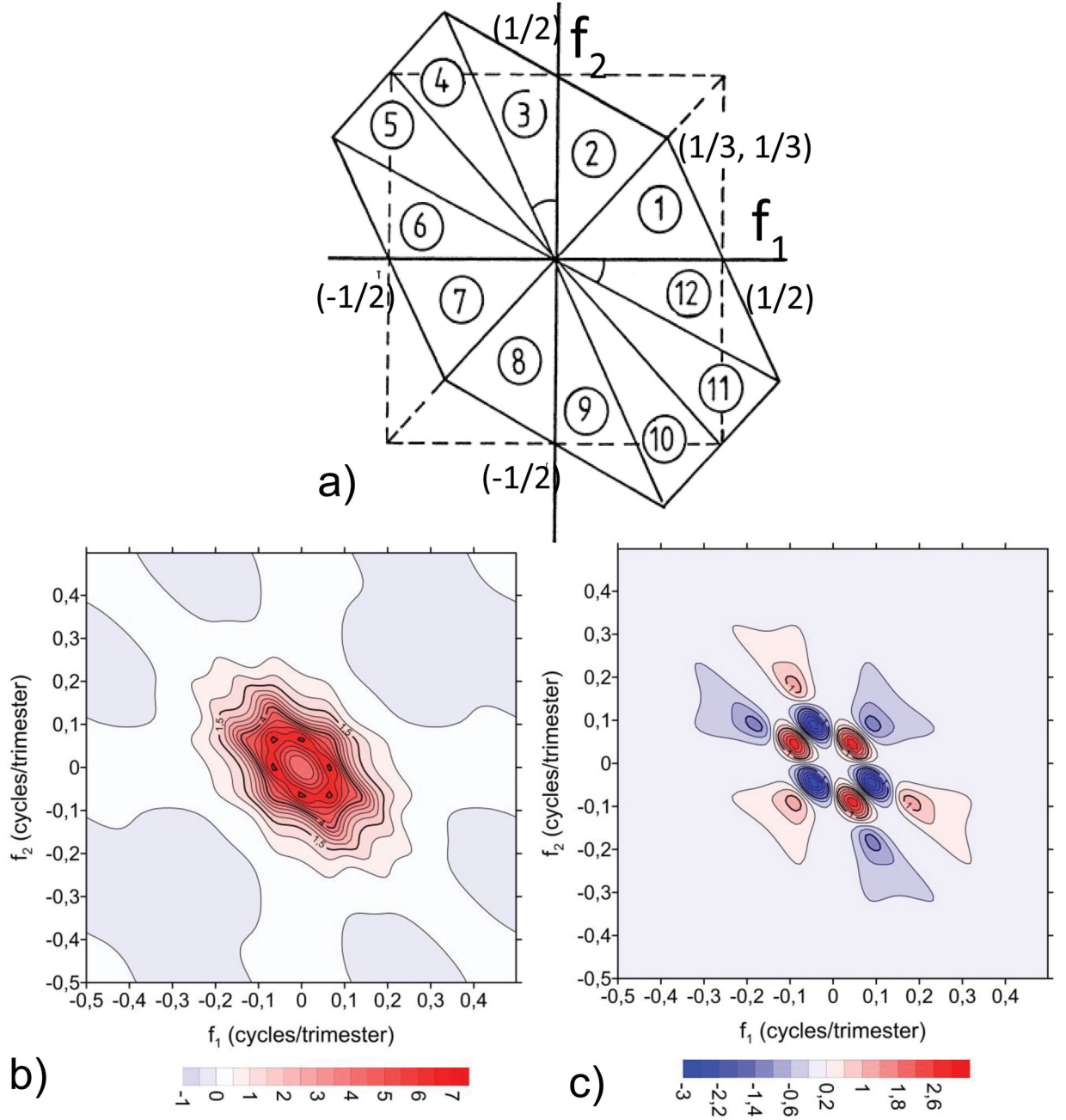


Fig. 6. Principal domain (triangle 1) in the spectral plane and its symmetric replicas (a) (adapted from Rao and Gabr (1984)). Real (b) and Imaginary (c) parts of the bispectrum of the NGAR(5) model. Note the symmetries associated with the 12 sectors shown in panel a).

5.3.2. Empirical smoothed bispectrum. To estimate the empirical bispectrum using the smoothed estimator (19), we start by identifying the ideal window function M_2 , according to (A4) in Appendix A. The values of the bispectral fluctuations $\sigma_{\hat{s}_{3,x}}$ and the average confidence interval half-size (cihs) of the bispectrum (given by the square root of the l.h.s. of (A4)) are given in Table 2 for several values of M_2 . In order to separate peaks, the condition

$\sigma_{\hat{s}_{3,x}} > cihs$ must be satisfied. As expected, smaller bandwidths (larger M_2) lead to larger bispectral fluctuations and larger bispectrum estimation errors via *cish*. From Table 2, the largest M_2 around which $\sigma_{\hat{s}_{3,x}} > cihs$ is $M_2 \sim 30$, which is used below.

Figure 7 shows the real (Fig. 7a) and imaginary (Fig. 7b) parts of the smoothed bispectrum of the 3.4 El Niño index for the most relevant part of the first quadrant.

Both parts show several peaks. In particular, both present high absolute values for $f_1, f_2 \in [0.04, 0.07]$ *cpt*, as for null NGAR(5) model (Fig. 6) but with much larger amplitude. As expected, the average diameter of peaks is of the order of $\frac{\sqrt{b_2}}{M_2} = 0.036$ *cpt*. The integral of $Re(\hat{S}_{3,x})$ (Fig. 7a) is the estimated skewness as Eq. (15). According to Eq. (11), the superposition of Fourier components along the triplet of frequencies $f_1, f_2, f_3 = f_1 + f_2$ where $Re()$ is positive (negative), will mostly generate extreme

Table 2. Average fluctuations $\sigma_{\hat{s}_{3,x}}$ of the bispectrum and the average confidence interval half-size (CIHS) for several values of the window length M_2 (Eq. (A4)).

M_2	10	20	30	40	80
$\sigma_{\hat{s}_{3,x}}$	1.16	1.67	2.00	2.27	3.69
CIHS	0.70	1.37	2.09	2.79	5.59

positive (negative) values, i.e. El Niño (La Niña) events. The integral of positive and negative values of $Re(\hat{S}_{3,x})$ over the frequency domain is 0.54 and -0.08 respectively (adding up to the observed skewness 0.46). The local maxima of the real part, mostly contributing to El Niño extremes, lie near the frequency triplet ($f_1 = f_2 = 0.05, f_3 = 0.1$ *cpt*) and the band ($f_1 + f_2 = f_3 = 0.165$ *cpt*), corresponding to local maxima of the power spectrum (Fig. 3), (e.g. quadratic phase synchronization Jajcay et al., 2018). On the other hand, the local minima of the real part, mostly contributing to La Niña extremes, lie near the frequency triplets ($f_1 = f_2 = 0.018, f_3 = 0.036$ *cpt*) and ($f_1 = 0.05, f_2 = 0.018, f_3 = 0.063$ *cpt*), which are again close to relative power maxima (see Fig. 3).

Figure 7c shows the squared bispectrum amplitude, providing the bispectral contribution to the predictability through (16), which agree quite well with Timmermann (2003). Its maxima occur near the local extremes of the

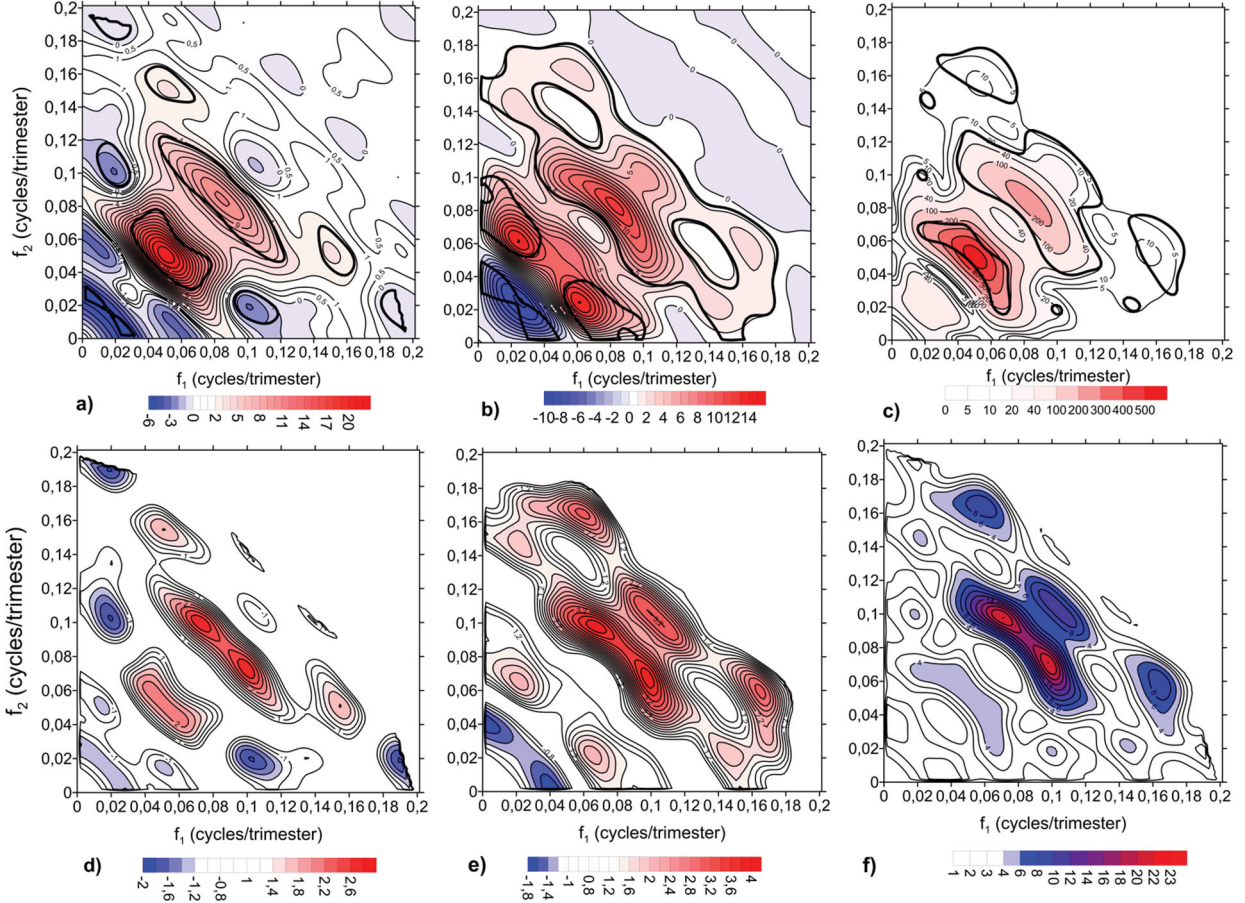


Fig. 7. Real (a) and Imaginary (b) part of the empirical bispectrum using a window lag ($M_2 = 30$). Squared amplitude of the smoothed bispectrum (c). In figures a–c, significant regions at 20% significance level (or lower) appear within thick contours. Real (d) and imaginary (e) parts of the standardized bispectrum deviation $T_{x,\hat{x}}(f_1, f_2)$ of El Niño index, and corresponding sum of the squared real and imaginary parts (f). Bifrequencies for which the null hypothesis H_0 is rejected at 20% significance levels (or lower) are color-shaded ($|T_{x,\hat{x}}| > 1.3$ for each part). Values are restricted to the most significant part of bispectrum.

real or imaginary parts of $\hat{S}_{3,x}$ (Fig. 7a,b). The most relevant region for nonlinear predictability occurs for frequencies satisfying $f_1 + f_2 \in [0.07, 0.1]$ cpt. Another maximum is observed for $f_1 + f_2 \in [0.16, 0.18]$ cpt, producing oscillations with periods of 5–6 trimesters, and suggests possible source of the high nonlinear predictability for lags 5–6 trm, as diagnosed by Cox test in Fig. 5c.

5.3.3. Bispectrum and bicovariance near the origin. The bispectrum is relevant for the bicovariance behaviour near the origin. In fact, $\gamma_x(\tau_1, \tau_2)$ can be approximated by a Taylor expansion:

$$\begin{aligned} \gamma_x(\tau_1, \tau_2) = & \gamma_x(0, 0) + \sum_{p=1,2} \frac{\partial \gamma_x}{\partial \tau_p} \tau_p + \frac{1}{2} \sum_{p,q=1,2} \frac{\partial^2 \gamma_x}{\partial \tau_p \partial \tau_q} \tau_p \tau_q \\ & + \frac{1}{6} \sum_{p,q,r=1,2} \frac{\partial^3 \gamma_x}{\partial \tau_p \partial \tau_q \partial \tau_r} \tau_p \tau_q \tau_r + O(\tau_1^a \tau_2^b); \quad a + b = 4 \end{aligned} \quad (22)$$

where derivatives are computed at $\tau_1 = \tau_2 = 0$. Using Eq. (14), we get:

$$\frac{\partial \gamma_x}{\partial \tau_p} = - \iint_{BD} \text{Im}[\Gamma_{3,x}(f_1, f_2)] 2\pi f_p df_1 df_2 \quad (23)$$

$$\frac{\partial^2 \gamma_x}{\partial \tau_p \partial \tau_q} = - \iint_{BD} \text{Re}[\Gamma_{3,x}(f_1, f_2)] (2\pi)^2 f_p f_q df_1 df_2 \quad (24)$$

$$\frac{\partial^3 \gamma_x}{\partial \tau_p \partial \tau_q \partial \tau_r} = \iint_{BD} \text{Im}[\Gamma_{3,x}(f_1, f_2)] (2\pi)^3 f_p f_q f_r df_1 df_2. \quad (25)$$

where bicovariance symmetries lead to symmetries at the origin: $\frac{\partial \gamma_x}{\partial \tau_p} = 0$, i.e. a local bicovariance extreme

(see Fig. 4b) and $\frac{\partial^2 \gamma_x}{\partial \tau_1 \partial \tau_1} = \frac{\partial^2 \gamma_x}{\partial \tau_2 \partial \tau_2}$, $\frac{\partial^3 \gamma_x}{\partial \tau_1 \partial \tau_1 \partial \tau_1} = \frac{\partial^3 \gamma_x}{\partial \tau_2 \partial \tau_2 \partial \tau_2}$, $\frac{\partial^3 \gamma_x}{\partial \tau_1 \partial \tau_1 \partial \tau_2} = \frac{\partial^3 \gamma_x}{\partial \tau_2 \partial \tau_2 \partial \tau_1}$.

The partial derivatives at the origin, estimated with the smoothed El Niño bispectrum yield $\frac{\partial^2 C_x}{\partial \tau_p \partial \tau_p} = -0.264$ ($p = 1, 2$), which explains most of the symmetric decrease of $C_x(\tau_1, \tau_2)$ near the origin (see Fig. 4b). The term $\frac{\partial^3 C_x}{\partial \tau_p \partial \tau_p \partial \tau_p} = -0.114$, however explains the asymmetry of that decrease, which is stronger for positive than negative lags yielding the bicovariance minimum $C_x(0, 4) = -0.4$.

5.3.4. Bispectrum from frequency-band partitions. A coarse-grained description of the bispectrum can be achieved by classifying the triplets $f_1, f_2, f_3 = f_1 + f_2$ into sets of frequencies forming a partition of $[0, 1/2]$. Each triplet is then characterized by the number of frequencies in each set. We consider the simple partition of the frequency interval using a cutoff frequency $0 < f_{cut} < 1/2$, separating low (S) and high (F) frequencies, with the corresponding decomposition $x(t) = s(t) + f(t)$ (see

Fig. 1). Figure 8a shows the 4 obtained subdomains namely, SSS, SSF, SFF and FFF, yielding an expansion into 4 terms of third-order statistics, e.g.,

$$\begin{aligned} E(x^3) = & E(s^3) + 3E(s^2 f) + 3E(s f^2) + E(f^3) \\ = & SSS + SSF + SFF + FFF, \end{aligned} \quad (26)$$

Figure 8b shows the terms in the r.h.s. of Eq. (26) as a function of f_{cut} . A reasonable criterion of discrimination among the different components is to choose f_{cut} that maximizes the sum $|SSS| + |SSF| + |SFF| + |FFF|$, which takes place at $f_{cut} = 0.082$ cpt (3.04 years). This yields inter- (slow S) and intra-triennial (fast F) variations with respective 82% and 38% explained variance, with a well-marked scale separation lying at a local minimum of the smoothed power spectrum (see Fig. 3), and a minimum value of $E(s^3)$ (-0.066).

To see the impact of spectral decomposition on the different terms of skewness, we compare in Table 3 the terms of Eq. (26), derived directly from the time series to those obtained from the partial integrals of the smoothed bispectrum (Fig. 7a). Table 3 shows that the values are quite close, except for the negative value of $E(s^3)$. The underestimation obtained from the smoothed bispectrum is suggested to be due to the weak resolution of low frequencies.

Extremes are classified according to the dominant term (SSS, SSF, SFF or FFF). From Table 3, extreme events of La Niña must be explained by the slow component $s(t)$ (e.g. 1887, 1917, 1956, 2000) or by phase synchronization of $s(t)$ and $f(t)$ (mostly of SFF type, e.g. 1973, 1988, 2008, and 2011) (Fig. 1). On the other hand, extreme events of El Niño, are mainly due to slow-fast component interactions, namely of SSF type (e.g. 1877, 1918, 1930, 1958, 2015) and SFF type (e.g. 1926, 1951, 1965, 1972, 1982–83, 1992, 1997, 2002) or from fast components only (FFF type) (e.g. 1923, 1977, 2006, 2009). However, some El Niño events (e.g. 1905, 1940, 1986–87), have occurred due to long persistence of the slow component (SSS type) (Fig. 1). Note also that there are few cases of phase polarity between fast and slow components (e.g. 1974) that lead to weak El Niño index.

In order to determine temporal changes of the bispectrum, we assess the third moment and its decomposition, Eq. (26), both in the full period (FULL) and along the three half-centuries: 1870–1919 (HC1), 1920–1969 (HC2) and 1970–2018 (HC3) (Fig. 1). Moreover, we evaluate the above statistics during El Niños ($x(t) > 0$) and La Niñas ($x(t) < 0$) to examine the variability of extremes and corresponding spectral contributions. For a given term in Eq. (26), for instance, SSS, its average $E(SSS)$ decomposes as: $E(SSS) = E(SSS)_+ + E(SSS)_-$, where

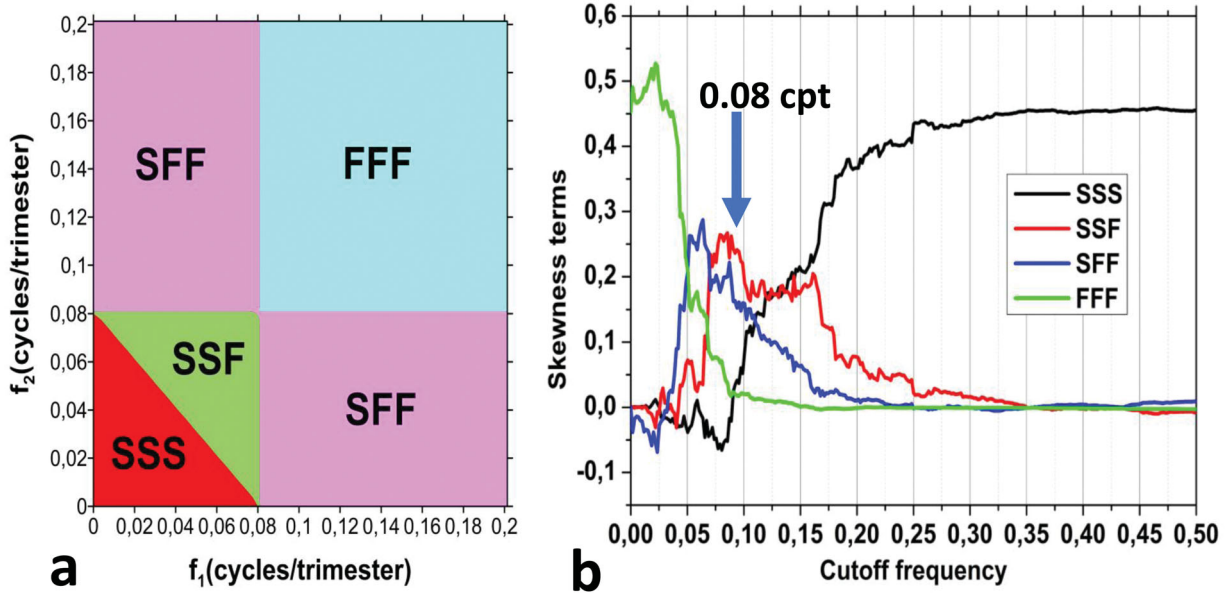


Fig. 8. (a) Subdomains SSS, SSF, SFF and FFF obtained from a frequency partition using a cut-off frequency $f_{cut} = 0.08$ cpt (~ 3 years). (b) Contributing terms to El Niño 3.4 index skewness (Eq. (26)).

Table 3. Contributing terms to skewness, Eq. (26), obtained from time series and from partial integrals of the smoothed bispectrum: total, negative and positive contributions. The most relevant integral positive (negative) contributions are marked in bold, representing the most relevant types of interaction contributing to extreme El Niño (La Niña) events.

	SSS	SSF	SFF	FFF
Time series	-0.066	0.263	0.185	0.071
$Re(\hat{S}_{3,x})$	-0.023	0.216	0.212	0.052
$Re(\hat{S}_{3,x}) < 0$	-0.037	0.000	-0.031	-0.012
$Re(\hat{S}_{3,x}) > 0$	0.014	0.216	0.243	0.063

$E(SSS)_+ = E(SSS \mid x > 0)prob(x > 0)$ and $E(SSS)_- = E(SSS \mid x \leq 0)prob(x \leq 0)$, giving the contributions to $E(SSS)$ during El Niños and La Niñas, respectively. Table 4 summarizes the results.

First, the most recent half century (HC3) has on average, the most extreme episodes of La Niña and El Niño, as observed from the high absolute values of sk_- and sk_+ . The amplification of La Niñas comes mostly from a clear increase of self-slow interaction $SSS = -0.06$ and cross interaction $SFF = -0.4$ whereas amplification of El Niños comes from the enhancement of the SFF term (0.6), as compared to the previous two half centuries. This suggests changes and decadal variability of the ENSO skewness, its bivarariance and bispectrum (Wu and Hsieh, 2003), associated to changes in the preferential Fourier phase couplings (Schulte et al., 2019). This was accompanied by an ENSO

regime shift, near 1970 towards more nonlinearity (An and Wang, 2000; An and Jin, 2004; An, 2009).

5.3.5. Statistical significance of the empirical bispectrum.

It is important to check the acceptance or rejection of the null bispectrum H_0 of NGAR(5) (Fig. 6b,c). The spectral method used here is based on a variation of the Hinich (1982) test of linearity. We anticipate that both the local (A6) and the integrated (A7) spectral-based tests of nonlinearity in the bi-frequency domain reject the null H_0 in consistency with the nonlinearity Cox test in the time domain (see Section 4.2.2) and hence other sources of non-Gaussianity (e.g. deterministic nonlinearity and multiplicative noise) shall be necessary. Furthermore, as shown in Appendix A the asymptotic bias (A2), variance (A3) as well as the asymptotic Gaussian PDF (Rao and Gabr, 1984) of the smoothed estimator are not good approximations because of the small number of degrees freedom ($N_{eff} = 66$) of the time series and hence cannot be used to test the null H_0 . This could be alleviated by using, e.g. a very long run of a climate model simulation. We use a Monte-Carlo strategy by computing the statistics of bispectrum by generating 10,000 surrogates of the NGAR(5) model forced by a noise prescribed by its first three moments. In order to easily obtain noise realizations, we consider noises produced by polynomial independent standard Gaussian noises, by relating the coefficients of monomial expectations to the imposed noise moments. For instance, the first trial noise: $\sigma_w w(t) = aw_1(t) + b[w_1^2(t) - 1]$ where w_1 is a standard

Table 4. Expected values (yellow) of the third moment of the normalized El Niño 3.4 index (denoted as sk) and its contributions SSS , SSF , SFF and FFF , as detailed in Eq. (26) for the full period and for the three half centuries 1870–1919 (HC1), 1920–1969 (HC2) and 1970–2018 (HC3). The expectations and their bispectral contributions are further decomposed into contributions conditioned to La Niñas (blue) and conditioned to El Niños (red). Note that $sk = sk_+ + sk_-$ (idem for skewness terms).

	sk	SSS	SSF	SFF	FFF	sk−	SSS−	SSF−	SFF−	FFF−	sk+	SSS+	SSF+	SFF+	FFF+
FULL	0.5	−0.1	0.2	0.2	0.1	−0.6	−0.4	0	−0.2	−0.1	1.1	0.3	0.2	0.4	0.2
HC1	0.7	0.1	0.2	0.3	0.1	−0.5	−0.3	0	−0.1	−0.1	1.2	0.4	0.2	0.4	0.2
HC2	0.2	0	0.1	0.3	0	−0.3	−0.2	0	−0.1	−0.1	0.5	0.2	0.1	0.4	0.1
HC3	0.5	−0.3	0.5	0.2	0.1	−1.0	−0.6	0.2	−0.4	−0.2	1.5	0.3	0.3	0.6	0.3

Gaussian white noise, leads to $\sigma_w^2 = a^2 + 2b^2 = 0.275$ and $\sigma_w^3 E(w^3) = 8b^3 + 6a^2b = 0.129$, yielding $a = 0.5122$, and $b = 0.0794$, and its excess kurtosis is 1.073.

Here, we choose: $\sigma_w w(t) = aw_1(t) + b[w_2^2(t) - 1]$ where w_1, w_2 are independent standard Gaussians with $\sigma_w^2 = a^2 + 2b^2 = 0.275$ and $\sigma_w^3 E(w^3) = 8b^3 = 0.129$, yielding $a = 0.3840$, and $b = 0.2525$ and a excess kurtosis of 2.475. The results are quite robust to changes in the noise model. Other possible, though less practical, noises are generated by maximum entropy constrained by the four first moments (Pires et al., 2010). We then compute the Monte-Carlo ensemble average $E[\hat{S}_{3,\bar{x}}]$ and variance $var[\hat{S}_{3,\bar{x}}]$, of the real and imaginary parts of the smoothed bispectrum. Noise high-order moments (greater than 2) appear only to influence the high-order moments of the smoothed spectrum (e.g. skewness and kurtosis) which are not relevant for the linearity test devised here.

The deviation of El Niño smoothed bispectrum $\hat{S}_{3,x}(f_1, f_2)$ (Fig. 7a,b) from that of NGAR(5) model is assessed by the test statistic (standardized deviation) (Eq. (A6)): $T_{x,\bar{x}}(f_1, f_2) \equiv \frac{\hat{S}_{3,x}(f_1, f_2) - E[\hat{S}_{3,\bar{x}}]}{(var[\hat{S}_{3,\bar{x}}])^{1/2}}$. Under H_0 , its real and imaginary parts are approximately standard Gaussian. We also limited the tests to frequencies with higher bispectrum amplitude, roughly corresponding to $|f_1|, |f_2| < 0.2$ (as in Fig. 7a,c). Figure 7 shows the real (Fig. 7d) and imaginary (Fig. 7e) parts of $T_{x,\bar{x}}(f_1, f_2)$ where significant regions (at $\alpha = 20\%$ significance level), are color-shaded. Fig. 7d shows that most peaks of $Re(\hat{S}_{3,x})$ (Fig. 7a) are significant. In particular, the low-frequency region (SSS type), producing La Niña events for $|f_1|, |f_2|, |f_3| < 0.06$ cpt, is significant (i.e. rejecting H_0). The other positive and negative bispectrum extremes discussed in Section 5.3.2 are also significant at $\alpha = 10\%$. The imaginary part of the test (Fig. 7e) and the squared amplitude (Fig. 7f) are also highly significant in most of the relevant regions of the bispectrum with significance levels reaching 5%. The most significant region of nonlinear predictability holds approximately for $f_1 + f_2$ within $[0.16, 0.18]$ cpt where bispectrum is significant at $\alpha = 5\%$, producing oscillations with periods of the order

5–6 trimesters. This suggests a possible source of the high nonlinear predictability for lags 3–6 trm, as diagnosed by Cox test (Fig. 5c), and for the nonlinear curtailing of El Niño Spring barrier (see Fig. 5a).

The local test $T_{x,\bar{x}}(f_1, f_2)$ computed on a frequency basis may lead, ambiguously, either to the rejection or to the acceptance of linearity, depending on f_1, f_2 (see Fig. 7d–e). This is suggested by the fact that finite N -sized samples generated by the non-Gaussian NGAR(5) model led to local frequency tests $T_{\bar{x},\bar{x}}(f_1, f_2)$ where linearity is falsely rejected (not shown). Therefore, in order to overcome this difficulty and enhance test robustness, we propose the integrated test of nonlinearity (A7) given by $T_{int\ x,\bar{x}} \equiv \sum_{(f_1, f_2) \in L} |T_{x,\bar{x}}(f_1, f_2)|^2$ over a representative lattice L (Fig. A3). We found that nonlinearity cannot be rejected (at 5% level), thanks to the highly significant regions of El Niño bispectrum (Fig. 7d–e and 7a–c).

5.3.6. Normalized bispectrum and bicoherency. An independent test of linearity, beyond that of previous section and diagnostic of phase synchronization comes from the normalized bispectrum or bicoherence spectrum (Kim and Powers, 1979; Nikias and Raghuvver, 1987; Hinich and Wolinsky, 2005; Rao et al., 2012). It is obtained by prewhitening $x(t)$ to yield a non-Gaussian white noise $y(t)$, and reconstructed by inverse FT of $X_y(f) \equiv \frac{X_x(f)}{[\Gamma_x(f)]^{1/2}} = \sqrt{N}e^{i[\Theta_x(f)]}$. The test is then:

$$\Gamma_{3,y}(f_1, f_2) \equiv \frac{\Gamma_{3,x}(f_1, f_2)}{(\Gamma_x(f_1)\Gamma_x(f_2)\Gamma_x(f_3))^{1/2}} \quad (27)$$

We stress that the phases of $X_y(f)$ and $X_x(f)$ are the same (i.e. $\Theta_x(f)$), leading to a nonvanishing correlation $cor(x, y) = \sigma_x^{-1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sqrt{\Gamma_x(f)} df$.

A linear process i.e. $x(t) = \sum_k \alpha(k)w(t-k)$ where w is a white noise yields $\Gamma_x(f) = |X_x(f)|^2 E(w^2)$. By using the result of Section 5.1.2, the normalized bispectrum (27) becomes $\Gamma_{3,y}(f_1, f_2) = sk(w)e^{i[\Theta_x(f_1) + \Theta_x(f_2) - \Theta_x(f_3)]}$ where $\Theta_x(\cdot)$ is the phase of the DFT of sequence $\alpha(k)$. Therefore, the amplitude of $\Gamma_{3,y}$ is uniform,

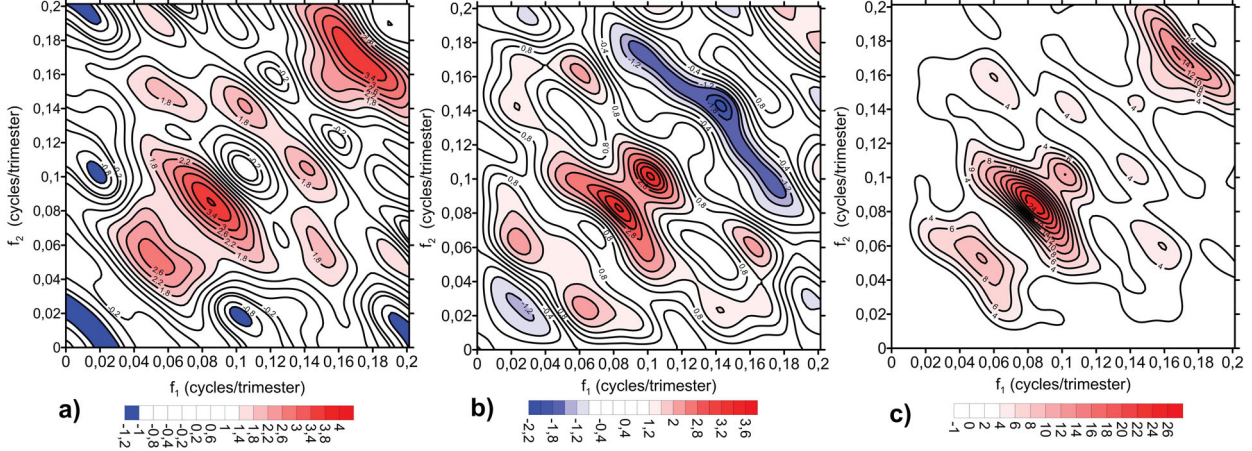


Fig. 9. Real (a) and Imaginary (b) parts of the normalized smoothed bispectrum of El Niño index, ($M_2=30$) and its squared amplitude (c). Regions statistically significant at 20% level are colour-shaded.

i.e. $|\Gamma_{3,y}(f_1, f_2)| = sk(w)$ which is precisely the Hinich (1982) null hypothesis of linearity.

In the case of El Niño, we get an estimated prewhitened non-Gaussian noise $y(t)$ by using the more reliable theoretical maximum entropy NGAR(5) spectrum for the normalization in (27), instead of the empirical smoothed spectrum. The lag correlation of the resulting noise $y(t)$ is very close to zero and thus can be considered a white noise. Its skewness is 0.282 and the excess kurtosis is 0.531. The correlation with the signal $x(t)$ is quite high: 0.75, coming mainly from extreme events.

The smoothed normalized bispectrum, using the same window length $M = M_2 = 30$, is hereby denoted $\hat{S}_{3,y}$. Significant peaks (at 20% significance level) of $\hat{S}_{3,y}$, both of the real (Fig. 9a) and imaginary (Fig. 9b) parts are located nearly at the same bifrequencies as the non-normalized bispectrum $\hat{S}_{3,x}$ (Fig. 7a,b) though peaks are attenuated by normalization and a new peak appears at higher frequencies $f_1 = f_2 = 0.18$ (2.5 years), ($f_3 = 0.36$) cpts (2.8 years). The squared bicoherency $|\hat{S}_{3,y}(f_1, f_2)|^2$ (Fig. 9c) is clearly nonuniform, showing regions of nonacceptance of the null bispectrum of NGAR(5) and hence rejecting the linearity hypothesis.

6. Stochastic modelling

6.1. The method

6.1.1. The motivation. ENSO has been extensively studied to look for a deeper understanding of the underlying physics and complexity (see recent reviews of Chunzai, 2018; Timmermann et al., 2018 and references therein), improve predictability (see the review of Tang et al., 2018), as well as to get accurate statistics (e.g. pdf, extremes etc.). This was done through different dynamical

models (physical-based deterministic models), statistical models (e.g. linear inverse modelling by Penland (1996) and Privalsky and Muzylev, 2013) and models based on machine learning (Dijkstra et al., 2019).

However, even complex models may exhibit biases of various statistics. The present top-down approach of fitting simple stochastic models to observations, attempts learn signal-noise relationships from models, enabling parametrizations of the nonlinear and complex effects of nonobserved variables and hence reproducing a set of relevant stochastic properties (e.g. spectrum, bispectrum).

A number of stochastic univariate models of ENSO have been fitted such as: smooth transition autoregressive (STAR) models (Hall et al., 2001; Ubilava and Helmers, 2013), autoregressive conditional heteroscedasticity type models (ARCH) (Ahn and Kim, 2005), and threshold AR models (De Gooijer, 2017). This section aims at fitting a minimal univariate stochastic model for El Niño 3.4. index driven by a multiplicative Gaussian delayed noise, able to reproduce the observed empirical spectrum, skewness and bispectrum as well as assess the impact it has on predictability, compared to benchmark linear models.

6.1.2. The model formulation. The models fitted here, belong to a class in which the simulated scalar state $\tilde{x}(t)$, at integer t , is driven by a noise $u(t) = \sigma_w w(t)$ where $w(t)$ is a standard Gaussian white noise and σ_w is a positive constant. The simulated state $\tilde{x}(t)$ depends (through a function F) on: a) the simulated previous state values at $t < 0$, represented in delay coordinate (Takens et al., 1981) by $\tilde{\mathbf{x}}(t-1) \equiv [\dots \tilde{x}(t-2), \tilde{x}(t-1)]^T$; b) the previous noise values at $t < 0$: $\mathbf{u}(t-1) \equiv [\dots u(t-2), u(t-1)]^T$ and c) a parameter vector $\boldsymbol{\theta}$. Throughout this section bold letters refer to vectors and italic to scalars. The model writes thus as:

$$\tilde{x}(t) = F[\tilde{x}(t-1), \mathbf{u}(t-1), \boldsymbol{\theta}] + u(t) \quad (28)$$

The model (28) will also be used in forecast mode, and the τ -lag ($\tau \geq 1$) forecast valid at time at t , $x(t, \tau)$, is given by:

$$x(t, \tau) = F[x(t-1, \tau-1), \boldsymbol{\epsilon}(t-1, \tau), \boldsymbol{\theta}], \quad (29)$$

where $\mathbf{x}(t-1, \tau-1) \equiv [\dots, x(t-2, \tau-2), x(t-1, \tau-1)]^T$, with $x(t', \tau) = x(t')$ for $\tau \leq 0$. The noise vector in (29) becomes $\boldsymbol{\epsilon}(t-1, \tau) \equiv [\dots, \epsilon(t-2, \tau-1), \epsilon(t-1, \tau)]^T$ with $\epsilon(t', \tau') \equiv x(t', \tau'-1) - x(t', \tau'')$, where $\tau'' = \max(1, \tau')$. For instance the one-step forecast is $x(t, 1) = F[x(t-1, 0), \boldsymbol{\epsilon}(t-1, 1), \boldsymbol{\theta}]$ where $\mathbf{x}(t-1, 0) = \mathbf{x}(t-1) \equiv [\dots, x(t-2), x(t-1)]^T$ and $\boldsymbol{\epsilon}(t-1, 1) = \boldsymbol{\epsilon}(t-1) = [\dots, \epsilon(t-2), \epsilon(t-1)]^T$ with $\epsilon(t') \equiv x(t') - x(t', 1)$ is the one-step error forecasts, being thus consistent with (28).

The models of form (28) include AR linear models, fitted in Section 3.3.2. However, in order to reproduce nonlinear and non-Gaussian ENSO behaviour, we have considered bilinear models (Haggan and Oyetunji, 1980; Rao and Gabr, 1984; Rao 1981), which differ from AR processes by the addition to them of lagged bilinear (BL) terms, denoted ARBL(p_1, p_2):

$$F[\tilde{x}(t-1), \mathbf{u}(t-1), \boldsymbol{\theta}] = \sum_{k=1}^{p_1} a_k \tilde{x}(t-l_k) + \sum_{k=1}^{p_2} b_k \tilde{x}(t-r_k) u(t-s_k) + \alpha, \quad (30)$$

characterized by its correlated-additive-multiplicative (CAM) noise (Usoro, 2015), where $l_k, r_k, s_k \geq 1$ are lags, with $\boldsymbol{\theta} \equiv (a_1, \dots, a_{p_1}, b_1, \dots, b_{p_2}, \alpha)$. These models can produce non-Gaussian statistics and nonvanishing bicovariances and bispectra (Rao and Gabr, 1984). Note that the restricted case of lags $r_k = 0, s_k = 0$ (Monahan, 2020) are excluded from model of Eq. (30), to allow inverting $u(t)$ from past values for forecasting. Note also that the case $r_k > s_k$ leads to sub-diagonal bilinear models whereas the case $r_k \leq s_k$ corresponds to diagonal/super-diagonal bilinear models. In the former case, for example, the intervening noise is independent of the most recent state whereas the latter case corresponds to nonlinear feedbacks, having in general nonvanishing time averages due to correlation between states and past noises. We stress that models (30) have a nonlinear Volterra development in terms of lagged noises, which in a certain way parametrizes nonlinearities which are not present in a deterministic form in function F . However, other type of models could be fitted, e.g. adding quadratic terms in the deterministic forcing but which can lead to instabilities in simulations. Another difficulty with nonlinear deterministic part, is the very wide class of nonlinearity: which nonlinearity: if

polynomial-what degree? However, we investigate it in another study.

6.1.3. Model fitting. In order to optimize predictability and reproduce data statistics, we apply a hybrid fitting algorithm that minimizes a cost function J_{hyb} which is the weighted sum of the normalized one-step forecast residuals J_{fit} and the normalized squared distance between a set of observed and simulated statistics (average, autocovariance and bi-covariance) J_{stat} :

$$J_{hyb}(\boldsymbol{\theta}) = c_{fit} J_{fit}(\boldsymbol{\theta}) + c_{stat} J_{stat}(\boldsymbol{\theta}, \sigma_w(\boldsymbol{\theta})) \quad (31)$$

where c_{fit} and c_{stat} are positive weights and $\sigma_w(\boldsymbol{\theta}) = \sigma_x (J_{fit}(\boldsymbol{\theta}))^{1/2}$ is the RMS of the forecast residuals for a model using parameters $\boldsymbol{\theta}$ with $\sigma_x = 1$ (observations are standardized).

The term J_{fit} is given by the one-step forecast error:

$$J_{fit}(\boldsymbol{\theta}) = \frac{1}{(N - N_\gamma) \sigma_x^2} \sum_{t=N_\gamma}^{N-1} [x(t) - F[\mathbf{x}(t-1), \boldsymbol{\epsilon}(t-1), \boldsymbol{\theta}]]^2 \quad (32)$$

where $N_\gamma \equiv \max_{k,k'}(l_k, r_{k'}, s_{k'})$.

Iterative minimization algorithms of $J_{fit}(\boldsymbol{\theta})$ for general bilinear models are discussed in Pham and Tran (1981), (Rao and Gabr 1984), Grahm 1995, Guegan and Pham (1989), Gabr 1998 and Falguerolles and Francis (1992). Traditionally, the method of moments is used to obtain implicit relationships between the parameters and lagged moments (e.g. Sesay and Rao, 1988; Tang and Mohler, 1988; Kim et al., 1990), where, in most cases parameters are difficult to be expressed as a function of moments. Here we apply a method where statistics are estimated from a long simulation of (Eq. (28)) with initial conditions: $\tilde{x}(t) = w(t) = 0$; $t = -N_{past}, \dots, -N_{past} + N_\gamma$, and $w(t) \sim N(0, 1)$; $t = -N_{past} + N_\gamma + 1, \dots, N_{sim}$. Statistics are computed for $t = 1, \dots, N_{sim}$ with $N_{sim} = 30,000$ and the initial $N_{past} = 1000$ values are discarded from statistics as spin-up.

The term J_{stat} , involving the mean, autocovariance and bicovariance from observations, ($S_{obs}, C_{obs}(\cdot)$, and $B_{obs}(\cdot, \cdot)$), and simulations ($S_{sim}, C_{sim}(\cdot)$, and $B_{sim}(\cdot, \cdot)$), is

$$J_{stat}(\boldsymbol{\theta}, \sigma_w) = J_S(\boldsymbol{\theta}, \sigma_w) + J_C(\boldsymbol{\theta}, \sigma_w) + J_B(\boldsymbol{\theta}, \sigma_w), \quad (33)$$

where

$$J_S(\boldsymbol{\theta}, \sigma_w) = \frac{(S_{sim} - S_{obs})^2}{\sigma_x^2} \quad (34)$$

$$J_C(\boldsymbol{\theta}, \sigma_w) = \sum_{|\tau| \leq \tau_{max}} [C_{sim}(\tau) - C_{obs}(\tau)]^2 \frac{b_1 \lambda(\tau)^2}{\tau_{max} \sigma_x^4} \sim \frac{b_1}{\tau_{max} \sigma_x^4} \int_{-1/2}^{1/2} [\hat{S}_{2,\tilde{x}}(f) - \hat{S}_{2,x}(f)]^2 df, \quad (35)$$

Table 5. Statistics of the AR(5) and bilinear ARBL(5,1), ..., ARBL(5,5) models with the corresponding bilinear-term lags (r_{p_2}, s_{p_2}) along with the terms J_{fit} and J_{stat} for simple and hybrid fitting in addition to the normalized AIC relative to hybrid fitting. Correlation skills cor_τ are also shown at lags $\tau = 1, 2, 3$ trimesters, assessed over the period 1970–2018 for models obtained by hybrid fitting calibrated in the period 1870–1969.

Model	(r_{p_2}, s_{p_2})	$J_{fit}(simple)$	$J_{stat}(simple)$	$J_{fit}(hybrid)$	$J_{stat}(hybrid)$	NAIC	cor_1	cor_2	cor_3
AR(5)		0.2779	0.0199	0.2784	0.0192	−1.268	0.855	0.552	0.261
ARBL(5,1)	(2,1)	0.2742	0.0211	0.2754	0.0145	−1.277	0.859	0.579	0.309
ARBL(5,2)	(2,4)	0.2708	0.0109	0.2738	0.0055	−1.282	0.864	0.607	0.379
ARBL(5,3)	(1,3)	0.2645	0.0138	0.2693	0.0062	−1.296	0.864	0.609	0.382
ARBL(5,4)	(6,3)	0.2601	0.0146	0.2639	0.0047	−1.315	0.865	0.598	0.333
ARBL(5,5)	(5,1)	0.2566	0.0151	0.2604	0.0034	−1.327	0.867	0.618	0.349

Table 6. Coefficients and lags of the autoregressive and bilinear terms of the ARBL(5,5) model of El Niño index, obtained by hybrid fitting over the full period. The independent constant $\alpha = 0.0626$ and the additive driving noise is $u(t) = 0.5102 w(t)$.

k	a_k	b_k	(r_k, s_k)
1	1.130	0.289	(2,1)
2	−0.384	−0.154	(2,4)
3	−0.072	−0.114	(1,3)
4	0.159	0.100	(6,3)
5	−0.139	−0.108	(5,1)

and

$$J_B(\theta, \sigma_w) \sum_{|\tau_1|, |\tau_2| \leq \tau_{max}} [B_{sim}(\tau_1, \tau_2) - B_{obs}(\tau_1, \tau_2)]^2 \frac{b_2 \lambda(\tau_1, \tau_2)^2}{\tau_{max}^2 \sigma_x^6} \\ \sim \frac{b_2}{\tau_{max}^2 \sigma_x^6} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} [\hat{S}_{3,\bar{x}}(f_1, f_2) - \hat{S}_{3,x}(f_1, f_2)]^2 df_1 df_2, \quad (36)$$

with $\tau_{max} = 16$.

The used window lag functions $\lambda(\tau)$ and $\lambda(\tau_1, \tau_2)$ are scaled by $M = M_2 = \tau_{max}$. The Parseval Theorem applied to (35–36), shows that minimizing $J_C(\theta, \sigma_w)$, $J_B(\theta, \sigma_w)$ leads also to minimizing errors in the spectral domain.

In the analysis we compare two situations: $c_{fit} = 1$, $c_{stat} = 0$ (simple fitting) and the hybrid fitting where $c_{fit} = 1$; $c_{stat} = \Delta J_{fit} / \Delta J_{stat}$, given by the ratio of typical variations of J_{fit} and J_{stat} (hybrid fitting). The optimal parameters issued from simple and hybrid fittings are hereby denoted θ_{fit} and θ_{hyb} , respectively. Note that the term $c_{stat} J_{stat}$ reduces overfitting and the domination of one term over the other, see Appendix B for the description of the minimization of $J_{hyb}(\theta)$. We compare both fittings in terms of J_{fit} and J_{stat} for the AR(p_1) model and several ARBL(p_1, p_2) models. The statistical significance is given through the normalized Akaike Information Criterion: $NAIC = \log(J_{fit}) - 2dim\theta / (N - N_\gamma)$ that penalizes the number of model terms.

6.2. Results for El Niño index

6.2.1. *Fitting statistics.* We analyse and evaluate a sequence of models (Eq. (30)), for $p_1 = 5$, $l_k = k$; $k = 1, \dots, p_1$, with various p_2 values using lags $r_k \leq 6, s_k \leq 6$ (Table 5) with every new lag pair producing the largest J_{fit} decrease.

Table 5 shows that the mean residual squares J_{fit} decreases with increasing model complexity. For hybrid, NAIC decreases with increasing complexity (i.e., no overfitting). Moreover, the hybrid fitting is able to get improved statistics compared to that from the simple fitting ($J_{stat}(\theta_{hyb}) < J_{stat}(\theta_{fit})$), with reductions up to about one fifth for the ARBL(5,5) model. That reduction entails very tiny increments in the sum of residuals, i.e. $J_{fit}(\theta_{hyb}) > J_{fit}(\theta_{fit})$, of the order of 1%, but still keeping insignificant NAIC values.

The estimated of the ARBL(5,5) model are shown in Table 6 and $(\alpha, \sigma_w) = (0.0626, 0.5102)$.

6.2.2. *Simulated autocovariance and spectrum.* We note that the autoregressive part of the ARBL(5,5) model is quite similar to that of AR(5) model (Eq. (6)). The bilinear coefficients have smaller amplitude than the autoregressive coefficients. ARBL(5,5) contains two feedback terms corresponding to $(r_k, s_k) = (2, 4)$ and $(1, 3)$ whose nonzero mean values come from white noise squares and leading to a nonvanishing constant α . This parametrizes partially the bicovariance at the interseasonal scales which is at the origin of the nonlinear reduction of the Spring Predictability Barrier (see Section 4.2.2). The model recovers quite well the empirical autocovariance function (not shown), and also the smoothed spectrum (Fig. 10) using the window lag $M = 30$, with particularly similar peak in the band $f \sim 0.05 - 0.07$ cpt.

6.2.3. *Simulated bicovariance and bispectrum.* Most ($\sim 80\%$) uncertainty in $J_{hyb}(\theta_{hyb})$ comes from the bicovariance uncertainty through $J_B(\theta, \sigma_w)$. The simulated bicovariance (Fig. 11) reproduces very well the same pattern

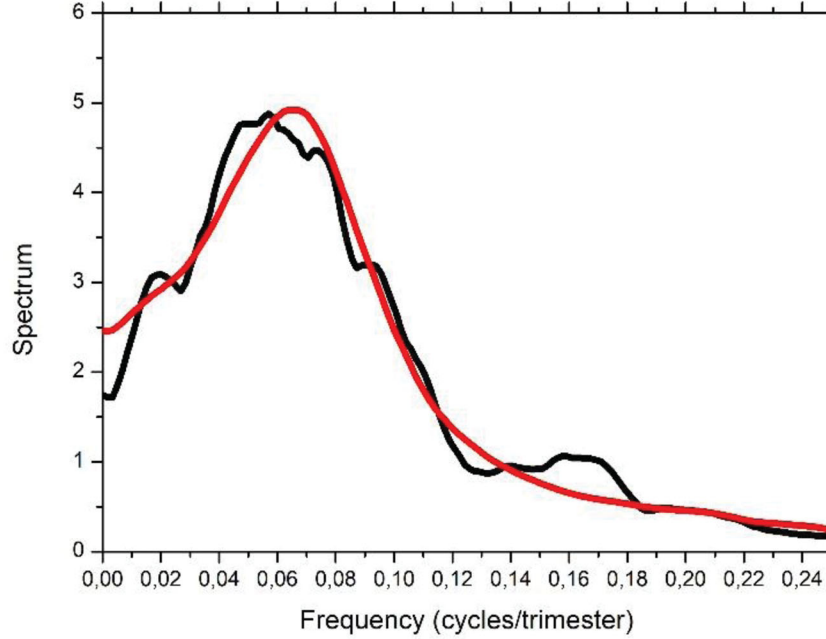


Fig. 10. Smoothed empirical (black) and ARBL(5,5) simulated (red) spectra using a window lag $M = 30$.

of the bicovariance inferred from observations (Fig. 4b), at least up to lags 12 trimesters. The simulation yields a skewness equal to 0.26, which is comparable with the observed skewness (0.46). Any simpler model with only one bilinear term is unable to yield positive skewness and negative bicovariances $C_{\tilde{x}}(0,3)$, $C_{\tilde{x}}(0,4)$, which are fundamental to get skillful El Niño predictions from Spring to the next Winter.

The smoothed bispectrum of the ARBL(5,5) simulation, using a window lag $M_2 = 30$, is shown in Fig. 12a (real part) and Fig. 12b (imaginary part), to be compared with the empirical one (Fig. 7a,b). The real part (Fig. 7a) is quite well reproduced, with negative values in the low-frequency band $|f_1 + f_2| < 0.06$, and positive local maxima near $f_1 = f_2 = 0.045$ cpts and $f_1 = f_2 = 0.085$ cpts (Fig. 12a). The decomposition of skewness (Eq. (26)), (with $f_{cut} = 0.06$) yields the values -0.036 , 0.062 , 0.154 and 0.082 respectively for the SSS, SSF, SFF and SSS contributions, which agree approximately with observations (Table 3). The imaginary part of the simulated bispectrum (Fig. 12b) is also roughly well reproduced, showing negative values within the region $f_1 + f_2 < 0.06$ (Fig. 7b) and positive elsewhere. The maximum is near $f_1 = f_2 \sim 0.07$ cpts. Note that the existence of a single peak is related to the simulated single-peak spectrum.

We shall remark that the bilinear models chosen here, have a linear deterministic part with nonlinearity coming indirectly through the CAM noise. However, other type of models could be fitted, e.g. adding quadratic terms in the deterministic forcing but which can lead to other

difficulties like instabilities and chaotic behaviour in simulations.

6.2.4. Predictability. In order to assess the predictability impact of the inclusion of bilinear terms in ARBL models, compared to the AR(5) model, we compute the correlation skill cor_τ , (Table 5) between observations and predictions for lags for $\tau = 1, 2, 3$ trimesters. Models are optimized by hybrid fitting in the training period 1870–1969 (100 years) and predictions are evaluated in the validation period 1970–2018 (49 years), where the most intense La Niñas and El Niños have been observed (see Section 5.3.4). All the tested models' predictions are skillful ($cor_\tau > 0.5$) for lags up to two trimesters.

Table 5 shows that, in general correlation skills increase with increasing complexity. For instance, the correlation skill of model ARBL(5,5) is about 2%, 7% and 9% larger than AR(5), respectively for lags of 1, 2 and 3 trimesters. The presence of at least two bilinear terms suggests the reason behind the improvement of the 2-trimester forecasts by ARBL(5,5), with respect to AR(5), for which some extreme El Niños (e.g. 1973, 1983, 1998, 2016) are more accurately predicted. The ARBL(5,5) model has no explicit deterministic nonlinearities, which are common to physically-based models. Here, nonlinearities are parametrized through the bilinear terms. Despite the simplicity of the ARBL(5,5) model, its correlation skills ($cor_1 = 0.87$, $cor_2 = 0.62$) are not dramatically smaller, in the same period than those of physically-based models ($cor_1 = 0.85 - 0.91$, $cor_2 = 0.68 - 0.80$) and to

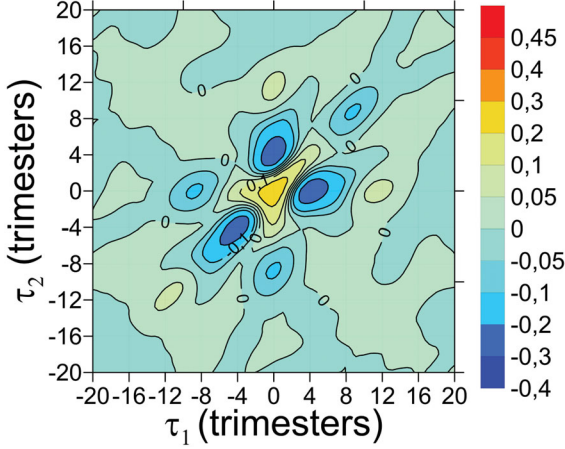


Fig. 11. Bicornelation of the ARBL(5,5) model, approaching the empirical bicornelation of El Niño index (to be compared with Fig. 4b).

those of a much more complex neural-network based model ($cor_1 = 0.92$, $cor_2 = 0.80$) as observed in Fig. 2a of (Ham et al., 2019).

7. Discussion and conclusion

El Niño Southern Oscillation (ENSO) is one of the most important coupled atmosphere–ocean system, exhibiting time scales ranging from seasons to decades and beyond, with a particularly worldwide teleconnection. Using different stochastic and/or dynamic approaches, most studies have emphasized and shown its intrinsic complexity, nonlinearity and non-Gaussianity. Most of those studies limited their investigations to the second order statistics in addition to skewness and/or kurtosis.

Here, we follow the same line of research by performing a data-driven systematic study of the third-order stochastic moments, both in the time and spectral domains, applied to the standardized trimonthly-average El Niño 3.4 index with a trimester sampling. Within the time domain, this comprises the bicovariance $\gamma_x(\tau_1, \tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)]$, in addition to nonlinear correlations for testing nonlinearity and nonlinear predictability for forecasting horizons from seasons up to a few years. The study uses a 149-year period (1870–2018) time series with its statistically significant skewness of 0.46, peaking mostly at the boreal winter (1.04). The analysis of bicovariance maxima reveals high negative nonlinear correlation, implying that El Niño or La Niña extremes are likely to be followed by La Niña one year later, whereas mild conditions, on the other hand, favour El Niño occurrence.

The analysis of the nonlinear predictability, on seasonal time scales, shows that such nonlinear correlation is

enhanced further when forecasts are issued at the NH Spring season (AMJ). This is linked to the persistence of many La Niñas starting in Spring up to next Winter and to the fact that strong Winter El Niños have only occurred under close climatological conditions in the previous boreal Spring. This strongly suggests that nonlinearity in the inter-seasonal timescale can contribute significantly to reduce the so-called El Niño Spring predictability barrier. Another equally important aspect is the fact that ENSO nonlinearity allows for the extension of predictability skill even for forecasting time of a few years, particularly when these forecasting time intervals satisfy phase synchronization and quadratic phase locking with certain dominant Fourier frequencies.

Similarly, within the spectral domain the bispectrum and bicoherence have been computed. As with power spectrum and variance, the bispectrum provides, in particular, the contribution of each bi-frequency bin to the observed skewness and squared bicovariance. This warrants the detection of combinations of El Niño Fourier components that mostly contribute to ENSO extremes by phase synchronization. The bispectrum also permits a test of nonlinearity in the spectral domain. The bispectrum has been estimated by a smoothed estimator using a window lag of 30 trimesters = 7.5 years, obtained from a trade-off between bispectrum bias, variance, and spectral resolution. To estimate the statistical significance of peaks, a conservative test has been adopted. The bispectrum of a non-Gaussian autoregressive null-hypothesis model NGAR(5) is tested and rejected at 5% significance level. We first obtained the coarse-grained spectral partition of the skewness by splitting the full signal into a slow component $s(t)$ (with periods larger than 3 years) and a fast component $f(t)$. The skewness is then decomposed into 4 components, namely $SSS = -0.066$, $SSF = 0.263$, $SFF = 0.185$ and $FFF = 0.071$, implying, in particular, that most El Niños result from interaction between inter- and intra-triennial timescales. Some decadal tendency towards SFF-type El Niños is apparent from the time series, which is consistent with the observed ENSO decadal variability. Note that if a maximum of the bispectrum real part is observed at (f_1, f_2) then a peak in the power spectrum is observed approximately at $f_3 = f_1 + f_2$. In particular, the leading bispectral maxima contributing to El Niño occurs at $f_1 = f_2 = 0.05$, $f_3 = 0.1$ *cpt* (periods of 5 and 2.5 years inside the SSF region) and near the band $f_1 + f_2 = 0.165$ *cpt*, crossing the SFF and FFF regions, with a maximum at $f_1 = f_2 = 0.082$, $f_3 = 0.165$ *cpt* (periods of 2.9 and 1.5 years). On the other hand, minima contributing mostly to La Niña extremes, lie near $f_1 = f_2 = 0.018$, $f_3 = 0.036$ *cpt* (periods of 14 and 7 years) and $(f_1, f_2) = (0.05, 0.018)$, $f_3 = 0.063$ *cpt* (periods of 5, 7 and 4 years), both inside the SSS region.

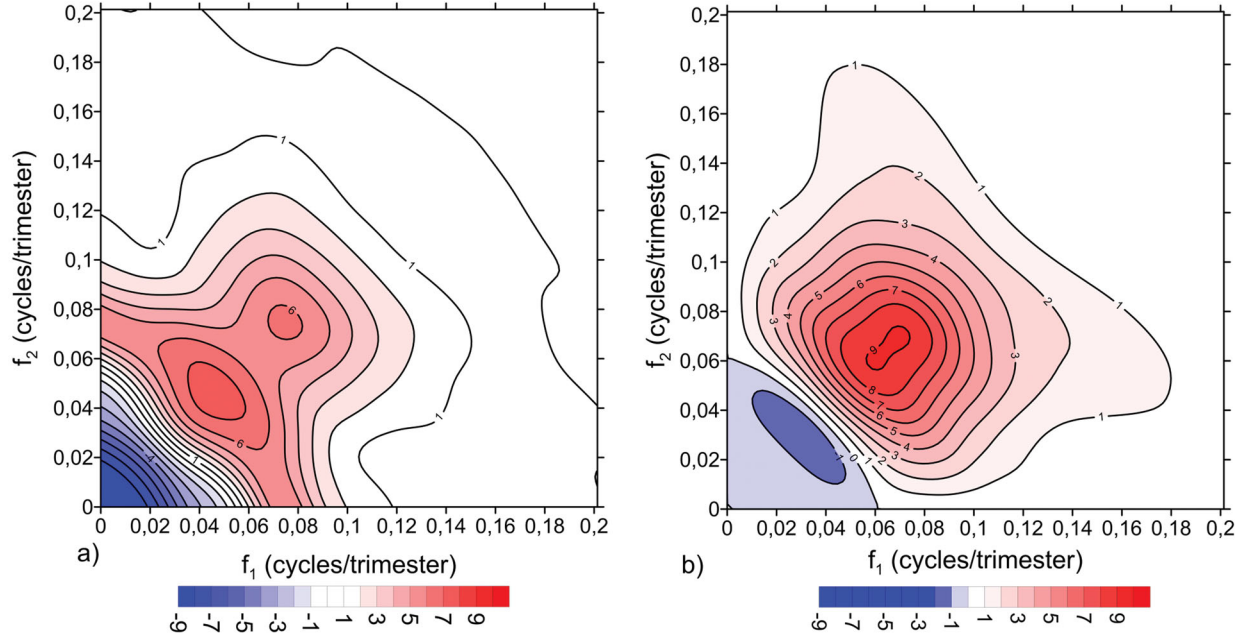


Fig. 12. Real (a) and Imaginary (b) parts of smoothed bispectrum of the ARBL(5,5) using a window lag $M_2=30$.

Lastly, a minimal stochastic model was constructed, which was able to reproduce the main features of the spectrum and bispectrum, and yielded robust improvement of the predictability skill, compared to an autoregressive AR(5) model. The model was selected from a large class of bilinear models with correlated-additive-multiplicative lagged noise. To gain predictability skill with the right stochastic properties, a hybrid fitting approach is used by minimizing a combination of forecasting squared residuals and squared deviations from empirical third-order statistics. The bilinear model yields forecast improvements, particularly at lags of 1, 2 and 3 trimesters with 2%, 7% and 9% of correlation skill increment respectively, suggested to be linked to the attenuation of El Niño predictability Spring barrier and to the more accurate prediction of super El Niños.

This study contributes to the understanding of ENSO predictability and modelling from the perspective of the bispectrum and phase synchronization. In a changing climate, this is especially relevant for the study and predictability of ENSO extremes resulting from resonant-type interaction. The study also provides the possibility to investigate other ENSO indices such as El-Niño Modoki or other Niño indices, and check whether other processes are at play. In particular, forecast skill of ENSO, based on the developed models are of great importance for seasonal (and longer) timescales forecasting. A systematic analysis of these topics is beyond the scope of this manuscript and is left for future research.

Data availability statement

The data that support the findings of this study are available from the NOAA website: https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.anom.data

Acknowledgments

This research was developed at IDL with the support of the Portuguese Foundation for Science and Technology through the FCT – project UIDB/50019/2020 – IDL – Instituto Dom Luiz (IDL) and FCT – project JPIOCEANS/0001/2019 (ROADMAP: ‘The Role of ocean dynamics and Ocean–Atmosphere interactions in Driving climate variations and future Projections of impact-relevant extreme events’). We acknowledge the International Meteorological Institute (IMI) at Stockholm University for hosting Carlos Pires, and also the constructive discussion with Adam Monahan. Two anonymous referees provided constructive comments, which helped improve the manuscript. Many thanks for the undeniable support of the family.

Disclosure statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- Ahn, J. H. and Kim, H. S. 2005. Nonlinear modeling of El Niño/southern oscillation index. *J. Hydrol. Eng.* 10, 8–15. doi:10.1061/(ASCE)1084-0699(2005)10:1(8)
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi:10.1109/TAC.1974.1100705
- An, S. I. 2009. A review of interdecadal changes in the nonlinearity of the El Niño-southern oscillation. *Theor. Appl. Climatol.* 97, 29–40. doi:10.1007/s00704-008-0071-z
- An, S. I. and Jin, F.-F. 2004. Nonlinearity and asymmetry of ENSO. *J. Clim.* 17, 2399–2412. doi:10.1175/1520-0442(2004)017<2399:NAAOE>2.0.CO;2
- An, S. I. and Wang, B. 2000. Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency. *J. Clim.* 13, 2044–2055. doi:10.1175/1520-0442(2000)013<2044:ICOTSO>2.0.CO;2
- Ashley, R., Patterson, D. and Hinich, M. 1986. A diagnostic test for nonlinear serial dependence in time series fitting errors. *J. Time Ser. Anal.* 7, 165–178. doi:10.1111/j.1467-9892.1986.tb00500.x
- Berner, J., Achatz, U., Batté, L., Bengtsson, L., Cámara, A. D. L. and co-authors. 2017. Stochastic parameterization: toward a new view of weather and 574 climate models. *Bull. Am. Meteorol. Soc.* 98, 565–588. doi:10.1175/BAMS-D-15-00268.1
- Berner, J., Sardeshmukh, P. D. and Christensen, H. M. 2018. On the dynamical mechanisms governing El Niño-southern oscillation irregularity. *J. Clim.* 31, 8401–8419. doi:10.1175/JCLI-D-18-0243.1
- Bianucci, M., Capotondi, A., Mannella, R. and Merlino, S. 2018. Linear or nonlinear modeling for ENSO dynamics? *Atmosphere* 9, 435. doi:10.3390/atmos9110435
- Birkelund, Y. and Hanssen, A. 1999. Multitaper estimators for bispectra. *Proc. IEEE Workshop on Higher-Order Statistics*, Caesarea, Israel, pp. 207–211.
- Birkelund, Y. and Hanssen, A. 2000. Adaptive bispectral estimation using Thomson's multitaper approach. *Proc. IEEE Adaptive Systems for Signal Processing, Communication and Control Symposium*, Lake Louise, Alberta, Canada, pp. 283–288.
- Biswas, M., Chandrasekar, A. and Goswami, B. N. 1995. Bispectra of a tropical coupled ocean-atmosphere system. *Curr. Sci.* 68, 1236–1243. <https://www.jstor.org/stable/24096599>
- Boucharel, J., Dewitte, B., Garel, B., Penhoat, Y. and Du. 2009. ENSO's non-stationary and non-Gaussian character: the role of climate shifts. *Nonlin. Process. Geophys.* 16, 453–473. DOI.org/10.5194/npg-16-453-2009 doi:10.5194/npg-16-453-2009
- Brillinger, D. 1965. An introduction to polyspectra. *Ann. Math. Statist.* 36, 1351–1374. doi:10.1214/aoms/1177699896
- Brillinger, D. and Rosenblatt, M. 1967. Asymptotic theory of k-th order spectra. In: *Spectral Analysis of Time Series* (ed. B. Harris). John Wiley, New York, pp. 153–188.
- Burgers, G. and Stephenson, D. B. 1999. The 'normality' of ENSO. *Geophys. Res. Lett.* 26, 1027–1030. doi:10.1029/1999GL900161
- Chunzai, W. 2018. A review of ENSO theories. *Natl. Sci. Rev.* 5, 813–825. doi:10.1093/nsr/nwy104.
- Cox, D. R. 1991. Long-range dependence, nonlinearity and time irreversibility. *J. Time Ser. Anal.* 12, 329–335. doi:10.1111/j.1467-9892.1991.tb00087.x
- De Gooijer, J. G. 2017. *Elements of Nonlinear Time Series Analysis and Forecasting: Springer Series in Statistics*. Springer-Verlag, New York. doi:<https://doi.org/10.1007/978-3-319-43252-6>.
- Deser, C., Alexander, M. A., Xie, S. P. and Phillips, A. S. 2010. Sea surface temperature variability: patterns and mechanisms. *Ann. Rev. Mar. Sci.* 2, 115–143. doi:10.1146/annurev-marine-120408-151453
- Dijkstra, H. A., Petersik, P., Hernández-García, E. and López, C. 2019. The application of machine learning techniques to improve El Niño prediction skill. *Front. Phys.* 7, 153. doi:10.3389/fphy.2019.00153
- Duan, W. and Wei, C. 2013. The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model. *Int. J. Climatol.* 33, 1280–1292. doi:10.1002/joc.3513
- Falguerolles, A. and Francis, B. 1992. Algorithmic approaches for fitting bilinear models. In: *Computational Statistics* (eds. Y. Dodge and J. Whittaker). Physica, Heidelberg, pp. 77–82. doi:https://doi.org/10.1007/978-3-662-26811-7_12
- Fan, H., Huang, B., Yang, S., Li, Z. and Deng, K. 2019. Seasonally-dependent impact of easterly wind bursts on the development of El Niño events. *Clim. Dyn.* 53, 1527–1546. doi:10.1007/s00382-019-04688-2
- Fedorov, A. V. and Philander, S. G. 2000. Is El Niño changing? *Science* 288, 1997–2002. doi:10.1126/science.288.5473.1997
- Franzke, C., Majda, A. J. and Branstator, G. 2007. The origin of nonlinear signatures of planetary wave dynamics: mean phase space tendencies and contributions from non-Gaussianity. *J. Atmos. Sci.* 64, 3987–4003. doi:10.1175/2006JAS2221.1
- Frauen, C., Dommenget, D. and Tyrrell, N. 2014. Analysis of the nonlinearity of El Niño-southern oscillation 33721 teleconnections. *J. Clim.* doi:10.1175/JCLI-D-13-00757.1
- Gabr, M. M. 1998. Robust estimation of bilinear time series models. *Commun. Stat. Theory Methods* 27, 41–53. doi:10.1080/03610929808832649
- Grahn, T. 1995. A conditional least squares approach to bilinear time series estimation. *J. Time Ser. Anal.* 16, 509–529. doi:10.1111/j.1467-9892.1995.tb00251.x
- Granger, C. W. J. and Anderson, A. P. 1978. *Introduction to Bilinear Time Series Models*. Vandenhoeck & Ruprecht, Göttingen.
- Guegan, D. and Pham, D. T. 1989. A note on the estimation of the parameters of the diagonal bilinear model by the method of least squares. *Scand. J. Stat.* 16, 129–136.
- Haggan, V. and Oyetunji, O. 1980. On the selection of subset autoregressive time series models. UMIST Tech. Report 124. Dept. of Maths, UMIST.
- Hall, A., Skalin, J., & Teräsvirta, T. 2001. A nonlinear time series model of El Niño. *Environmental Modelling & Software*, 16(2), 139–146. doi: 10.1016/S1364-8152(00)00077-3
- Ham, Y.-G., Kim, J.-H. and Luo, J.-J. 2019. Deep learning for multi-year ENSO forecasts. *Nature* 573: 568–572. doi:10.1038/s41586-019-1559-7
- Hannachi, A. and Iqbal, W. 2019. On the nonlinearity of winter northern hemisphere atmospheric variability. *J. Atmos. Sci.* 76, 333–356. doi:10.1175/JAS-D-18-0182.1

- Hannachi, A., Straus, D., Franzke, C., Corti, S. and Woollings, T. 2017. Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere. *Rev. Geophys.* 55, 199–234. doi:10.1002/2015RG000509
- Hinich, J. and Wolinsky, M. A. 1988. A test for aliasing using bispectral analysis source. *J. Am. Stat. Assoc.* 83, 499–502. doi:10.1080/01621459.1988.10478623
- Hinich, M. J. 1982. Testing for Gaussianity and linearity of a stationary time series. *J. Time Ser. Anal.* 3, 169–175. doi:10.1111/j.1467-9892.1982.tb00339.x
- Hinich, M. J. and Wolinsky, M. A. 2005. Normalizing bispectra. *J. Stat. Plann. Inference* 130, 405–411. doi:10.1016/j.jspi.2003.12.022
- Hocke, K. and Kämpfer, N. 2008. Bispectral analysis of the long-term recording of surface pressure at Jakarta. *J. Geophys. Res.* 113, D10113. doi:10.1029/2007JD009356
- Jajcay, N., Kravtsov, S., Sugihara, S. G., Tsonis, A. A. and Paluš, M. 2018. Synchronization and causality across time scales in El Niño Southern Oscillation. *Clim. Atmos. Sci.* 1, 33. doi:10.1038/s41612-018-0043-7
- Jenkins, G. M. and Watts, D. G. 1968. *Spectral Analysis and its Applications*. Holden-Day, San Francisco, xviii p. 525.
- Jiang, N., Neelin, J. D. and Ghil, M. 1995. Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. *Clim. Dyn.* 12, 101–112. doi:10.1007/BF00223723
- Kim, Y. C. and Powers, E. J. 1979. Digital Bispectral Analysis and Its Applications to Nonlinear Wave Interactions. *IEEE Trans. Plasma Sci.* 7, 120–131. doi:10.1109/TPS.1979.4317207
- Kim, K. Y. 2002. Investigation of ENSO variability using cyclostationary EOFs of observational data. *Meteorol. Atmos. Phys.* 81, 149–168. doi:10.1007/s00703-002-0549-7
- Kim, W. K., Billard, L. and Basawa, I. V. 1990. Estimation of the first order diagonal bilinear time series model. *J. Time Ser. Anal.* 11, 215–230. doi:10.1111/j.1467-9892.1990.tb00053.x
- Kondrashov, D., Kravtsov, S., Robertson, A. W. and Ghil, M. 2005. A hierarchy of data-based ENSO models. *J. Clim.* 18, 4425–4444. doi:10.1175/JCLI3567.1
- Kovach, C. K., Oya, H. and Kawasaki, H. 2018. The bispectrum and its relationship to phase-amplitude coupling. *Neuroimage*. 173, 518–539. doi:10.1016/j.neuroimage.2018.02.033
- Kravtsov, S. 2012. An empirical model of decadal ENSO variability. *Clim. Dyn.* 39, 2377–2391. doi:10.1007/s00382-012-1424-y
- Lii, K. S. and Rosenblatt, M. 1982. Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* 10, 1195–1208. doi:10.1214/aos/1176345984
- Lovejoy, S. 2018. Spectra, intermittency, and extremes of weather, macroweather and climate. *Sci. Rep.* 8, 1269.
- Marquardt, D. W. 1963. An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* 11, 431–441. doi:10.1137/0111030
- Martinez-Villalobos, C., Newman, M., Vimont, D. J., Penland, C. and Neelin, J. D. 2019. Observed El Niño-La Niña asymmetry in a linear model. *Geophys. Res. Lett.* 46, 9909–9919. doi:10.1029/2019GL082922
- Monahan, A. H. 2020. Bispectral unfolding of the skewness of correlated additive and multiplicative noise processes. *Chaos* 30, 023126. doi:10.1063/1.5125787
- Müller, D. 1987. Bispectra of sea-surface temperature anomalies. *J. Phys. Oceanogr.* 17, 26–36. doi:10.1175/1520-0485(1987)017<0026:BOSSTA>2.0.CO;2
- Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F. F., Wakata, Y. and co-authors. 1998. ENSO theory. *J. Geophys. Res.* 103, 14261–14290. doi:10.1029/97JC03424
- Nikias, C. L. and Raghuveer, M. R. 1987. Bispectrum estimation: a digital signal processing framework. *Proc. IEEE* 75, 869–891. doi:10.1109/PROC.1987.13824
- Penland, C. 1996. A stochastic model of Indo-Pacific sea-surface temperature anomalies. *Phys. D* 98, 534–558. doi:10.1016/0167-2789(96)00124-8
- Pham, D. T. and Tran, L. T. 1981. On the first order bilinear time series models. *J. Appl. Prob.* 18, 617–627. doi:10.2307/3213316
- Pires, C. and Hannachi, A. 2017. Independent subspace analysis of the sea surface temperature variability: non-Gaussian sources and sensitivity to sampling and dimensionality. *Complexity* 2017, 1–23. doi:10.1155/2017/3076810
- Pires, C. and Perdigão, R. 2015. Non-Gaussian interaction information: estimation, optimization and diagnostic application of triadic wave resonance. *Nonlin. Process. Geophys.* 22, 87–108. doi:10.5194/npg-22-87-2015
- Pires, C. A., Talagrand, O. and Bocquet, M. 2010. Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Phys. D* 239, 1701–1717. doi:10.1016/j.physd.2010.05.006
- Priestley, M. B. 1981. *Spectral Analysis and Time Series 1*. Academic, New York.
- Privalsky, V. and Muzylev, S. 2013. An experimental stochastic model of the El Niño-Southern oscillation system at climatic time scales. *Univ. J. Geosci.* 1, 28–36. <http://www.hrpub.org>.
- Proistosescu, C., Rhines, A. and Huybers, P. 2016. Identification and interpretation of nonnormality in atmospheric time series. *Geophys. Res. Lett.* 43, 5425–5434. doi:10.1002/2016GL068880
- Raghuveer, M. R. and Nikias, C. L. 1986. Bispectrum estimation via AR modeling. *Signal Process.* 10, 35–48. doi:10.1016/0165-1684(86)90063-0
- Rao, S. S., Rao, T. S. and Rao, C. R. 2012. *Time Series Analysis: Methods and Applications*, Vol. 30, Elsevier Science and Technology, North-Holland.
- Rao, T. S. 1981. On the theory of bilinear time series models. *J. R. Statist. Soc. B* 43, 2, 244–255.
- Rao, T. S. and Gabr, M. M. 1984. Bilinear time series models. In: *An Introduction to Bispectral Analysis and Bilinear Time Series Models. Lecture Notes in Statistics*, Vol. 24, Springer, New York, NY.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V. and co-authors. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* 108, 4407. 2003. doi:10.1029/2002JD002670
- Richardson, A. M. and Hodgkiss, W. S. 1994. Bispectral analysis of underwater acoustic data. *J. Acoust. Soc. Am.* 96, 828–837. doi:10.1121/1.410321
- Rusticelli, E., Ashley, R. A., Dagum, E. B. and Patterson, D. M. 2008. A new bispectral test for nonlinear serial dependence. *Econ. Rev.* 28, 279–293. doi:10.1080/07474930802388090

- Sardeshmukh, P. D. and Sura, P. 2009. Reconciling non-Gaussian climate statistics with linear dynamics. *J. Clim.* 22, 1193–1207. doi:10.1175/2008JCLI2358.1
- Schulte, J., Policelli, F. and Zaitchik, B. 2020. A skewed perspective of the Indian rainfall–ENSO relationship. *Hydrol. Earth Syst. Sci.*, 24, 5473–5489. doi:10.5194/hess-24-5473-2020
- Sesay, S. A. O. and Rao, T. S. 1988. Yule–Walker type difference equations for higher-order moments and cumulants for bilinear time series models. *J. Time Ser. Anal.* 9, 385–401. doi:10.1111/j.1467-9892.1988.tb00478.x
- Stein, K., Timmermann, A., Schneider, N., Jin, F.-F. and Stuecker, M. F. 2014. ENSO seasonal synchronization theory. *J. Clim.* 27, 5285–5310. doi:10.1175/JCLI-D-13-00525.1
- Stuecker, M. F., Timmermann, A., Jin, F.-F., McGregor, S. and Ren, H.-L. 2013. A combination mode of the annual cycle and the El Niño/Southern oscillation. *Nat. Geosci.* 6, 540–544. doi:10.1038/ngeo1826
- Sun, F. and Yu, J.-Y. 2009. A 10–15-yr modulation cycle of ENSO intensity. *J. Clim.* 22, 1718–1735. doi:10.1175/2008JCLI2285.1
- Sura, P. and Hannachi, A. 2015. Perspectives of non-Gaussianity in atmospheric synoptic and low-frequency variability. *J. Clim.* 28, 5091–5114. doi:10.1175/JCLI-D-14-00572.1
- Sura, P. and Sardeshmukh, P. 2008. A global view of non-Gaussian SST variability. *J. Phys. Oceanogr.* 38, 639–647. doi:10.1175/2007JPO3761.1
- Takens, F., Rand, D. A. and Young, D. S., eds. 1981. *Dynamical Systems and Turbulence (Warwick 1980)*, Lecture Notes in Mathematics, Vol. 898, Springer, Berlin, 365 pp.
- Tang, Y., Zhang, R. H., Liu, T., Duan, W., Yang, D. and co-authors. 2018. Progress in ENSO prediction and predictability study. *Natl. Sci. Rev.* 5, 826–839. doi:10.1093/nsr/nwy105
- Tang, Z. and Mohler, R. R. 1988. Bilinear time series: theory and applications. In *Lecture Notes in Control and Information Sciences* (eds. M. Thoma and A. Wyner), 106, pp. 43–58. Springer-Verlag, Berlin.
- Timmermann, A. 2003. Decadal ENSO amplitude modulations: a nonlinear paradigm. *Glob. Planet. Change* 37, 135–156. doi:10.1016/S0921-8181(02)00194-7
- Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W. and co-authors. 2018. El Niño–Southern oscillation complexity. *Nature* 559, 535–545. doi:https://doi.org/10.1038/s41586-018-0252-6
- Timmermann, A., Voss, H. U. and Pasmanter, R. 2001. Empirical dynamical system modeling of ENSO using nonlinear inverse techniques. *J. Phys. Oceanogr.* 31, 1579–1598. doi:10.1175/1520-0485(2001)031<1579:EDSMOE>2.0.CO;2
- Ubilava, D. and Helmers, C. G. 2013. Forecasting ENSO with a smooth transition autoregressive model. *Environ. Model. Softw.* 40, 181–190. doi:10.1016/j.envsoft.2012.09.008
- Usono, A. E. 2015. Comparative analysis of linear and bilinear time series models. *American J. Math. Stat.* 5, 265–271.
- von Storch, H. and Zwiers, F. W. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, 484 p. doi:https://DOI.org/10.1017/CBO9780511612336
- Wang, C. 2018. A review of ENSO theories. *Natl. Sci. Rev.* 5, 813–825. doi:10.1093/nsr/nwy104
- Wang, C., Deser, C., Yu, J.-Y., DiNezio, P., and Clement, A. 2016. El Niño–Southern Oscillation (ENSO): a review. In *Coral Reefs of the Eastern Pacific* (eds. P. Glynn, D. Manzello, and I. Enochs). Springer Science Publisher, Berlin, pp. 85–106.
- Weiss, G. 1975. Time-reversibility of linear stochastic processes. *J. Appl. Prob.* 12, 831–836. doi:10.2307/3212735
- Wilks, D. S. 2011. *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, Academic Press, Amsterdam.
- Woollings, T., Hannachi, A. and Hoskins, B. 2010. Variability of the North Atlantic eddy-driven jet stream. *QJR. Meteorol. Soc.* 136, 856–868. doi:10.1002/qj.625
- Wu, A. and Hsieh, W. W. 2003. Nonlinear interdecadal changes of the El Niño–Southern oscillation. *Clim. Dyn.* 21, 719–730. doi:10.1007/s00382-003-0361-1
- Yeh, S.-W. and Kirtman, B. P. 2004. Tropical Pacific decadal variability and ENSO amplitude modulation in a CGCM. *J. Geophys. Res.* 109, C11009. doi:10.1029/2004JC002442
- Yishuai, J., Zhengyao, L. and Zhengyu, L. 2020. Controls of spring persistence barrier strength in different ENSO regimes and implications for 21st century changes. *Geophys. Res. Lett.* 47, 11. doi:10.1029/2020GL088010
- Zhang, Y., Wallace, J. M. and Battisti, D. S. 1997. ENSO-like interdecadal variability: 1900–93. *J. Clim.* 10, 1004–1020. doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2

Appendix A: Statistics of the smoothed bispectrum estimator

Let us denote the error of the smoothed bispectrum estimator (19) as:

$$\delta\hat{S}_{3,x}(f_1, f_2) \equiv \hat{S}_{3,x}(f_1, f_2) - \Gamma_{3,x}(f_1, f_2). \quad (\text{A1})$$

Equation (A1) follows (for large $N, M_2, N/M_2^2$), a complex Gaussian PDF (Brillinger, 1965) with independent real and imaginary parts. The estimator's bias is approximately given by:

$$E[\delta\hat{S}_{3,x}(f_1, f_2)] \sim \frac{-\hat{\lambda}''(0)}{4\pi^2 M_2^2} \left(\frac{\partial^2 \Gamma_{3,x}}{\partial f_1^2} + \frac{\partial^2 \Gamma_{3,x}}{\partial f_2^2} - \frac{\partial^2 \Gamma_{3,x}}{\partial f_1 \partial f_2} \right), \quad (\text{A2})$$

Asymptotically, the real and imaginary parts of the estimator (19) have equal variances (Rao and Gabr, 1984), for different f_1, f_2, f_3 , given by

$$\text{var}[Re(\hat{S}_{3,x})] = \text{var}[Im(\hat{S}_{3,x})] = 0.5 \frac{M_2^2}{Nb_2} \Gamma_x(f_1) \Gamma_x(f_2) \Gamma_x(f_3) \quad (\text{A3})$$

which decreases both with N and the two-dimensional bandwidth $\frac{\sqrt{b_2}}{M_2}$ where $\frac{1}{b_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{\lambda}(u) \hat{\lambda}(v) \hat{\lambda}(u-v)]^2 du dv$, and hence larger spectral resolution implies larger variance.

The optimal value of M_2 minimizing the bispectrum MSE is frequency dependent (Rao and Gabr, 1984). Instead, we use an overall criterion such that typical bispectrum fluctuations $\sigma_{\hat{S}_{3,x}}$ are larger than the average confidence interval half-size (cish), which writes as:

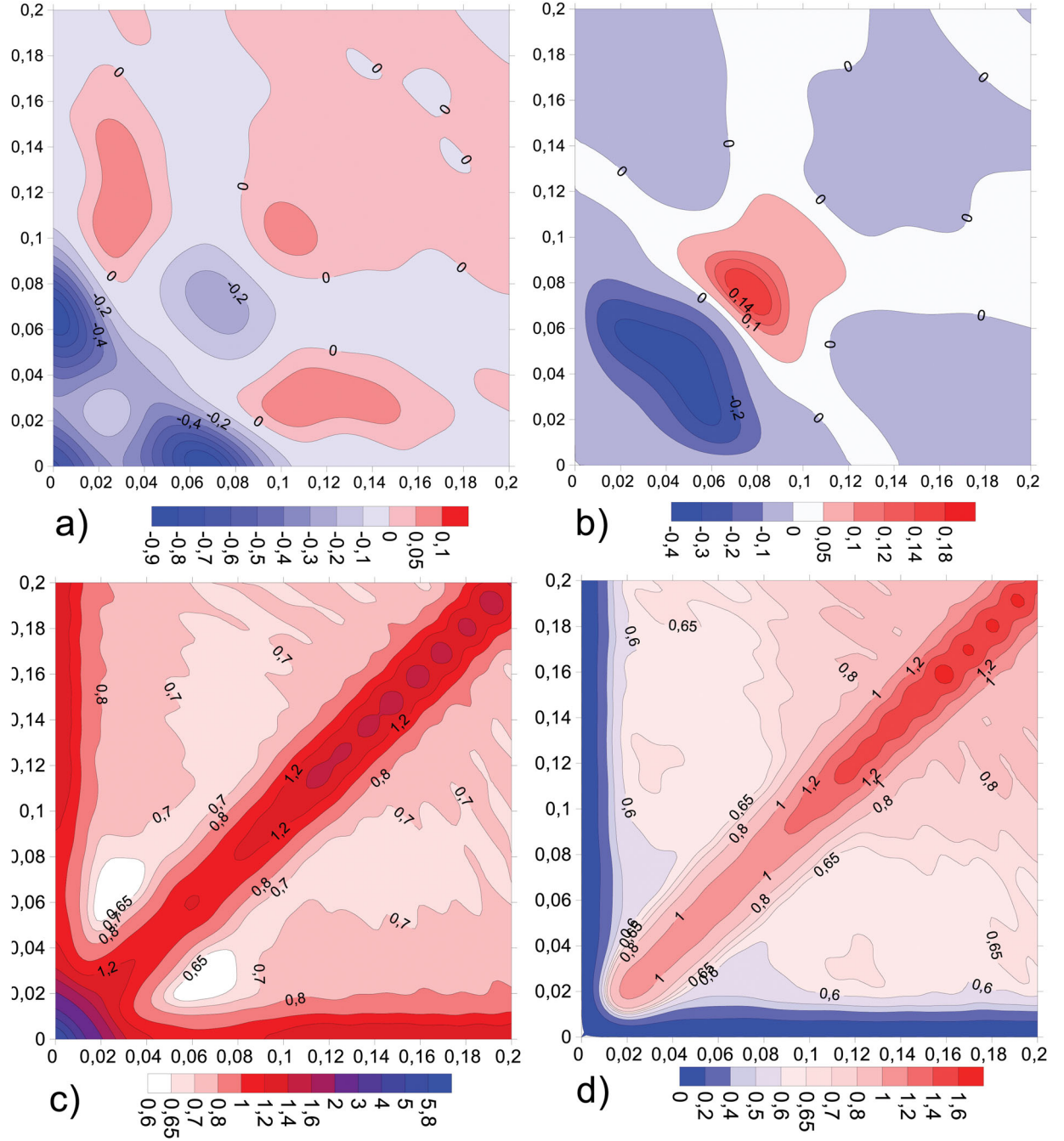


Fig. A1. Real (a) and imaginary (b) parts of bias $E[\delta\hat{S}_{3,\bar{x}}]$. Real (c) and imaginary (d) values of the ratio $\text{var}_{\text{rat}}(\hat{S}_{3,\bar{x}})$ between empirical variances $\text{var}(\hat{S}_{3,x})$ and total asymptotic variance.

$$\sigma_{\hat{S}_{3,x}}^2 \equiv \max \left[\overline{\text{Re}(\hat{S}_{3,x} - sk(x)^3)^2}, \overline{\text{Im}(\hat{S}_{3,x})^2} \right] > (\Phi_{1-\frac{\alpha}{2}})^2 \frac{M_2^2}{Nb_2} \overline{\Gamma_x(f_1)\Gamma_x(f_2)\Gamma_x(f_3)} = \text{cish}^2 \quad (\text{A4})$$

where the overbar means integration over the domain of bifrequencies, and $\Phi_q(\cdot)$ is q th quantile of the standard

normal. Then, for the best bispectral resolution we chose the largest M_2 satisfying (A4).

Here, we are looking for a statistical test of the null bispectrum of NGAR(5). For that, we need good bias and variance approximations of the smoothed bispectrum. To achieve this, even for a reduced number of temporal degrees of freedom (the case here), we have generated 10,000 Monte-Carlo N -sized time series or

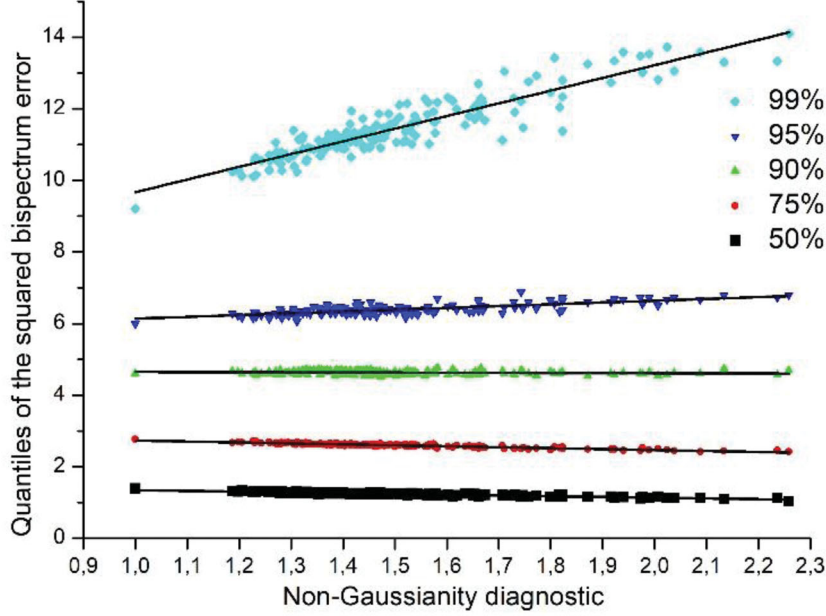


Fig. A2. Values of the non-Gaussianity diagnostic $Y_{\bar{x}}$ (equal to 1 under Gaussianity) versus the 50%, 75%, 90%, 95% and 99% half-quantiles $\frac{1}{2}q_p(|T_{\bar{x},\bar{x}}|^2)$ of the squared bispectrum normalized error, collected over all the 173 bifrequencies of lattice L . A linear adjustment is shown for each quantile.

surrogates $\tilde{x}(t)$ based on the null H_0 , and constructed ensemble statistics. The bispectrum error is $\delta\hat{S}_{3,\bar{x}}(f_1, f_2) = \hat{S}_{3,\bar{x}}(f_1, f_2) - \Gamma_{3,\bar{x}}(f_1, f_2)$ with bias $E[\delta\hat{S}_{3,\bar{x}}] = E[\hat{S}_{3,\bar{x}}] - \Gamma_{3,\bar{x}}$. Monte-Carlo biases of the NGAR(5) bispectrum (Fig. A1a,b) are positive (negative) near the local bispectrum minima (maxima) (see Figs. 7a,b), in agreement with the asymptotic bias (A2). Exception holds near the origin due possibly to under sampling of low frequencies.

The deviation between the asymptotic variance (A3) and the Monte-Carlo variance $\text{var}[\hat{S}_{3,\bar{x}}]$ (for the real and imaginary parts) is assessed through the ratio:

$$\text{var}_{\text{rat}}(\hat{S}_{3,\bar{x}}) \equiv \frac{\text{var}(\hat{S}_{3,\bar{x}})}{\frac{M_2^2}{Nb_2} \Gamma_{\bar{x}}(f_1) \Gamma_{\bar{x}}(f_2) \Gamma_{\bar{x}}(f_3)}, \quad (\text{A5})$$

shown in Fig. A1c,d. It can be verified from this figure that when the distance to the principal domain edges or vertices slightly surpasses the bandwidth $\frac{\sqrt{b_2}}{M_2} = 0.036 \text{ cpt}$, then the ratio is nearly constant, varying in the range 0.65–0.80, which is larger than the value 0.5 observed in asymptotic conditions, and hence (A3) underestimates the estimator's variance (due to sampling ($N_{\text{eff}} \ll N$)).

In order to evaluate the distance between the estimated $\hat{S}_{3,x}(f_1, f_2)$ and the H_0 bispectrum, we compute the standardized error (real and imaginary parts):

$$T_{x,\bar{x}}(f_1, f_2) \equiv \frac{\hat{S}_{3,x}(f_1, f_2) - E[\hat{S}_{3,\bar{x}}]}{[\text{var}(\hat{S}_{3,\bar{x}})]^{1/2}} \quad (\text{A6})$$

Asymptotically, under H_0 , both real and imaginary parts of $T_{x,\bar{x}}$ are independent standard Gaussian, and $|T_{x,\bar{x}}|^2 \equiv \text{Re}(T_{x,\bar{x}})^2 + \text{Im}(T_{x,\bar{x}})^2 \sim \chi_2^2$. However, for finite samples, $T_{x,\bar{x}}$ tends to be non-Gaussian and leptokurtic and hence the diagnostic $Y_{\bar{x}} \equiv \frac{\text{var}(|T_{x,\bar{x}}|^2)}{2E(|T_{x,\bar{x}}|^2)}$ gets values larger than 1. For the purpose of H_0 tests, we also compute the Monte-Carlo quantiles $q_p(|T_{x,\bar{x}}|^2)$. Therefore, the H_0 bispectrum is rejected at the frequency pair (f_1, f_2) if $T_{x,\bar{x}}(f_1, f_2)$ or $|T_{x,\bar{x}}|^2$ fall outside their H_0 nonrejection intervals.

The non-Gaussianity assessment of $T_{x,\bar{x}}$ is presented in Fig. A2, showing $Y_{\bar{x}}$ versus $\frac{1}{2}q_p(|T_{x,\bar{x}}|^2)$, for $p = 50\%, 75\%, 90\%, 95\%$ and 99% , of the local test $|T_{x,\bar{x}}|^2$. The leftmost points in Fig. A2 are set to the Gaussian case ($Y_{\bar{x}} = 1$) and quantiles are those of $\frac{1}{2}\chi_2^2$. Moreover, quantiles are well adjusted by a linear fitting of $Y_{\bar{x}}$. Quantiles up to $p = 95\%$ are well approximated by those of $\frac{1}{2}\chi_2^2$, but less for the whereas the 99% . Consequently, to test H_0 by using $T_{x,\bar{x}}$ and $|T_{x,\bar{x}}|^2$, with 5% significance or larger, we can use the χ_2^2 .

The local test (A6) used in Section 5.3.5 is not very powerful because H_0 can be rejected at random frequencies

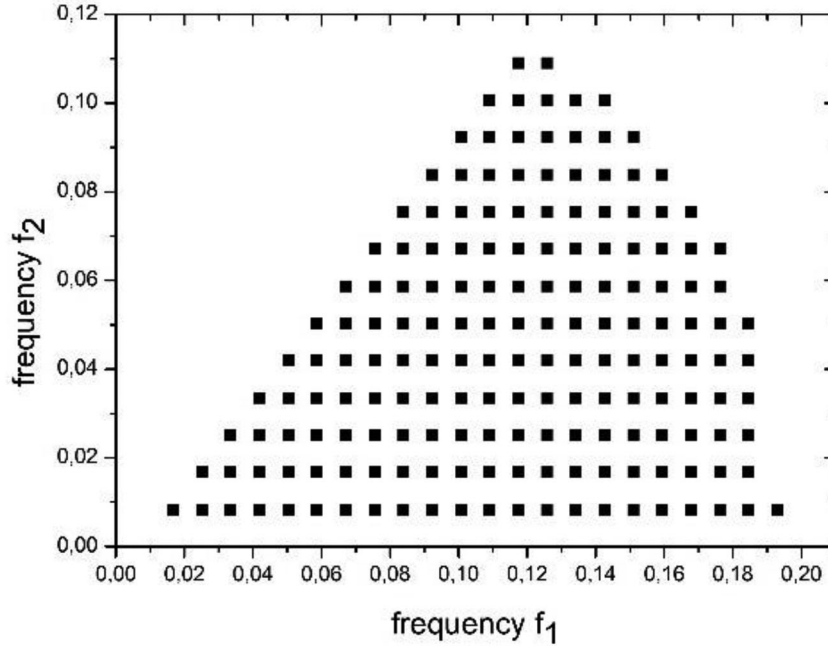


Fig. A3. Position of bifrequencies in the set L used in the integral test of linearity.

even for finite H_0 samples. To overcome this an integrated test is used based on the total squared amplitudes:

$$T_{int \ x, \bar{x}} \equiv \sum_{(f_1, f_2) \in L} |T_{x, \bar{x}}(f_1, f_2)|^2, \quad (A7)$$

along n_L bifrequencies of a regularly sampled lattice L , with frequency step, $\Delta f = \frac{5}{N} = 0.0084$ cpm, i.e. $\frac{1}{4}$ of the two-dimensional (2D) bandwidth $\frac{\sqrt{b_2}}{M_2}$ to get a good bispectrum representation. $T_{x, \bar{x}}$ values become uncorrelated when the distance between bifrequencies is typically greater than the 2D bandwidth. The lattice L (Fig. A3) covers the significant region, where $\frac{\Gamma_{\bar{x}}(f_1)\Gamma_{\bar{x}}(f_2)\Gamma_{\bar{x}}(f_3)}{\text{Max}_{PD}(\Gamma_{\bar{x}}(f_1)\Gamma_{\bar{x}}(f_2)\Gamma_{\bar{x}}(f_3))} \geq \varepsilon_L = 0.01$, yielding $n_L = 173$ bifrequencies.

Under H_0 , the test (A7) has an average $2n_L = 346$ and variance equal to 16527, much larger than $\text{var}(\chi_{2n_L}^2) = 692$, due to correlated bispectrum errors and due to sampled bispectra very far from the theoretical bispectrum. The upper quantiles of $T_{int \ x, \bar{x}}$ for probabilities 80%, 95%, 99% and 99.5% are 489, 563, 804 and 903 respectively, which are the rejection thresholds of NGAR(5) H_0 bispectrum, used in Section 5.3.5.

Appendix B: Minimization of the cost function of the hybrid fitting

B.1. Method

The minimization of $J_{hyb}(\theta)$ is based on a Quasi-Newton method, and to prevent bad-conditioning,

the guess correction is performed along the low-rank subspace spanned by the leading eigenvectors of the Hessian matrix H :

$$\theta_{k+1} = \theta_k - \delta [H_{lr}(J_{hyb}(\theta_k))]^{-1} \text{grad}_{lr}(J_{hyb}(\theta_k)) \quad (B1)$$

where $\delta \in]0, 1]$ weights the iteration step, $H_{lr} = R_{lr} D_{lr} R_{lr}'$ is the low-rank Hessian diagonalisation, and $\text{grad}_{lr} = R_{lr} \text{grad}$ is the low-rank gradient. The iterative procedure stops when the relative difference between consecutive values is less than 10^{-6} . The gradient and the Hessian matrix are obtained by time-averages of auxiliary linear models that we describe below.

B.2. Gradient of the cost function

The cost function (Eq. (37)) of the hybrid fitting is:

$$J_{hyb}(\theta) = c_{fitt} J_{fitt}(\theta) + c_{stat} J_{stat}(\theta, \sigma_w(\theta)) \quad (B2)$$

where $\sigma_w \theta = \sigma_x(J_{fitt}(\theta))^{1/2}$ and $\theta = (\theta_1, \dots, \theta_{dim\theta})'$. The partial derivative of J_{hyb} with respect to θ_i is

$$\frac{\partial J_{hyb}}{\partial \theta_i} = c_{fitt} \frac{\partial J_{fitt}}{\partial \theta_i} + c_{stat} \left(\frac{\partial J_{stat}}{\partial \theta_i} + \frac{\partial J_{stat}}{\partial \sigma_w} \frac{\partial \sigma_w}{\partial \theta_i} \right) \quad (B3)$$

where $\frac{\partial \sigma_w}{\partial \theta_i} = \frac{\sigma_w}{2J_{fitt}} \frac{\partial J_{fitt}}{\partial \theta_i}$. The derivatives of J_{fitt} are

$$\frac{\partial J_{fitt}}{\partial \theta_i} = \frac{2}{(N - N_\gamma) \sigma_x^2} \sum_{t=N-N_\gamma}^{N-1} \varepsilon(t) \frac{\partial \varepsilon(t)}{\partial \theta_i}, \quad (B4)$$

where $\epsilon(t) = x(t) - F[\mathbf{x}(t-1), \boldsymbol{\varepsilon}(t-1), \boldsymbol{\theta}]$ and $\frac{\partial \epsilon(t)}{\partial \theta_i} = -\left(\frac{\partial F}{\partial \theta_i} + \frac{\partial F}{\partial \boldsymbol{\varepsilon}(t-1)} \cdot \frac{\partial \boldsymbol{\varepsilon}(t-1)}{\partial \theta_i}\right)$ with the dot standing for inner product along all components of the vector $\boldsymbol{\varepsilon}(t-1)$ of past innovations. Those derivatives are random variables governed by an AR linear model whose random coefficients depend on the time-series values. Therefore $\frac{\partial J_{\text{stat}}}{\partial \theta_i}$ is approximated by a time average of a product containing the derivative $\frac{\partial \epsilon(t)}{\partial \theta_i}$.

The derivatives of J_{stat} are

$$\begin{aligned} \frac{\partial J_{\text{stat}}}{\partial \theta_i} = & \frac{2(\mathbf{S}_{\text{sim}} - \mathbf{S}_{\text{obs}})}{\boldsymbol{\sigma}_x^2 N_{\text{sim}}} \sum_{t=1}^{N_{\text{sim}}} \frac{\partial \tilde{x}(t)}{\partial \theta_i} + \frac{2\mathbf{b}_1}{\tau_{\text{max}} \boldsymbol{\sigma}_x^4} \sum_{|\tau| \leq \tau_{\text{max}}} \lambda(\tau)^2 [\mathbf{C}_{\text{sim}}(\tau) - \mathbf{C}_{\text{obs}}(\tau)] \frac{\partial \mathbf{C}_{\text{sim}}(\tau)}{\partial \theta_i} \\ & + \frac{2\mathbf{b}_2}{\tau_{\text{max}}^2 \boldsymbol{\sigma}_x^6} \sum_{|\tau_1|, |\tau_2| \leq \tau_{\text{max}}} \lambda(\tau_1, \tau_2)^2 [\mathbf{B}_{\text{sim}}(\tau_1, \tau_2) - \mathbf{B}_{\text{obs}}(\tau_1, \tau_2)] \frac{\partial \mathbf{B}_{\text{sim}}(\tau_1, \tau_2)}{\partial \theta_i} \end{aligned} \quad (\text{B5})$$

with

$$\frac{\partial \mathbf{C}_{\text{sim}}(\tau)}{\partial \theta_i} = \frac{1}{N_{\text{sim}} - \tau} \sum_{t=1}^{N_{\text{sim}} - \tau} \left[\tilde{x}(t) \frac{\partial \tilde{x}(t+\tau)}{\partial \theta_i} + \tilde{x}(t+\tau) \frac{\partial \tilde{x}(t)}{\partial \theta_i} \right] \quad (\text{B6})$$

$$\frac{\partial \mathbf{B}_{\text{sim}}(\tau_1, \tau_2)}{\partial \theta_i} = \frac{1}{N - \max(\tau_1, \tau_2)} \sum_{t=1}^{N - \max(\tau_1, \tau_2)} \left[\tilde{x}(t) \tilde{x}(t+\tau_1) \frac{\partial \tilde{x}(t+\tau_2)}{\partial \theta_i} + \tilde{x}(t) \tilde{x}(t+\tau_2) \frac{\partial \tilde{x}(t+\tau_1)}{\partial \theta_i} + \tilde{x}(t+\tau_1) \tilde{x}(t+\tau_2) \frac{\partial \tilde{x}(t)}{\partial \theta_i} \right] \quad (\text{B7})$$

where $\tilde{x}(t) = F[\tilde{\mathbf{x}}(t-1), \mathbf{u}(t-1), \boldsymbol{\theta}] + \sigma_w w(t)$, $\mathbf{u}(t-1) = \sigma_w \mathbf{w}(t-1)$ and $\frac{\partial \tilde{x}(t)}{\partial \theta_i} = \left(\frac{\partial F}{\partial \theta_i} + \frac{\partial F}{\partial \tilde{\mathbf{x}}(t-1)} \cdot \frac{\partial \tilde{\mathbf{x}}(t-1)}{\partial \theta_i}\right)$ applied to a generic time. An identical formula applies to $\frac{\partial J_{\text{stat}}}{\partial \sigma_w}$, depending on partial derivatives $\frac{\partial \tilde{x}(t)}{\partial \sigma_w} = \left(\frac{\partial F}{\partial \mathbf{u}(t-1)} \cdot \mathbf{w}(t-1) + \frac{\partial F}{\partial \tilde{\mathbf{x}}(t-1)} \cdot \frac{\partial \tilde{\mathbf{x}}(t-1)}{\partial \sigma_w}\right) + w(t)$ where dot stands for inner products along components of the standard Gaussian past innovations $\mathbf{w}(t-1)$ and components of the past states $\tilde{\mathbf{x}}(t-1)$. Finally, we must see that $\frac{\partial J_{\text{stat}}}{\partial \theta_i}$ is made by time-averages along the simulation period of a sum of products containing the random derivatives $\frac{\partial \tilde{x}(t)}{\partial \theta_i}$.

The particular form of the derivatives $\frac{\partial u(t)}{\partial \theta_i}$, $\frac{\partial \tilde{x}(t)}{\partial \theta_i}$, $\frac{\partial \tilde{x}(t)}{\partial \sigma_w}$ depend on the stochastic model. In the case of bilinear models (Eq. (29)), Rao and Gabr (1984) have obtained similar formulas, for $t = 0, \dots, N-1$:

$$\frac{\partial u(t)}{\partial a_k} = -\left(x(t-l_k) + \sum_{i=1}^{p_2} b_i x(t-r_i) \frac{\partial u(t-s_i)}{\partial a_k}\right) \quad (\text{B8})$$

$$\frac{\partial u(t)}{\partial b_k} = -\left(x(t-r_k) u(t-s_k) + \sum_{i=1}^{p_2} b_i x(t-r_i) \frac{\partial u(t-s_i)}{\partial a_k}\right) \quad (\text{B9})$$

$$\frac{\partial u(t)}{\partial \alpha} = -\left(1 + \sum_{i=1}^{p_2} b_i x(t-r_i) \frac{\partial u(t-s_i)}{\partial \alpha}\right), \quad (\text{B10})$$

using initial conditions as pseudo-observations, $x(t) = u(t) = \frac{\partial u(t)}{\partial \theta_i} = 0$ for $t = -N_\gamma - 1, \dots, -1$. The derivatives of the states are, for $t = 0, \dots, N_{\text{sim}}$:

$$\frac{\partial \tilde{x}(t)}{\partial a_k} = x(t-l_k) + \sum_{i=1}^{p_1} a_i \frac{\partial \tilde{x}(t-l_i)}{\partial a_k} + \sum_{i=1}^{p_2} \sigma_w b_i \frac{\partial \tilde{x}(t-r_i)}{\partial a_k} w(t-s_i) \quad (\text{B11})$$

$$\frac{\partial \tilde{x}(t)}{\partial b_k} = \sigma_w x(t-r_k) w(t-s_k) + \sum_{i=1}^{p_1} a_i \frac{\partial \tilde{x}(t-l_i)}{\partial b_k} + \sum_{i=1}^{p_2} \sigma_w b_i \frac{\partial \tilde{x}(t-r_i)}{\partial b_k} w(t-s_i) \quad (\text{B12})$$

$$\frac{\partial \tilde{x}(t)}{\partial \alpha} = 1 + \sum_{i=1}^{p_1} a_i \frac{\partial \tilde{x}(t-l_i)}{\partial \alpha} + \sum_{i=1}^{p_2} \sigma_w b_i \frac{\partial \tilde{x}(t-r_i)}{\partial \alpha} w(t-s_i) \quad (\text{B13})$$

$$\frac{\partial \tilde{x}(t)}{\partial \sigma_w} = w(t) + \sum_{i=1}^{p_1} \sigma_w a_i \frac{\partial \tilde{x}(t-l_i)}{\partial \sigma_w} + \sum_{i=1}^{p_2} \sigma_w b_i \frac{\partial \tilde{x}(t-r_i)}{\partial \sigma_w} w(t-s_i), \quad (\text{B14})$$

using initial conditions as pseudo-observations, $\tilde{x}(t) = \alpha$; $w(t) = \frac{\partial \tilde{x}(t)}{\partial \theta_i} = 0$, $\frac{\partial \tilde{x}(t)}{\partial \alpha} = 1$ for $t = -N_{\text{past}}, \dots, 0$, and the derivatives are then substituted into the derivatives of J_{stat} .

B.3. Hessian matrix of the cost function

The Hessian matrix is obtained from:

$$\frac{\partial^2 J_{hyb}}{\partial \theta_i \partial \theta_j} = c_{fitt} \frac{\partial^2 J_{fitt}}{\partial \theta_i \partial \theta_j} + c_{stat} \left[\frac{\partial^2 J_{stat}}{\partial \theta_i \partial \theta_j} + \frac{\partial J_{stat}}{\partial \sigma_w} \frac{\partial^2 \sigma_w}{\partial \theta_j \partial \theta_i} + \frac{\partial \sigma_w}{\partial \theta_i} \frac{\partial^2 J_{stat}}{\partial \theta_j \partial \sigma_w} + \frac{\partial \sigma_w}{\partial \theta_j} \frac{\partial^2 J_{stat}}{\partial \theta_i \partial \sigma_w} + \frac{\partial \sigma_w}{\partial \theta_i} \frac{\partial \sigma_w}{\partial \theta_j} \frac{\partial^2 J_{stat}}{\partial \sigma_w \partial \sigma_w} \right] \quad (B15)$$

where $\frac{\partial^2 \sigma_w}{\partial \theta_i \partial \theta_j} = \frac{\sigma_w}{2J_{fitt}} \frac{\partial^2 J_{fitt}}{\partial \theta_i \partial \theta_j} - \frac{\sigma_w}{4J_{fitt}^2} \frac{\partial J_{fitt}}{\partial \theta_i} \frac{\partial J_{fitt}}{\partial \theta_j}$. Because the functions involved are sums of squares, we use the approximation of Marquardt (1963), i.e., products of first derivatives. The second derivatives of J_{fitt} are approximated as

$$\frac{\partial^2 J_{fitt}}{\partial \theta_i \partial \theta_j} \sim \frac{2}{(N - N_\gamma) \sigma_x^2} \sum_{t=N-N_\gamma}^{N-1} \frac{\partial u(t)}{\partial \theta_i} \frac{\partial u(t)}{\partial \theta_j} \quad (B16)$$

The second derivatives of J_{stat} are then approximated as:

$$\begin{aligned} \frac{\partial^2 J_{stat}}{\partial \theta_i \partial \theta_j} \sim & \frac{2}{\sigma_x^4 N_{sim}^2} \left[\sum_{t=1}^{N_{sim}} \frac{\partial \tilde{x}(t)}{\partial \theta_i} \right] \left[\sum_{t'=1}^{N_{sim}} \frac{\partial \tilde{x}(t')}{\partial \theta_j} \right] + \frac{2b_1}{\tau_{max} \sigma_x^4} \sum_{|\tau| \leq \tau_{max}} \lambda(\tau)^2 \frac{\partial C_{sim}(\tau)}{\partial \theta_i} \frac{\partial C_{sim}(\tau)}{\partial \theta_j} \\ & + \frac{2b_2}{\tau_{max}^2 \sigma_x^6} \sum_{|\tau_1|, |\tau_2| \leq \tau_{max}} \lambda(\tau_1, \tau_2)^2 \frac{\partial B_{sim}(\tau_1, \tau_2)}{\partial \theta_i} \frac{\partial B_{sim}(\tau_1, \tau_2)}{\partial \theta_j} \end{aligned} \quad (B17)$$

The same formula applies for mixed derivatives with respect to σ_w .

Appendix C: Symbols and acronyms (by order of appearance)

x_{raw} =	Raw time series
x_{cent} =	Detrended and centered time series
$\sigma(x)$ $sk(x)$ $ekurt(x)$ =	Standard deviation, skewness and excess kurtosis of signal x
$x(t)$ =	Standardized time series
$s(t)$, $f(t)$ =	Slow and fast components of the standardized time series
N =	Time series size
$E(\cdot)$ =	Expectation operator
N_{eff} =	Number of effective degrees of freedom
$C_x(\tau)$ =	Empirical autocovariance
$X_x(f) \equiv A_x(f) e^{i\Theta_x(f)}$ =	DFT of the signal $x(t)$ and its polar form
$S_x(f)$ =	Periodogram
$\Gamma_x(f)$ =	Exact spectrum
$\gamma_x(\tau)$ =	Exact autocovariance
$\hat{S}_x(f)$ =	Smoothed spectrum
$\lambda(\tau)$ =	Window function
$\hat{\lambda}(u)$ =	Standardized window function
M =	1D window length
b_1 =	Standardized 1D bandwidth
ν =	Degrees of freedom of the qui-square
$q_p(\cdot)$ =	Quantile por probability p and PDF $(\cdot)$
$w(t)$ =	Zero centered white noise (not necessarily Gaussian nor purely random)
σ_w =	Standard deviation of the noise
$\sigma_{w,p}^2$ =	Noise variance from fitting data with an AR(p) model
$\gamma_x(\tau_1, \tau_2)$ =	Exact bicovariance
$C_x(\tau_1, \tau_2)$ =	Empirical bicovariance
Δ =	Forecast lag in unit time steps
$x_{nl}(t)$ =	Nonlinear component of $x(t)$
TCox Δ	Cox test for lag Δ and quantile 95% under the linearity hypothesis
TCox – 95 %	
$\delta(\cdot)$	Kronecker delta

$\Gamma_{3,w}(f_1, f_2) =$	Exact bispectrum
$\Gamma_{3,y}(f_1, f_2) =$	Normalized exact bispectrum
$S_{3,x}(f_1, f_2) =$	Bi-periodogram
$\Lambda(\tau_1, \tau_2) =$	2D window function used in the smoothed bispectrum
$\hat{S}_{3,x} =$	Smoothed bispectrum
$\hat{S}_{3n,x} =$	Smoothed normalized bispectrum
$\delta\hat{S}_{3,x} =$	Error of the smoothed bispectrum estimator
$var_{ar} =$	Total smoothed bispectrum error variance under asymptotic conditions
$\overline{(\cdot)} =$	Average over the bifrequency domain
$\sigma_{\hat{S}_{3,x}} =$	Typical bispectrum fluctuations
$\tilde{x}(t) =$	Null hypothesis model process
$H_0 =$	Null hypothesis or the null hypothesis model
$\Phi_{\cdot} =$	Standard Gaussian quantile function
$T_{x,\tilde{x}}(f_1, f_2) =$	Standardized bispectrum difference between a model x and $\tilde{x}$
$var_{rat}(\hat{S}_{3,\tilde{x}}) =$	Ratio between empirical and total asymptotic bispectrum variance
$T_{int\ x,\tilde{x}} =$	Sum of squared amplitudes of $T_{x,\tilde{x}}$ over a certain set L , used to test H_0
$\Delta f =$	Frequency step used in the lattice set L
$\varepsilon_L =$	Minimum threshold fraction of the asymptotic error variances accepted in L
$n_L =$	Cardinal of L
$A(f) =$	Frequency function intervening in the AR model
$f_{cut} =$	Cutoff frequency
$\mathbf{Y}_{\tilde{x}} =$	Diagnostic of non-Gaussianity of bispectrum errors
$\tilde{x}(t) =$	Simulated state at time t
$u(t) =$	Driving noise of a simulating model
$\tilde{\mathbf{x}}(t-1) =$	Vector of past simulated states
$\mathbf{u}(t-1) =$	Vector of past driving noises or innovations
$J_{fit}(\boldsymbol{\theta}) =$	Sum of the squared residuals as function of parameters in vector $\boldsymbol{\theta}$
$\varepsilon(t) =$	forecasting error at lag 1
$\boldsymbol{\varepsilon}(t-1) =$	Vector of past forecasting errors
N_{past}	Size of the trail sector of a simulating time series
$N_{\gamma} =$	Maximum lag used in a simulating model
$N_{sim} =$	Size of the sector of a simulating time series, used to compute statistics
$S_{obs} \quad S_{sim} =$	Averages from the observed (obs) and simulated (sim) time series
$C_{obs}(\tau) \quad C_{sim}(\tau) =$	Noncentered lagged moments of second order for obs. and sim. data
Bobs (T1, T2)	Noncentered lagged moments of third order for obs. and sim.
Bsim (T1, T2)	
$J_S, J_C, J_B =$	Normalized squared deviation between the obs. and sim. statistics: average, lagged moments of second and third order respectively
$J_{stat} =$	Weighted sum of J_S, J_C, J_B assessing the average deviations between obs. and sim. statistics.
$J_{hyb} =$	Weighted sum of J_{fit} and J_{stat} with weights c_{fit} and c_{stat} , to minimize in the hybrid fitting
$cor_{lag} =$	Correlation skill between observations and predictions at a certain lag
cpt	Cycles per trimester
PD	Principal domain, i.e. the triangle of vertices (0,0), (0,1/2) and (1/3,1/3) in the bifrequency domain.
NGAR(5)	Non-Gaussian autoregressive model or order 5, used to build the null hypothesis bispectrum
NAIC	Normalized Akaike Information Criterion