Nonlin. Processes Geophys. Discuss., 1, 1539–1602, 2014
www.nonlin-processes-geophys-discuss.net/1/1539/2014/
doi:10.5194/npgd-1-1539-2014

Nonlinear Processes in Geophysics
Discussions
Open Access

Non-Gaussian interaction information: estimation, optimization and diagnostic application of triadic wave resonance

C. A. L. Pires[1,*] **and R. A. P. Perdigão**[2]

[1]Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal
[2]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

Received: 2 July 2014 – Accepted: 9 September 2014 – Published: 2 October 2014

Correspondence to: C. A. L. Pires (clpires@fc.ul.pt)

Published by Copernicus Publications on behalf of the European Geosciences Union & the American Geophysical Union.

Abstract

Non-Gaussian multivariate probability distributions, derived from climate and geofluid statistics, allow for nonlinear correlations between linearly uncorrelated components, due to joint Shannon negentropies. Triadic statistical dependence under pair-wise (total or partial) independence is thus possible. Synergy or interaction information among triads is estimated. We formulate an optimization method of triads in the space of orthogonal rotations of normalized principal components, relying on the maximization of third-order cross cumulants. Its application to a minimal one-dimensional, periodic, advective model, leads to enhanced triads that occur between oscillating components of circular or locally confined wave-trains satisfying the triadic wave resonance condition.

1 Introduction

Climate and geofluid dynamics often display highly complex and nonlinear interactions between system variables, expressed on a physical or spectral basis. Existence of interactions mean that continuous or discrete state vector components statistically constrain the values and decrease the freedom of other components by some physical process with exchanges of extensive variables between system partitions (e.g. mass, energy, momentum), or by impulse propagation (e.g. waves, solitons) or by other more abstract processes. Statistical constraints are synonym of linear and nonlinear correlations and Bayesian-conditional probabilities among variables, which can occur either at short or at long distances, i.e. teleconnections (Wallace and Gutzler, 1981; Van den Dool et al., 2000) where short/long is defined according to a given metrics defined on the state vector. Examples of that in the spectral space are the kinetic energy turbulent cascade (Batchelor, 1953) and the wave resonant interactions (Hammack, 1993) or in the physical space the non-linear El Niño teleconnections (Monahan, 2001; Hsieh et al., 2006). Information Theory (Shannon, 1948; Cover and Thomas,

1991) proposes a consistent quantitative statistical measure (in *bits* or *nats*) of those constraints by assessing the mutual information (MI) and the MI multivariate version, the multi-information (MII) or total correlation (Schneidman et al., 2003) that is shared among variables. The information flow is transferred across a certain channel, which is interpreted as an interaction process (Liang and Kleeman, 2007). Under that information theory framework, the absence of interaction means the statistical independence among variables and a vanishing MII. For dynamical systems, time-extending interactions are measured by the concept of Granger causality (Granger, 1969) by computing MIIs in the delayed coordinate phase space. Beyond the referred papers, there are many applications of Information Theory in nonlinear geophysics, such as in predictability (DelSole, 2005; Abramov et al., 2005), in stochastic theory of multiscale systems (Majda et al., 2005), in ensemble forecasting (Roulston and Smith, 2002) and data assimilation (Xu, 2007; Zupanski et al., 2007).

Let us consider the system state vector as a N-dimensional random vector (RVec) $\boldsymbol{X} \equiv (X_1, \ldots, X_N)'$ of N continuous- (or discrete-) valued scalar variables (in italic) where prime denotes transpose. The components can also be interpreted as nodes of a complex network (Donges et al., 2009) where multiple interactions lie. It is often useful to substitute $\boldsymbol{X}$ by vectors of lower dimension, that: (a) explain much of total variance or total Shannon entropy and (b) have no trivial redundancies between components (e.g. by being linearly uncorrelated), thus making the source separation of the signal. That is possible for continuous-valued RVecs by use of Principal component Analysis (PCA) (Hannachi et al., 2007), Independent Space Analysis (ICA) (Hyvärinen and Oja, 2000), Independent Space Analysis (ISA) (Theis, 2005, 2006; Pires, 2014) and other Blind Source Separation (BSS) methods. However, many of those techniques perform sub-optimally because the issuing variables or factors (e.g. the PCs from a PCA) still have some statistical dependence (non-null MII). This occurs for the vector of PCs derived from a RVec with a non-Gaussian joint probability density function (PDF). Although PCs are linearly uncorrelated, they can exhibit nonlinear correlations (Pearson correlations between functions not written as linear affine combinations of components). Non-

Gaussianity (NGty) is evident from climatic observations and model runs through the presence of asymmetries, multimodalities, clusters and kurtoses of both marginal and multiple-projected PDFs (Hannachi and O'Neill, 2001; Sura et al., 2005; Franzke and Majda, 2007; Pires and Perdigão, 2007). Non-Gaussian (NG) sources include the non-linear drifting as well as the random non-Gaussian and non-stationary forcings of the dynamical equations and the deterministic chaos.

The papers motivation is the fact that the non-Gaussian world (of NG PDFs) allows for much more complex, including multiple (triadic, quartic etc.) interactions that are not explained by linear correlations or second-order statistical moments alone since uncorrelatedness and joint Gaussianity imply global independence.

In the paper, we will focus on sets of three random variables (triads). In fact, three uncorrelated non-Gaussian variables (Y_1, Y_2, Y_3) (non-Gaussian triad) may produce non-null MII values due to nonlinear correlations between functions mixing the three variables. An even more constrained situation is possible, supposing three variables that, while being pair-wise independent (zero dyadic MIs), are not independent as a trivariate whole. In this case, the variables form a perfect triad, due to their interaction synergy leading to an emergent cooperative phenomenon (e.g. a triangle of lovers who meet together only in pairs and in an equitable way, ensuring no preference of any couple). That synergy is measured by the interaction information (IT), which is defined for a triad as $I_t(Y_1, Y_2, Y_3) \equiv I(Y_1, Y_2, Y_3) - I(Y_1, Y_2) - I(Y_1, Y_3) - I(Y_2, Y_3)$, in which $I(\ldots)$ is MII (McGill, 1954; Tsujishita, 1995). IT is the MII remaining term that one obtains after discarding all the information coming from proper subsets of variables (here, the dyads).

IT can even be negative in case of statistical redundancy (e.g. when interactions come into play, e.g. as the aforementioned love triangle moves from pair-wise meetings to a threesome). The IT concept is generalized to any number of variables (Jakulin and Bratko, 2004) and applied to many domains (e.g. economy, sociology, management, medicine, neurology, sampling design), as one tries to find, from an available pool of candidates, playing factors that maximize a given criterion of useful interactivity

(e.g. efficiency) (Timme et al., 2013) or that are responsible for an identified situation (e.g. genes explaining a disease). Both MII and IT can be split into the Gaussian and the non-Gaussian terms (Pires and Perdigão, 2012, 2013), coming respectively from the best Gaussian joint PDF fit and from joint non-Gaussianity (NGty). The NG terms dominate when used variables are uncorrelated.

Now, we present conditions under which IT is maximized. For that, let us consider the partition of each triad variable (Y_1, Y_2, Y_3) into N_c equiprobable categories or symbols. Then, IT is maximized when the bulk of the total probability occurs in N_c^2 out of the N_c^3 disjoint events, which are described by a Latin-square look-up table (e.g. Sudoku game for $N_c = 9$). An example, to be put in evidence in practice holds when $N_c = 2$ and categories $(Y_{1,c}, Y_{2,c}, Y_{3,c})$ satisfy a Boolean exclusive disjunction or a logic equivalence, hereby denoted respectively by $Y_{3,c} = Y_{1,c} \oplus Y_{2,c}$ and $Y_{3,c} = Y_{1,c} \leftrightarrow Y_{2,c}$.

The conditions underneath high IT values are difficult by construction, since dyadic MIs have to be minimized or even vanished, which can be quite rare or difficult to find in nature. That is the case of high-dimensional systems (many degrees of freedom), where multiple quasi-independent interactions (e.g. the brain's network of neurons) tend to produce Gaussian PDFs by invoking the Central Limit Theorem or when variable's time-averages are taken for describing the system's low frequency variability (LFV). Therefore, NGty, nonlinear correlations and multiple interactions are somehow hidden in the high-dimensional space.

In order to get highly synergetic triads, we apply a method relying on the projection pursuit (PP) (Friedman and Tukey, 1975) technique, aiming to maximize NGty or the Negentropy (Shannon entropy deficit with respect to the best Gaussian joint PDF fit). Therefore, first the variables are changed into a vector of uncorrelated variables of unit isotropic variances (e.g. the normalized PCs), thus preserving uncorrelatedness under orthogonal rotations. Then, the IT maximum is sought among three-dimensional (3-D) projections of orthogonally rotated standardized PCs taken from a subset in accordance to a given criterion (e.g. those PCs which maximize IT or the PCs of leading variance). However, IT is a very difficult functional, depending on the

full $\boldsymbol{X}$ probability density function (PDF) whose estimation is quite difficult due to the "curse of dimensionality" (CC) (Bellman, 1957). Therefore, in order to overtake the CC effect (Bocquet et al., 2010), we maximize, by gradient-descent methods (Gilbert and Lemaréchal, 1989) an IT proxy functional written in terms of the rotation matrix or equivalently in terms of the set of generalized Euler angles spanning the space of all orthogonal rotations (Raffenetti and Ruedenberg, 1969; Hoffman et al., 1972). The functional, to be used and maximized consists of a truncated version of the IT Edgeworth expansion, relying on cross cumulants of order higher-than-three (Comon, 1994). The functional vanishes under joint Gaussian conditions. For the sake of geometric interpretation, it writes as the square of a three-order joint cumulant $E(Y_1 Y_2 Y_3)$ (E stands by expectation), between the unit-variance, uncorrelated, centered components of the triad. Moreover, it is proportional to a triadic correlation, denoted as $\mathrm{cor}_3(Y_1, Y_2, Y_3)$, consistently satisfying the Schwarz inequality. That cumulant writes as a third order polynomial form of rotation matrix coefficients and as a linear form of three-order joint cumulants of the PCs.

For the application of the IT concept and its optimization in the rotations space, we have estimated the triadic IT applied to PCs of the output of the 40-variable Lorenz model (1995), hereafter denoted called Lorenz-95. The 40 variables are in fact longitudinally equidistant nodes in a latitude circle, along which a one-dimensional flow is set to take place. This is thus a minimal one-dimensional (1-D) model of fluid motion with nonlinear advection, linear dissipation and constant forcing, integrated along a latitude circle with circular periodic conditions. Circular symmetry of statistics leads to twice-degenerated sinusoidal empirical orthogonal functions (EOFs), each one with an associated wave-number. Therefore, PCs are Discrete Fourier Transforms (DFTs). We show that non-null values of the moment $E(Y_1 Y_2 Y_3)$, between normalized PCs and of the estimated IT: $I_t(Y_1, Y_2, Y_3)$, are obtained when the integer zonal wave-numbers k_1, k_2, k_3 satisfy the wave resonance condition (WRC): $k_1 \pm k_2 \pm k_3 = 0$ (Bartello, 1995; Hammack, 1993). It consists of the build-up of a third wave from the quasi-linear interaction of two weak amplitude waves satisfying the dispersion relationship.

Optimized triads following the aforementioned method are also computed. They correspond to locally confined wave-trains satisfying the WRC for the number of local oscillations. Therefore, the triad optimization can be characterized as a generalized WR diagnostic under the studied model's framework.

After this introduction, the paper presents the concepts of multi-information and interaction information in Sect. 2, followed by the optimization implementation of triads (Sect. 3). The application to the Lorenz-95 model appears in Sect. 4 ending with the Conclusion (Sect. 5). Some proofs and MII estimators are included in the Appendices A and B.

2 Multi-information and interaction information

2.1 General context

Without loss of generality, let $\boldsymbol{X} \equiv (X_1, \ldots, X_N)'$ be a zero average N-dimensional physical RVec, composed of components X_i, $i = 1, \ldots, N$ in italic, and $\mathbf{C}_{\boldsymbol{X},\boldsymbol{X}}$, $\mathbf{C}_{1,\boldsymbol{X},\boldsymbol{X}}$ respectively the covariance and correlation matrices of $\boldsymbol{X}$. The PCA of $\boldsymbol{X}$ comes from the factorization $\mathbf{C}_{\boldsymbol{X},\boldsymbol{X}} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}'$, where $\mathbf{W}$ is the orthogonal matrix of normalized EOFs in columns, satisfying $\mathbf{W}^{-1} = \mathbf{W}'$ and $\boldsymbol{\Lambda} \equiv \mathbf{Diag}(\lambda_1, \ldots, \lambda_N)$ is the diagonal matrix sorted by decreasing PC variances, i.e. $\lambda_1 \geq \ldots \geq \lambda_N$. The vector of PCs and of the standardized (unit-variance) PCs are respectively given by $\boldsymbol{X}_{\mathrm{PC}} \equiv \mathbf{W}^{\mathrm{T}}\boldsymbol{X}$ and $\boldsymbol{X}_{\mathrm{u}} \equiv \boldsymbol{\Lambda}^{-1/2}\boldsymbol{X}_{\mathrm{PC}}$. The PCs are uncorrelated though they can be statistically dependent, i.e. they interact if they have non-vanishing nonlinear correlations.

The goal of the paper is to find small subsets of $D \ll N$ rotated uncorrelated components such that: (a) they explain a relevant fraction of the total variance, (b) they exhibit high nonlinear multivariate correlations with a simple geometric and physical interpretation in terms of the dynamics governing $\boldsymbol{X}$, (c) the correlations emerge from synergies of the D components, hence not being explained by proper subsets. Without loss of generality, we restrain rotations to a set of normalized PCs chosen according

to a statistical or physical criterion (e.g. a set of variance-explained leading PCs), by writing $\boldsymbol{X}_{\mathrm{u}}$ as a concatenation or a union set of components or still a projection (depending on the context) written in the form $\boldsymbol{X}_{\mathrm{u}} = \boldsymbol{X}_{\mathrm{u,\,signal}} \| \boldsymbol{X}_{\mathrm{u,\,noise}}$, where $\boldsymbol{X}_{\mathrm{u,\,signal}}$ is the part to be rotated (the signal) of dimension $N_{\mathrm{rot}} \leq N$, corresponding to a certain set S of PC indices and $\boldsymbol{X}_{\mathrm{u,\,noise}}$ is the remaining projection (the generalized noise), in particular the null set. We can write consistently $\boldsymbol{X}_{\mathrm{u,\,signal}} = \cup_{i \in S}\{X_{\mathrm{u},i}\}$, where the union symbol has the meaning of concatenation.

Therefore, the rotated vector is $\boldsymbol{Y} = \boldsymbol{Y}_{\mathrm{rot}} \| \boldsymbol{X}_{\mathrm{u,\,noise}}$ with $\boldsymbol{Y}_{\mathrm{rot}} \equiv \mathbf{R}\boldsymbol{X}_{\mathrm{u,\,signal}}$, where $\mathbf{R}$ is a N_{rot}-order orthogonal matrix i.e. $\mathbf{R}' = \mathbf{R}^{-1}$. The components of $\boldsymbol{Y}_{\mathrm{rot}} \equiv (Y_1, \ldots, Y_{N_{\mathrm{rot}}})'$ are uncorrelated given by inner products $Y_i = \boldsymbol{v}_i' \boldsymbol{X}_{\mathrm{u,\,signal}}$ where $\boldsymbol{v}_i \equiv (R_{i1}, R_{i2}, \ldots, R_{iN_{\mathrm{rot}}})'$; $i = 1, \ldots, N_{\mathrm{rot}}$ are ortho-normalized vectors satisfying to $\boldsymbol{v}_i' \boldsymbol{v}_j = \delta_{ij}$, $i, j = 1, \ldots, N_{\mathrm{rot}}$ where δ_{ij} is the Kronecker Delta. Finally, let us write $\boldsymbol{Y}_{\mathrm{rot}} \equiv \boldsymbol{Y}_{\mathrm{proj}} \| \boldsymbol{Y}_{\mathrm{comp}}$ where $\boldsymbol{Y}_{\mathrm{proj}} \equiv \boldsymbol{P}_D \boldsymbol{Y}_{\mathrm{rot}} = (Y_1, \ldots, Y_D)'$ is the D-dimensional projection (D-tuple) onto the leading D components and $\boldsymbol{Y}_{\mathrm{comp}} \equiv (Y_{D+1}, \ldots, Y_{N_{\mathrm{rot}}})'$ is the corresponding orthogonal projection.

The components Y_k, $k = 1, \ldots, N_{\mathrm{rot}}$ are a linear combination of loadings multiplying the physical component perturbations in the form

$$Y_k = \sum_{i=1}^{N} W_{Y_i,k} X_i; \quad k = 1, \ldots, N_{\mathrm{rot}}; \tag{1a}$$

$$W_{Y_i,k} = \sum_{j \in S} \frac{1}{\lambda_j^{1/2}} R_{k,j} W_{i,j}; \quad k = 1, \ldots, N_{\mathrm{rot}} \tag{1b}$$

The loading vectors $(W_{Y_1,k}, \ldots, W_{Y_N,k})'$ are in general not orthogonal in the physical space unless the rotation is composed of partial rotations over different subsets of PCs of degenerated (equal) variance in each subset. We will focus in particular on triads ($D = 3$) of components (Y_1, Y_2, Y_3) of maximum interactivity. They are optimized among

the restrictions of the state vector to a certain signal projection and among the set of orthogonal rotations **R**.

2.2 Basic information-theory definitions

The assessment of statistical interactivity due to linear and nonlinear correlations rely on concepts of information theory (Shannon, 1948) and measures of non-Gaussianity. Given its relevance for the paper, we briefly present some definitions and notations, valid for any N-dimensional RVec $\boldsymbol{A} \equiv (A_1, \ldots, A_N)'$, in particular for $\boldsymbol{X}$ or $\boldsymbol{Y}$ defined above.

Definition 1: the "Gaussianized" vector $\boldsymbol{A}'_{\mathrm{g}} \equiv (A_{1\mathrm{g}}, \ldots, A_{N\mathrm{g}})'$ is the N-dimensional RVec composed of marginal Gaussian anamorphoses (Wackernagel, 1995) of each $\boldsymbol{A}$ component, $A_k \rightarrow A_{k\mathrm{g}} = G(A_k) \sim N(0,1)$, $k = 1, \ldots, N$, which are changed into standard Gaussian distributed variables of zero mean and unit variance.

Definition 2: the Shannon differential entropy (SDE) (Shannon, 1948) of $\boldsymbol{A}$ is $H(\boldsymbol{A}) \equiv -E(\log \rho_{\boldsymbol{A}}) = -\int \rho_{\boldsymbol{A}} \log \rho_{\boldsymbol{A}} \mathrm{d}\boldsymbol{A}$, where $\rho_{\boldsymbol{A}}$ is the PDF of $\boldsymbol{A}$. The entropy $H(\boldsymbol{A})$ under statistical independence of the components is $H_i(\boldsymbol{A}) \equiv \sum_{k=1}^{N} H(A_k)$, i.e. the sum of marginal SDEs. The SDE of the multivariate Gaussian PDF with the same average and covariance matrix as $\boldsymbol{A}$ writes: $H_{\mathrm{g}}(\boldsymbol{A}) \equiv H_{\mathrm{g}i}(\boldsymbol{A}) - F_{\mathrm{g}}(\boldsymbol{A})$, where $H_{\mathrm{g}i}(\boldsymbol{A}) \equiv N h_{\mathrm{g}} + 1/2 \sum_{k=1}^{N} \log[\mathrm{var}(A_k)]$ is $H_{\mathrm{g}}(\boldsymbol{A})$ for independent components, $h_{\mathrm{g}} = 1/2 \log(2\pi e)$ is the Shannon entropy of the standard Gaussian with PDF $\varphi(X) \equiv (2\pi)^{-1/2} \exp(-1/2 X^2)$ and $F_{\mathrm{g}}(\boldsymbol{A}) \equiv -1/2 \log(\det \mathbf{C}_{1,\boldsymbol{A},\boldsymbol{A}}) \geq 0$, dependent on the correlation matrix $\mathbf{C}_{1,\boldsymbol{A},\boldsymbol{A}}$, with $F_{\mathrm{g}}(\boldsymbol{A})$ being zero for a scalar or for uncorrelated components.

Definition 3: the negentropy of $\boldsymbol{A}$ is a non-negative quantity $J(\boldsymbol{A}) \equiv H_{\mathrm{g}}(\boldsymbol{A}) - H(\boldsymbol{A}) \geq 0$, being the deficit of SDE with respect to the Gaussian PDF with the same

average and covariance matrix. The negentropy for independent components is $J_i(\boldsymbol{A}) \equiv H_{\mathrm{g}i}(\boldsymbol{A}) - H_i(\boldsymbol{A}) \geq 0$, being the sum of marginal negentropies. J is invariant for any invertible linear affine transformation of $\boldsymbol{A}$. For $\boldsymbol{A}_{\mathrm{g}}$ we get $J_i(\boldsymbol{A}_{\mathrm{g}}) = 0$ despite having $J(\boldsymbol{A}_{\mathrm{g}}) \geq 0$.

Definition 4: the multi-information (MII) (Schneidman et al., 2003) or total correlation of $\boldsymbol{A}$ is a non-negative quantity given by

$$I_i(\boldsymbol{A}) \equiv H_i(\boldsymbol{A}) - H(\boldsymbol{A}) \tag{2a}$$

$$\equiv I(A_1, A_2, \ldots, A_N) \tag{2b}$$

$$\equiv E\{\log[\rho_{\boldsymbol{A}} / (\rho_{A_1} \rho_{A_2} \cdots \rho_{A_N})]\} \geq 0, \tag{2c}$$

measuring the full statistical dependence among components of $\boldsymbol{A}$. For $N = 2$, we recover the mutual information (MI) $I(A_1, A_2) = \int \rho_{A_1,A_2} \log[\rho_{A_1,A_2} / (\rho_{A_1} \rho_{A_2})] \mathrm{d}A_1 \mathrm{d}A_2$. The subscript i in I_i is omitted when all components are explicitly written.

MII is invariant for scalar one-to-one changes of components, in particular for $\boldsymbol{A}_{\mathrm{g}}$, from which we obtain $I_i(\boldsymbol{A}) = I_i(\boldsymbol{A}_{\mathrm{g}})$. This allows for decomposing the MII into two consistent positive terms: $I_i(\boldsymbol{A}) = I_{i\mathrm{g}}(\boldsymbol{A}) + I_{i\mathrm{ng}}(\boldsymbol{A})$ (Pires and Perdigão, 2012, 2013), respectively the Gaussian and the non-Gaussian MIIs, where $I_{i\mathrm{g}}(\boldsymbol{A}) = I_{i\mathrm{g}}(\boldsymbol{A}_{\mathrm{g}}) \equiv F_{\mathrm{g}}(\boldsymbol{A}_{\mathrm{g}}) \geq 0$, is uniquely dependent on "Gaussian" correlations, or correlations between "Gaussianized" components, and $I_{i\mathrm{ng}}(\boldsymbol{A}) = I_{i\mathrm{ng}}(\boldsymbol{A}_{\mathrm{g}}) = J(\boldsymbol{A}_{\mathrm{g}}) \geq 0$ is the part of MII due the joint negentropy of the PDF with Gaussian marginals. Each term $I_{i\mathrm{g}}$ and $I_{i\mathrm{ng}}$ is invariant for monotonic, (though not for non-monotonic in general) changes of marginals. Thanks to the Minimum Mutual Information Principle (Globerson et al., 2009), $I_{i\mathrm{g}}$ is the minimum MII, under imposed Gaussian marginals and a correlation matrix $\mathbf{C}_{1,\boldsymbol{A}_{\mathrm{g}},\boldsymbol{A}_{\mathrm{g}}}$ between Gaussianized components.

2.3 Multi-Information of a projected and rotated vector

In what follows, we present some results about the multi-information and non-Gaussianity crossing components of the sought projection $\boldsymbol{Y}_{\text{proj}}$, introduced before, and how it depends on the rotated vector $\boldsymbol{Y}_{\text{rot}}$.

Theorem 1: the MII of $\boldsymbol{Y}_{\text{proj}}$ expands as

$$I_i(\boldsymbol{Y}_{\text{proj}}) = J_i(\boldsymbol{X}_{\text{u, signal}}) + I_i(\boldsymbol{X}_{\text{u, signal}}) - [J_i(\boldsymbol{Y}_{\text{proj}}) + J_i(\boldsymbol{Y}_{\text{comp}}) + I_i(\boldsymbol{Y}_{\text{comp}}) + I(\boldsymbol{Y}_{\text{proj}}, \boldsymbol{Y}_{\text{comp}})]. \tag{3}$$

The proof comes in Appendix A. Therefore, from Eq. (3), the MII of $\boldsymbol{Y}_{\text{proj}}$ varies in the opposite sense of: (a) the marginal negentropies of rotated components (projected ones and their orthogonal complement), (b) the MII of the non-projected variables composing $\boldsymbol{Y}_{\text{comp}}$ and (c) the mutual information between the projected and the non-projected parts: $I(\boldsymbol{Y}_{\text{proj}}, \boldsymbol{Y}_{\text{comp}}) \equiv \int \rho_{\boldsymbol{Y}_{\text{rot}}} \log[\rho_{\boldsymbol{Y}_{\text{rot}}}/(\rho_{\boldsymbol{Y}_{\text{proj}}}\rho_{\boldsymbol{Y}_{\text{comp}}})]\mathrm{d}\boldsymbol{Y}_{\text{proj}}\mathrm{d}\boldsymbol{Y}_{\text{comp}}$. Since the MII is invariant for one-to-one changes of the intervening variables, in particular rotations, that is understood as the MI between the spanned subspaces span($\boldsymbol{Y}_{\text{proj}}$) and span($\boldsymbol{Y}_{\text{comp}}$). In particular, if $\boldsymbol{Y}_{\text{proj}} = \boldsymbol{Y}_{\text{rot}}$ i.e. is a full projection ($D = N$), the terms within squared brackets of Eq. (3) reduce to $J_i(\boldsymbol{Y}_{\text{rot}})$, under which the MII of $\boldsymbol{Y}_{\text{proj}}$ is minimized when the sum of marginal negentropies is maximum. That is the purpose of the Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000; Hastie et al., 2001).

The goal here is simply the opposite of ICA's, by maximizing $I_i(\boldsymbol{Y}_{\text{proj}})$ or more precisely the part of it dealing with synergies among components. This is discussed in the next section.

The following Theorem 2 shows the dominance of MII $I_i(\boldsymbol{Y}_{\text{proj}})$ by its non-Gaussian contribution. Since components of $\boldsymbol{Y}_{\text{rot}}$ are uncorrelated, one expects their Gaussian absolute correlations $c_{\text{g}i,j} \equiv |\text{cor}(Y_{i\text{g}}, Y_{j\text{g}})| = O(\delta) \ll 1$, $j \neq i \in \{1, \ldots, D\}$ to be a small residual coming from the static Gaussian anamorphoses. Therefore, under those

conditions one has

Theorem 2: $I_i(\boldsymbol{Y}_{\text{proj}})$ is expanded as

$$I_i(\boldsymbol{Y}_{\text{proj}}) = I_{i\text{g}}(\boldsymbol{Y}_{\text{proj}}) + I_{i\text{ng}}(\boldsymbol{Y}_{\text{proj}}); \tag{4a}$$

$$I_{i\text{g}}(\boldsymbol{Y}_{\text{proj}}) = O(\delta^2) \ll 1; \tag{4b}$$

$$I_{i\text{ng}}(\boldsymbol{Y}_{\text{proj}}) = J(\boldsymbol{Y}_{\text{proj, g}}). \tag{4c}$$

The scaling of $I_{i\text{g}}(\boldsymbol{Y}_{\text{proj}})$ comes from $\det(\boldsymbol{Y}_{\text{proj, g}}) = 1 - O(\delta^2)$. Equations (4a)–(4c) show that the non-Gaussian MII is dominant in general when components are uncorrelated, expressed as the joint negentropy of $\boldsymbol{Y}_{\text{proj, g}}$, also coinciding in this case to its compactness (Monahan and DelSole, 2009). That measures the level "featureness" of data concentration around a lower-dimensional manifold, different from a unit-sphere D-dimensional ball, corresponding to an isotropic Gaussian at which $J(\boldsymbol{Y}_{\text{proj, g}}) = 0$.

2.4 Interaction information in general

The components of $\boldsymbol{Y}_{\text{proj}}$ interact in a certain sense (e.g. physically), therefore they are statistically constrained with the amount of constraint being measured by information or the MII. Those interactions can emerge from proper subsets of components like local interactions within a complex network or may consist in global or synergetic emerging properties entailed by all members (components) as a whole like a collective cooperative phenomenon. That decomposition is accounted for by the concept of interaction information (IT) (McGill, 1954; Tsujishita, 1995; Jakulin and Bratko, 2004), which we present here for the generic N-dimensional RVec $\boldsymbol{A}$ introduced in Sect. 2.2.

The IT of $\boldsymbol{A}$, hereby denoted by $I_{\text{t}}(\boldsymbol{A})$ (where the subscript t distinguishes it from I_i), is the part of $I_i(\boldsymbol{A})$ not accounted for by any proper subset of components of $\boldsymbol{A}$, i.e. by any subset $\boldsymbol{B} \subset \boldsymbol{A}$. In particular for $\boldsymbol{A} = (A_1, A_2)'$, IT coincides with the MI i.e. $I_{\text{t}}(A_1, A_2) = I(A_1, A_2) \geq 0$. The IT is built according to the recurrence property

$$I_{\text{t}}(\boldsymbol{A}) = I_{\text{t}}(\boldsymbol{B}|C) - I_{\text{t}}(\boldsymbol{B}), \forall \boldsymbol{B}, C : \boldsymbol{A} = \boldsymbol{B} \cup \{C\},\ C \notin \boldsymbol{B}, \tag{5}$$

where $\boldsymbol{A}$ is $\boldsymbol{B}$ increased by a scalar component C and $I_t(\boldsymbol{B}|C)$ is the conditional IT using the conditional PDF $\rho_{\boldsymbol{B}|C}$ in its definition. Jakulin and Bratko (2003) show that I_t is given by adequate linear combinations of Shannon entropies and MIIs of the $\boldsymbol{A}$ subsets as

$$I_t(\boldsymbol{A}) = -(-1)^N \sum_{k=1}^{N} (-1)^k \sum_{\substack{\boldsymbol{B}\subseteq\boldsymbol{A} \\ |\boldsymbol{B}|=k}} H(\boldsymbol{B}) \tag{6a}$$

$$= (-1)^N \sum_{k=2}^{N} (-1)^k \sum_{\substack{\boldsymbol{B}\subseteq\boldsymbol{A} \\ |\boldsymbol{B}|=k}} I_i(\boldsymbol{B}). \tag{6b}$$

In the sums of Eqs. (6), subsets $\boldsymbol{B}$ of $\boldsymbol{A}$ are arranged according to their cardinality $|\boldsymbol{B}|$. Positive values of IT for $\dim(\boldsymbol{A}) = |\boldsymbol{A}| \geq 3$ mean synergy across components, while negative IT values indicate redundancy (Timme et al., 2013). The Gaussian and non-Gaussian terms from the decomposition of IT, denoted respectively as I_{tg}, I_{tng}, come from its MII decomposition in Eqs. (6), using the Definition 4. Moreover, the MII is the full sum of ITs

$$I_i(\boldsymbol{A}) = \sum_{\boldsymbol{B}\subseteq\boldsymbol{A}} I_t(\boldsymbol{B}), \tag{7}$$

which shows that the sum of synergies and redundancies are globally positively constrained. An important corollary comes from Eqs. (6):

Corollary 1: if $\boldsymbol{A}$ is composed by uncorrelated components, from which $H = H_{gi} - J$, then the IT is a function of negentropies as:

$$I_t(\boldsymbol{A}) = (-1)^N \sum_{k=1}^{N} (-1)^k \sum_{\substack{\boldsymbol{B}\subseteq\boldsymbol{A} \\ |\boldsymbol{B}|=k}} J(\boldsymbol{B}). \tag{8}$$

The Eqs. (6)–(8) hold in particular for $\boldsymbol{A} = \boldsymbol{Y}_{\text{proj}}$ expressing how the synergies depend on MIIs and negentropies of subsets.

2.5 Interaction information between triads and its estimation

We now focus on the IT of triads i.e. of $\boldsymbol{A} = (A_1, A_2, A_3)'$. Considering Eqs. (5) and (6), it is given by:

$$I_t(A_1, A_2, A_3) \equiv I_3(A_1, A_2, A_3) \tag{9a}$$

$$\equiv I_i(A_1, A_2, A_3) - I_2(A_1, A_2, A_3) \tag{9b}$$

$$= I(A_i, A_j|A_k) - I(A_i, A_j) \tag{9c}$$

$$= I[(A_i, A_j), A_k] - I(A_i, A_k) - I(A_j, A_k), \tag{9d}$$

where $I_2(A_1, A_2, A_3) \equiv I(A_1, A_2) + I(A_1, A_3) + I(A_2, A_3)$ is the notation for the sum of all dyadic MIs and (i, j, k) is a generic permutation of $(1,2,3)$. In Eq. (9) the conditional MI depends on $\rho_{A_i,A_j|A_k}/(\rho_{A_i|A_k}\rho_{A_j|A_k})$ whereas $I(A_i, A_j|A_k)$ depends on $\rho_{A_i,A_j,A_k}/(\rho_{A_i,A_j}\rho_{A_k})$. The value of the triadic IT $I_3(A_1, A_2, A_3)$ coincides with MII if variables are pair-wise independent, i.e. $I_2(\boldsymbol{A}) = 0$ (*perfect triad*).

Small values of $I_2(\boldsymbol{A})$ are favored under uncorrelatedness (case of $\boldsymbol{A} = \boldsymbol{Y}_3 \equiv (Y_1, Y_2, Y_3)$), where $I_2(\boldsymbol{Y})$ is dominated by the non-Gaussian dyadic MIs. Synergetic triads occur when components are nonlinearly related while keeping weak or vanishing pair-wised dependence.

The IT I_3 develops also in terms of the Kirkwood superposition approximation $K_{A_1,A_2,A_3} \equiv (\rho_{A_1,A_2}\rho_{A_1,A_3}\rho_{A_2,A_3})/(\rho_{A_1}\rho_{A_2}\rho_{A_3})$ (Pearl, 1988) in the form $I_3(A_1, A_2, A_3) = E\left[\ln\left(\rho_{Y_1,Y_2,Y_3}/K_{A_1,A_2,A_3}\right)\right]$. When the L_1 norm of K_{A_1,A_2,A_3} equals to one, it is reduced to a 3-dimensional PDF that is uniquely based on bivariate and univariate PDFs. Under those conditions, the triadic IT is a Kullback–Leibler divergence (Cover and Thomas, 1991) between two PDFs, which is necessarily non-negative, i.e. $I_3 \geq 0$. It holds if at least one pair of variables is independent (e.g. A_1, A_2 independent, then $K_{A_1,A_2,A_3} =$

$\rho_{A_3}\rho_{A_1|A_3}\rho_{A_2|A_3}$ is a PDF), corresponding to less restrictive non-perfect triads, hereafter called *asymmetric triads*. The asymmetry of the dependende relationships introduces some directional causality, susceptible of physical interpretation. It yields successively less constrained measures of interactivity:

$$I_3 \le I_{3-\mathrm{as}2} \equiv I_3 + \max I_{2\mathrm{set}} \le I_{3-\mathrm{as}1} \tag{10a}$$

$$\equiv I_i - \min I_{2\mathrm{set}} \le I_i \tag{10b}$$

$$I_{2\mathrm{set}} = \{I(A_j, A_k) :\ j,k \in \{1,2,3\},\ j \ne k\}, \tag{10c}$$

where $I_{3-\mathrm{as}2} = I(A_{j'}, A_{k'}|A_{i'})$ and $I_{3-\mathrm{as}1} = I[(A_j, A_k), A_i]$ for some permutations (i', j', k'), (i, j, k) of $(1,2,3)$. The equalities hold when two or one (out of the three) dyadic MIs are null respectively. The four possibilities of triad interactivity are sketched by non-oriented graphs displayed in Fig. 1a–d.

The MIIs and entropies are subjected to several inequalities (Makarychev et al., 2002). That leads to a useful lower bound of I_t. For that, let us write $A_k = f(A_i, A_j) + w$, where f is a nonlinear one-to-one function of A_i for each A_j that is supposedly correlated to A_k and w is a noise. By playing with conditional MIs, we can write

$$I_\mathrm{t} \le I[(A_i, A_j), A_k] \tag{11a}$$

$$\le I(f, A_k) + I(w, A_j|f). \tag{11b}$$

If the noise w is independent of both A_i, A_j, then the second term of the rhs of Eq. (11b) vanishes and $I(f, A_k)$ can be approximated by its Gaussian part, i.e.

$$I(f, A_k) \sim -\frac{1}{2}\log\left\{1 - \mathrm{cor}_\mathrm{g}[f(A_i, A_j), A_k]^2\right\} \tag{12}$$

Perfect triads can be obtained by nonlinear mixing of globally independent variables. For instance, taking independent standard Gaussian RVs $A_1, A_2, W \sim N(0,1)$ and $A_3 = |W|\mathrm{sgn}(A_1 A_2) \sim N(0,1)$ leads to a perfect triad with bivariate Gaussian isotropic PDFs

and the joint non-Gaussian PDF:

$$\rho_{A_1,A_2,A_3}(A_1, A_2, A_3) = \begin{cases} 0 & \text{if } A_1A_2A_3 \le 0 \\ 2\varphi(A_1)\varphi(A_2)\varphi(A_3) & \text{if } A_1A_2A_3 > 0 \end{cases}, \tag{13}$$

with $I_3(A_1, A_2, A_3) = I(A_1, A_2, A_3) = \log 2$. The signals of the RVs satisfy a Boolean-like relationship: $\mathrm{sgn}(A_i) = \mathrm{sgn}(A_j)\mathrm{sgn}(A_k)$ and the correlations between any two and the third variable are $\mathrm{cor}(A_j A_k, A_i) = 8/(2\pi)^{3/2} \sim 0.51$, where i, j, k is any permutation of $(1,2,3)$. That triadic symbolic link leads to the twisting character of the PDF iso-surface $\rho_{A_1,A_2,A_3} = 0.01$ plotted in Fig. 2.

In a perfect triad like (A_1, A_2, A_3) satisfying Eq. (13), the unmixing operator recovering original independent variables $[A_1, A_2, |W| = A_3\mathrm{sgn}(A_1 A_2)]$ is nonlinear, which could only be possible by using nonlinear-ICA (Almeida, 2003, 2005). This means that linear-ICA, using rotated variables, is unable to unfold perfect triads into independent scalars, because the 2-D-manifold where the PDF concentrate has an intrinsic curvature that is preserved by the isometric rotation transformation.

For uncorrelated variables, i.e. $(A_1, A_2, A_3) = (Y_1, Y_2, Y_3)$, both the IT and MII come essentially from the non-Gaussian term, which is justified by:

Theorem 3: assuming that the absolute Gaussian correlations of (Y_1, Y_2, Y_3) are scaled as $|\mathrm{cor}(Y_{i\mathrm{g}}, Y_{j\mathrm{g}})| \equiv c_{\mathrm{g}i,j} = O(\delta) \ll 1$, we get

$$I_3(Y_1, Y_2, Y_3) = I_{3\mathrm{g}}(Y_1, Y_2, Y_3) + I_{3\mathrm{ng}}(Y_1, Y_2, Y_3) \tag{14a}$$

$$I_{3\mathrm{g}}(Y_1, Y_2, Y_3) = -\frac{1}{2}\ln\left[\frac{1 - c_{\mathrm{g}1,2}^2 - c_{\mathrm{g}1,3}^2 - c_{\mathrm{g}2,3}^2 + 2c_{\mathrm{g}1,2}c_{\mathrm{g}1,3}c_{\mathrm{g}2,3}}{\left(1 - c_{\mathrm{g}1,2}^2\right)\left(1 - c_{\mathrm{g}1,3}^2\right)\left(1 - c_{\mathrm{g}2,3}^2\right)}\right] \tag{14b}$$

$$= O(\delta^3) \ll 1. \tag{14c}$$

As a corollary, it comes that jointly Gaussian perfect triads are impossible since uncorrelatedness implies global independence. Consequently, synergetic perfect triads are necessarily jointly non-Gaussian.

Using Eq. (8), the IT between uncorrelated variables writes in terms of joint and marginal negentropies:

$$I_3(Y_1,Y_2,Y_3) = J(Y_1,Y_2,Y_3) - [J(Y_1,Y_2) + J(Y_1,Y_3) + J(Y_2,Y_3)] + J(Y_1) + J(Y_2) + J(Y_3) \quad (15)$$

that is totally due to non-Gaussianity of variables. For uncorrelated and marginally non-Gaussian variables, the Theorem 3 applies and $I_3 \approx I_{3ng}$. In practice, in Sect. 4, the triadic and dyadic MIIs necessary to compute IT will be numerically estimated (see Walters-Williams and Li (2009) for a panoply of MII estimators). For that, we first estimate the joint 2-D- or 3-D-PDFs by a kernel-based estimator and then the MIIs and IT by an integral Gauss quadrature formula with previous Gaussian anamorphoses of marginals (see details in Appendix B along with the MI estimator's bias depending on sample size).

For weak deviations from joint Gaussianity, the negentropies in Eq. (15) can be expanded in terms of self and cross cumulants (Comon, 1994), using the Edgeworth expansion of the joint PDF (Van Hulle, 2005). In particular, the second, third and fourth-order cumulants between centered variables (used here), are respectively

$$\kappa^{ij} \equiv E(Y_i Y_j); \quad (16a)$$

$$\kappa^{ijk} \equiv E(Y_i Y_j Y_k) \quad (16b)$$

$$\kappa^{ijkl} \equiv E(Y_i Y_j Y_k Y_l) - E(Y_i Y_j)E(Y_k Y_l) - E(Y_i Y_k)E(Y_j Y_l) - E(Y_i Y_l)E(Y_j Y_k) \quad (16c)$$

being invariant for any permutation of the indices $(i,j,k,l) \in \{1,2,3\}$. After some tedious algebra, we get a positive fourth-order approximation of I_3, depending on

squares of third- and fourth-order cross cumulants as

$$\begin{aligned} I_3(Y_1,Y_2,Y_3) = {} & (\kappa^{123})^2 \left\{ \frac{1}{2} + \frac{1}{12}[(\kappa^{111})^2 + (\kappa^{222})^2 + (\kappa^{333})^2] \right\} + \frac{1}{12}[(\kappa^{1123})^2 + \\ & (\kappa^{2213})^2 + (\kappa^{3312})^2] + \frac{1}{36}\{(\kappa^{111})^2[(\kappa^{223})^2 + (\kappa^{332})^2] + (\kappa^{222})^2[(\kappa^{113})^2 + (\kappa^{331})^2] + \\ & (\kappa^{333})^2[(\kappa^{112})^2 + (\kappa^{221})^2]\} + \frac{1}{24}[(\kappa^{112}\kappa^{233})^2 + (\kappa^{113}\kappa^{322})^2 + (\kappa^{221}\kappa^{133})^2 + \\ & (\kappa^{223}\kappa^{311})^2 + (\kappa^{331}\kappa^{122})^2 + (\kappa^{332}\kappa^{211})^2] + O\left(N_{\text{dof-Ed}}^{-3/2}\right) \end{aligned} \quad (17)$$

where Y_1, Y_2, Y_3 are supposed to behave as a sum of $N_{\text{dof-Ed}}$ iid. variables, which grows as far as variables are closer to Gaussianity where the Central Limit Theorem is invoked. The irst two big terms of Eq. (17) depends on trivariate cumulants while the hird and fourth big terms involve bivariate cumulants, thus showing that triadic interactivity can be due to combinations of nonlinear dyadic correlations. Equation (17) depends on the mixing cumulant $\kappa^{123} = E(Y_1 Y_2 Y_3)$. Therefore, thanks to the linear pair-wise uncorrelation and variable normalization, that is proportional to $\text{cor}(Y_3, Y_1 Y_2)$, $\text{cor}(Y_2, Y_1 Y_3)$ and $\text{cor}(Y_1, Y_2 Y_3)$. In order to define a quantity that is independent of the particular order of those products, we introduce here a consistent measure of trivariate correlation between centered, normalized and pair-wise uncorrelated variables Y_1, Y_2, Y_3, satisfying the Schwarz inequality. The measure is given by the cubic-root of the product of those three correlations as

$$\text{cor}_3(Y_1,Y_2,Y_3) \equiv \frac{E(Y_1 Y_2 Y_3)}{\left[\text{var}(Y_1 Y_2)\text{var}(Y_1 Y_3)\text{var}(Y_2 Y_3)\right]^{1/6}} \in [-1,1] \quad (18)$$

and will be tested in practice.

Recurring to Eq. (12), we will also test the approximation

$$I_3(Y_1,Y_2,Y_3) \sim -\frac{1}{2}\log\left[1 - \text{cor}_3(Y_1,Y_2,Y_3)^2\right]. \quad (19)$$

The goodness of that is measured by the difference between $|\mathrm{cor}_3|$ and the equivalent correlation triadic correlation, obtained by inverting I_3 from Eq. (19):

$$\mathrm{cor}_{3\mathrm{eq}}(Y_1, Y_2, Y_3) = [1 - \exp(-2I_3(Y_1, Y_2, Y_3)]^{1/2}. \quad (20)$$

2.6 Discrete triadic interactions

Maximum triadic interactivity comes essentially from certain categorical (discrete) deterministic relationships, without the need to consider continuous PDFs. Those relationships can be further enhanced in a coarse-grained description of the joint PDF. For that, we can consider a partition of the support sets of any of the marginal variables into $N_\mathrm{c} \geq 2$ equiprobable subsets, indexed by N_c different symbols (e.g. digits, letters). Subsets are not necessarily connected (e.g. unions of disconnected intervals). The corresponding coarse-graining transformation is

$$Y_i \in \mathbb{R} \rightarrow Y_{i,\mathrm{c}} = F_{i,\mathrm{c}}(Y_i) \in B_{N_\mathrm{c}} = \{s_1, \ldots, s_{N_\mathrm{c}}\},\ i = 1,2,3 \quad (21)$$

where s_j, $j = 1, \ldots, N_\mathrm{c}$ are symbols (e.g. numbers or letters). Therefore, the full discrete PDF is characterized by $(N_\mathrm{c})^3$ probabilities $\Pr(Y_{1,\mathrm{c}} = i, Y_{2,\mathrm{c}} = j, Y_{3,\mathrm{c}} = k)$; $1 \leq i, j, k \leq N_\mathrm{c}$, under $3(N_\mathrm{c} - 1)$ independent probability constraints of type $\Pr(Y_{1,\mathrm{c}} = i) = \Pr(Y_{2,\mathrm{c}} = j) = \Pr(Y_{3,\mathrm{c}} = k) = 1/N_\mathrm{c}$, $\forall i, j, k$ and a global normalization constraint. The triadic interaction information in the discrete form is

$$I_{3\mathrm{c}}(Y_1, Y_2, Y_3) \equiv I_3(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) \quad (22a)$$

$$= I(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) - I_2(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}), \quad (22b)$$

where $I_2(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) \equiv I(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}) + I(Y_{1,\mathrm{c}}, Y_{3,\mathrm{c}}) + I(Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}})$. Thanks to the data processing inequality (Cover and Thomas, 1991), the MI always decreases in the coarse-graining process i.e. $I(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) \leq I(Y_1, Y_2, Y_3)$ and $I_2(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) \leq I_2(Y_1, Y_2, Y_3)$. However, the same inequality is not necessary for I_3 of non-perfect triads. The conditions of a perfect triad i.e. $I_3(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) = I(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}})$, require

the constraining condition of pair-wise independence, formulated through the $3N_\mathrm{c}^2$ conditions: $\Pr(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}) = \Pr(Y_{1,\mathrm{c}}, Y_{3,\mathrm{c}}) = \Pr(Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) = 1/N_\mathrm{c}^2$, $\forall Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}$, from which only $3(N_\mathrm{c} - 1)^2$ are independent, thus leading to $N_\mathrm{c}^3 - 3N_\mathrm{c}(N_\mathrm{c} - 1) - 1$ independent probabilities.

Let us analyze it for the simplest case of perfect triads in which $N_\mathrm{c} = 2$, $B_2 = \{1,2\}$. After some algebra, we get all the $8 = 2^3$ joint probabilities, uniquely dependent on a single free positive parameter $p \in [0, 1/4]$. Any of the 8 joint probabilities are: p or $(1/4 - p)$, respectively when (one or three) or (zero or two), out of the three variables are set to $2 \in B_2 = \{1,2\}$. The corresponding IT is $I_3(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) = 3\ln(2) + 4\left[p\log(p) + (1/4 - p)\log(1/4 - p)\right]$. IT is minimized at $I_3 = 0$ when all the 8 probabilities are equal to $p = 1/8$ (independent variables) and maximized at $I_3 = \log(2)$ for $p = 0$ or $p = 1/4$ from which only $4 = N_\mathrm{c}^2$ events have non-vanishing probabilities. There, any third variable, for instance $Y_{3,\mathrm{c}}$, is totally determined as function of the remaining pair $(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}})$. The lookup tables of $Y_{3,\mathrm{c}}$ for $p = 0$ and $p = 1/4$ are given in Table 1 in which all the $N_\mathrm{c} = 2$ symbols are used (without repetition) in any row and any column as consequence of the requirements of independence of row indices $(Y_{i,\mathrm{c}})$, column indices $(Y_{j,\mathrm{c}})$ and table-cell values $(Y_{k,\mathrm{c}})$ with i, j, k different.

These non-repetition conditions are precisely those of the Latin-Squares (Dénes and Keedwell, 1974) (LS) of N_c symbols (e.g. Sudoku game in which $N_\mathrm{c} = 9$). The more synergetic perfect triads, i.e. those maximizing IT are the Latin-Square categorical relationships between N_c iid symbols with the IT: $I_3(Y_{1,\mathrm{c}}, Y_{2,\mathrm{c}}, Y_{3,\mathrm{c}}) = \log N_\mathrm{c}$, (Bailey, 1996). This is a quite useful property, for example in the efficiency of sampling and experiment design (e.g. the design of schedules of sports leagues derives from a Latin square).

Still considering the above cases of Latin squares of $N_\mathrm{c} = 2$ symbols and by making the correspondence of B_2 elements as: $1 \rightarrow \mathrm{True} \rightarrow$; $2 \rightarrow \mathrm{False}$, the above tables are isomorphic to the Boolean logic tables, respectively of the exclusive disjunction $Y_{3,\mathrm{c}} \equiv Y_{1,\mathrm{c}} \oplus Y_{2,\mathrm{c}} = \mathrm{True}$, iff $Y_{1,\mathrm{c}} \neq Y_{2,\mathrm{c}}$ and of the logic equivalence $Y_{3,\mathrm{c}} \equiv Y_{1,\mathrm{c}} \leftrightarrow Y_{2,\mathrm{c}} = \mathrm{True}$, iff $Y_{1,\mathrm{c}} = Y_{2,\mathrm{c}}$ (see Table 1).

The emerging property common to both tables is that, for ergodic processes and during 3/4 of the time (0.75 of probability), only one out of the three variables $(Y_{1,c}, Y_{2,c}, Y_{3,c})$ are equal to a given symbol S (e.g. above or below median), while during 1/4 of the time (0.25 of probability) the three together exhibit that symbol S.

An interesting result is that squares of trivariate cumulants (e.g. $(\kappa^{123})^2$, $(\kappa^{1123})^2$) appearing in Eq. (17) grow as far the joint trivariate PDF is close to certain LS forms. For example, let us consider the 2-symbol LS described by the Boolean equivalence $Y_{3_c} = (Y_{1_c} \leftrightarrow Y_{2_c})$ or $Y_{3_c} = Y_{1_c} Y_{2_c}$ where $Y_{i_c} \in \{-1, 1\}, i = 1,2,3$ are equally probable classes, μ_i is the median of Y_i, $i = 1,2,3$, η_i is the median of Y_i^2. That LS holds for two different category attributions. One renders $\kappa^{123} - \mu_1\mu_2\mu_3 = E[(Y_1 - \mu_1)(Y_2 - \mu_2)(Y_3 - \mu_3)]$ maximum for $Y_{i_c} = \mathrm{sgn}(Y_i - \mu_i)$, $i = 1,2,3$; the other maximizes $\kappa^{1123} - \mu_2\kappa^{113} - \mu_3\kappa^{112} + \mu_2\mu_3(1 - \eta_1) = E[(Y_1^2 - \eta_1)(Y_2 - \mu_2)(Y_3 - \mu_3)]$ for $Y_{1_c} = \mathrm{sgn}(Y_1^2 - \eta_1)$, $Y_{i_c} = \mathrm{sgn}(Y_i - \mu_i)$, $i = 2,3$.

A suggestive example of a perfect triad is a "love triangle" whose intervening parties meet together only in pairs such that each one meets someone else half of the time and one is alone on the other halftime. Asymmetric triads are also easily imagined from the above example, playing with the different levels of allowed collusions between members. Let us associate the variable values: +, (−) to $(Y_{1,c}, Y_{2,c}, Y_{3,c})$ components when one meets someone (is alone). These constraints lead to $(Y_{1,c}, Y_{2,c}, Y_{3,c})$ being pair-wise independent, filling the conditions of a perfect triad with $I_3(Y_{1,c}, Y_{2,c}, Y_{3,c}) = \log(2)$. Once an external control drives or they decide forming of a whole group (threesome), there will be redundancy and $I_3(Y_{1,c}, Y_{2,c}, Y_{3,c}) = -2I(Y_{1,c}, Y_{2,c}, Y_{3,c})$ becomes negative.

The aforementioned situation is somehow similar to another triadic interaction; though not perfect stating that only two, out of the three categorical variables are set to equal values, say occurring in phase. Considering in this case, that all the six possible values of $(Y_{1,c}, Y_{2,c}, Y_{3,c})$: (+,+,−), (+,−,+), (−,+,+), (−,−,+), (−,+,−), (+,−,−), have

equal probabilities 1/6, we see that variables are not anymore pair-wise independent giving a value $I_2(Y_{1,c}, Y_{2,c}, Y_{3,c}) = \log(\frac{32}{27})$ and a positive IT $I_3(Y_{1,c}, Y_{2,c}, Y_{3,c}) = \log(\frac{108}{96})$.

3 Optimization of the interaction information

3.1 Proxies of the interaction information and its maximization

The aim of this section is to find relevant triads in the form $\boldsymbol{Y}_{\mathrm{proj}} = (Y_1, Y_2, Y_3)' = \boldsymbol{P}_3 \mathbf{R} \boldsymbol{X}_{\mathrm{u,\,signal}}$, in the space of orthogonally rotated components of the N_{rot}-dimensional space of spherized PCs of the signal part of the original physical field. The IT $I_3(Y_1, Y_2, Y_3)$ is therefore a functional depending both on the rotation matrix $\mathbf{R}$ and on the joint PDF of $\boldsymbol{X}_{\mathrm{u,\,signal}}$ which can be quite difficult to estimate from finite samples due to the dimension N_{rot} of $\boldsymbol{X}_{\mathrm{u,\,signal}}$ due to the curse of dimensionality (Bellman, 1957). Therefore, we will follow a projection pursuit (PP) (Friedman and Tukey, 1974; Huber, 1985) strategy, by considering a projection index, i.e. a functional of the $\boldsymbol{Y}_{\mathrm{proj}}$ components simulating $I_3(Y_1, Y_2, Y_3)$, or equivalently, a function $F_{\mathrm{opIT}}(\mathbf{R})$, explicitly written in terms of variables characterizing orthogonal rotations, the generalized Euler angles (Raffenetti and Ruedenberg, 1969; Hoffman et al., 1972). This technique is also used in dyadic source separation of climatic data (Pires, 2014). One possibility for $F_{\mathrm{opIT}}(\mathbf{R})$ is to consider a truncated form of the Edgeworth expansion of IT of Eq. (17).

However, for the sake of geometric simplicity, we consider that proxy to be proportional to the square of some nonlinear correlation between zero average functions: $F_1(\boldsymbol{Y}_{\mathrm{proj}})$ and $F_2(\boldsymbol{Y}_{\mathrm{proj}})$, where the dependencies of $F_1(\boldsymbol{Y}_{\mathrm{proj}}), F_2(\boldsymbol{Y}_{\mathrm{proj}})$ include the three variables. In synthesis, $F_{\mathrm{opIT}}(\mathbf{R})$ is positive satisfying

$$F_{\mathrm{opIT}}(\mathbf{R}) = E[F_Y(\boldsymbol{Y}_{\mathrm{proj}})]^2 \tag{23a}$$

$$= k\,\mathrm{cor}(F_1, F_2)^2 \quad k > 0 \tag{23b}$$

$$F_Y(\boldsymbol{Y}_{\mathrm{proj}}) \equiv F_1(\boldsymbol{Y}_{\mathrm{proj}}) F_2(\boldsymbol{Y}_{\mathrm{proj}}), \tag{23c}$$

where $E(F_1) = E(F_2) = 0$ and $F_{\mathrm{opIT}}(\mathbf{R}) = 0, \forall \mathbf{R}$ when the joint PDF of $\boldsymbol{X}_{\mathrm{u,signal}}$ is Gaussian. Under non-Gaussian conditions, the goal is thus to find $\max_{\mathbf{R}} F_{\mathrm{opIT}}(\mathbf{R})$, i.e. the maximum over the space of rotations, leading to the dominant synergetic triad in the subspace span($\boldsymbol{X}_{\mathrm{u,signal}}$). In that space, if $|\mathrm{cor}(F_1, F_2)|$ is high enough, then data is concentrated along some principal manifold (Hastie and Stuetzle, 1989) of dimension two, i.e. along a surface with a geometric shape described by the regression equation

$$F_2\,\sigma(F_2)^{-1} = \mathrm{cor}(F_1, F_2) F_1\,\sigma(F_1)^{-1} + w, \tag{24}$$

where w is some independent noise and σ means standard variation.

We will consider the proxy relying on third-order cumulants by taking $F_Y(Y_1, Y_2, Y_3) = Y_1 Y_2 Y_3$ in Eq. (23a), i.e.

$$F_{\mathrm{opIT}}(\mathbf{R}) = E[F_Y(Y_1, Y_2, Y_3)]^2 \tag{25a}$$

$$= E(Y_1 Y_2 Y_3)^2. \tag{25b}$$

3.2 Maximization of the interaction information over the space of rotations

In order to get $\max_{\mathbf{R}} F_{\mathrm{opIT}}(\mathbf{R})$, $\mathbf{R}$ needs to be written in terms of a control vector spanning all possible rotations of a N_{rot}-dimensional metric space. This is done by using the vector of generalized Euler angles with a dimension $N_{\mathrm{ang}} = 1/2 N_{\mathrm{rot}}(N_{\mathrm{rot}} - 1)$ (e.g. $N_{\mathrm{ang}} = 3$ for $N_{\mathrm{rot}} = 3$):

$$\boldsymbol{\theta}(\mathbf{R}) = (\theta_1, \ldots, \theta_k, \ldots, \theta_{N_{\mathrm{ang}}})^{\mathrm{T}} \tag{26}$$

Each $\theta_k \in [0, 2\pi[$ rad is the rotation angle on the plane spanned by orthogonal components $(X_{u,i(k)}, X_{u,j(k)})$, $i(k) \neq j(k)$, where k is a counting index determining univocally the pair $[i(k), j(k)]$. In particular for the indexation $[i(k), j(k)] = (1,2), \ldots, (1, N_{\mathrm{rot}}), (2,3), \ldots, (2, N_{\mathrm{rot}}), \ldots, (N_{\mathrm{rot}} - 1, N_{\mathrm{rot}})$, $k = 1, 2, \ldots, N_{\mathrm{ang}}$ and taking the function $M_1(k) \equiv 1/2[(2N_{\mathrm{rot}} - 1) + \sqrt{(2N_{\mathrm{rot}} - 1)^2 - 8k}]$, we get the indices: $i(k) = \lceil M_1(k) \rceil$,

given by the ceiling function (minimum integer majorant) of $M_1(k)$ and $j(k) = k - 1/2\{[i(k) - 1][2N_{\mathrm{rot}} - i(k)] - 2i(k)\}$. The elementary rotation $N_{\mathrm{rot}} \times N_{\mathrm{rot}}$ matrix $\mathbf{R}_k(\theta_k)$ corresponding to angle θ_k has vanishing entries, except for:

$$\mathbf{R}_{[i(k),i(k)],(k)} = \mathbf{R}_{[j(k),j(k)],(k)} = \cos(\theta_k) \tag{27a}$$

$$\mathbf{R}_{[i(k),j(k)],(k)} = -\mathbf{R}_{[j(k),i(k)],(k)} = \sin(\theta_k) \tag{27b}$$

$$\mathbf{R}_{(l,l),(k)} = 1, \quad l \neq i(k),\ l \neq j(k), \tag{27c}$$

where $\mathbf{R}_{[i(k),j(k)],(k)}$ is the entry on the $i(k)$th row and $j(k)$th column of $\mathbf{R}_k$. Finally, the generic rotation matrix is obtained by the product of all elementary rotation matrices (Raffenetti and Ruedenberg, 1969) as $\mathbf{R} = \mathbf{R}_1 \mathbf{R}_2 \ldots \mathbf{R}_{N_{\mathrm{ang}}}$. The proxy functions are thus functions of $\mathbf{R}(\boldsymbol{\theta})$.

Then, we find local maxima of $F_{\mathrm{opIT}}[\mathbf{R}(\boldsymbol{\theta})]$ using the Quasi-Newton descent-gradient technique on $-F_{\mathrm{opIT}}(\mathbf{R}(\boldsymbol{\theta}))$ implemented by the Fortran subroutine M1QN3 (Gilbert and Lemaréchal, 1989). The gradient is coded with respect to angle components. The rotation optimization in terms of Euler angles (Hoffman et al., 1972) is alternative to that relying on direct components of $\boldsymbol{Y}_{\mathrm{proj}}$ constrained by orthogonality relationships through Lagrange multipliers (Jennrich, 2001; Trendafilov, 2006). The function F_{opIT} is not necessarily globally convex, hence owing multiple local maxima like the contrast function (Comon, 1994) maximized in ICA. Therefore, a reasonable number N_{trial} of randomly chosen first guess angle vectors are chosen for the maximization. In order to assure that the absolute maximum is reached, the number N_{trial} must increase with N_{rot} since the number of local maxima also increases.

The derivative of $F_{\mathrm{opIT}}[\mathbf{R}(\boldsymbol{\theta})]$ with respect to θ_k is obtained by the derivative chain rule applied to $\theta_k \rightarrow \boldsymbol{Y}_{\mathrm{proj}} \rightarrow F_Y \rightarrow F_{\mathrm{opIT}}$ as

$$\frac{\partial F_{\mathrm{opIT}}}{\partial \theta_k} = 2E(F_Y) \sum_{i=1}^{3} E\left[\frac{\partial F_Y}{\partial Y_i}\frac{\partial Y_i}{\partial \theta_k}\right] \tag{28a}$$

$$= 2E(F_Y) \sum_{i=1}^{3}\sum_{p=1}^{N_{\mathrm{rot}}}\sum_{q=1}^{N_{\mathrm{rot}}}\sum_{l=1}^{N_{\mathrm{rot}}} \mathbf{R}_{i',p}\mathbf{R}_{i'',q}\mathbf{S}_{i,l(k)}E[X_{\mathrm{u},p}X_{\mathrm{u},q}X_{\mathrm{u},l}] \tag{28b}$$

$$\frac{\partial Y_i}{\partial \theta_k} = \sum_{l=1}^{N_{\mathrm{rot}}} \mathbf{S}_{i,l(k)}X_{\mathrm{u},l} \tag{28c}$$

where (i, i', i'') in Eq. (28b) is a permutation of $(1,2,3)$ and $\mathbf{S}_{i,l,(k)}$ (line i, column l) are entries of the matrix $\mathbf{S}_k \equiv (\boldsymbol{P}_3\mathbf{R}_1 \ldots \frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k} \ldots \mathbf{R}_{N_{\mathrm{ang}}})$, where the non-vanishing components of $\frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k}$ are:

$$\left(\frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k}\right)_{i(k),i(k)} = \left(\frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k}\right)_{j(k),j(k)} \tag{29a}$$

$$= -\sin(\theta_k); \tag{29b}$$

$$\left(\frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k}\right)_{i(k),j(k)} = -\left(\frac{\mathrm{d}\mathbf{R}_k}{\mathrm{d}\theta_k}\right)_{j(k),i(k)} \tag{29c}$$

$$= \cos(\theta_k) \tag{29d}$$

The gradient depends on pre-computed third-order moments of un-rotated variables hence saving time in the maximization. The algorithm can be generalized to an arbitrary control function.

4 Triadic interactions in the 40-variable Lorenz model

4.1 Model physics

The previous information-theoretic analysis for the statistical diagnosing of triadic interactions will be applied to a minimal one-dimensional (1-D) fluid motion model, retaining the nonlinear advective part of Navier–Stokes equations (NSEs) and being subjected to constant external forcing and linear dissipation. Under that framework, we will directly relate triadic correlations with the physical-meaning conditions of triadic wave resonance (Bartello, 1995), thus giving the potential relevance of those diagnostics when applied to more complex models or to observational datasets.

For that purpose, we use the 40-variable Lorenz-95 model (Lorenz, 1995) that is written in terms of $N = 40$ discretized, equally spaced values $X_i(t), i = 1, \ldots, N$ of the fluid velocity field $u(x,t)$ around a latitude circle with circular periodic conditions, where $x \in [0, 2\pi[$ is longitude and t is time. The model equations are

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = X_{i-1}(X_{i+1} - X_{i-2}) - X_i + F; \quad i = 1, \ldots, N,\ X_0 \equiv X_N,\ X_{-1} \equiv X_{N-1}, \tag{30}$$

where F is the constant forcing. Coefficients of Eqs. (30) (e.g. one unit for the dissipative term) are adjusted in such a way that one time unit roughly represents 5 days of large-scale atmospheric circulation (Lorenz, 1995). Here, we take $F = 8$, also used in predictability (Lorenz and Emanuel, 1998) and data assimilation studies (Evenson and Fario, 1997; Van Leeuwen, 2010), leading to a chaotic attractor with an average perturbation-doubling period of 0.42 time units (2.1 days). The Lorenz-95 model has a single fixed unstable state $X_i = F, \forall i$ with correspondent perturbations of different wavelengths exchanging energy between them and propagating eastward (towards higher indices). The Eqs. (30) are invariant under transformations $i \rightarrow i + l$ due to the circular symmetry, hence the statistics over the attractor's PDF reflect it by the invariance of the marginal PDFs ρ_{X_i} with respect to i. Similarly, the 2-D and 3-D projected PDFs ρ_{X_i,X_l} and ρ_{X_i,X_l,X_n} with $i \le l \le n$, are arbitrary functions respectively of

type $f_1(l-i)$ and $f_2(l-i, n-l)$. Moreover, the quadratic terms are conservative leading to the kinetic energy equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\sum_{i=1}^{N} X_i^2\right) = F\sum_{i=1}^{N} X_i - \sum_{i=1}^{N} X_i^2 \tag{31}$$

Taking infinite time averages of Eqs. (30) and (31), i.e. averages over the attractor's ergodic PDF and denoting the average of X_i by μ_X, its variance by σ_X^2 and its shift-correlations by $c_{X,l} \equiv \mathrm{cor}(X_i, X_{i+l})$, we get, thanks to the circular symmetry:

$$0 \le \mu_X = F - \sigma_X^2(c_{X,1} - c_{X,2}) \tag{32a}$$

$$= (c_{X,1} - c_{X,2})^{-1} \le F \tag{32b}$$

$$0 \le \sigma_X^2 = \mu_X(F - \mu_X) \le \frac{1}{4}F^2. \tag{32c}$$

All the theory of Sects. 2 and 3 applies, with the state vector $\boldsymbol{X} \equiv (X_0, \ldots, X_{N-1})'$ substituted by the perturbed vector $\boldsymbol{X} - \overline{\boldsymbol{X}} = \boldsymbol{X} - \mathbf{1}\mu_X$ where $\mathbf{1}$ is a vector of "ones".

4.2 Model Fourier transform

Some conclusions are drawn by using a Fourier analysis of Eq. (30). The solutions of the model are N-shift periodic, hence a discrete Fourier transform (DFT) (http://en.wikipedia.org/wiki/Discrete_Fourier_transform) applies by expanding X_i in terms of

discrete complex exponentials:

$$X_i(t) = \frac{1}{N}\sum_{k=-N/2+1}^{N/2} T_{X,k}(t)\Psi_{k,i}; \tag{33a}$$

$$T_{X,k}(t) = \sum_{i=0}^{N-1} X_i(t)\Psi^*_{k,i}; \tag{33b}$$

$$\Psi_{k,i} = \exp\left(\mathrm{i}\frac{2\pi}{N}ik\right) \tag{33c}$$

where $\mathrm{i} = \sqrt{-1}$, the star denotes conjugate complex, k is the integer wave-number, $\Psi_{k,i}$ are orthogonal sinusoidal functions, collected in the matrix $\boldsymbol{\Psi}$ of column vectors, satisfying the normalization $\sum_{i=0}^{N-1}\Psi_{k',i}\Psi^*_{k,i} = N\delta_{k,k'}$ and $T_{X,k}(t)$ is the time-depending kth DFT verifying the invariance $T_{X,k} = (T_{X,-k})^* = T_{X,k\pm N}$. Representing the state and DFT vectors by $\boldsymbol{X} \equiv (X_0, \ldots, X_{N-1})'$ and $\boldsymbol{T_X} \equiv (T_{-N/2+1}, \ldots, T_{N/2})'$ respectively, we collect Eqs. (33a,b) in the condensed representation $\boldsymbol{X} = 1/N\boldsymbol{\Psi}\boldsymbol{T_X}$ where $\boldsymbol{T_X} = \boldsymbol{\Psi}^*\boldsymbol{X}$ with the orthogonality condition condensed as $\boldsymbol{\Psi}\boldsymbol{\Psi}^* = \boldsymbol{\Psi}^*\boldsymbol{\Psi} = N\mathbf{I}_N$ where $\mathbf{I}_N$ is the N-dimensional identity matrix and the star means adjoint ≡ transpose of conjugate complex.

Now, we will present fundamental results relying on single and cross DFT statistics, coming from the circular invariance of the state vector statistics. The instantaneous DFT of the l-shift of X_i is $T_{X_{i+l},k}(t) = T_{X_i,k}(t)\exp\left(\mathrm{i}\frac{2\pi}{N}lk\right), \forall l \in \mathbb{Z}$. However, its time average (hereafter denoted by overbar) is independent of l, leading to $\overline{T_{X_{i+l},k}} = \overline{T_{X_i,k}} = \overline{T_{X_i,k}}\exp\left(\mathrm{i}\frac{2\pi}{N}lk\right)$, hence the wavy-DFT averages vanish, i.e. $\overline{T_{X_i,k}}$, $k \ne 0$. The sole non-null DFT is that corresponding to the average: $\overline{T_{X_i,0}} = N\mu_X$.

Similarly, taking the product of two DFTs, the moment $\overline{T_{X_{i+l},k}T_{X_{i+l},k'}}$ shall be independent of l due to circular symmetry, leading to $\overline{T_{X_{i+l},k}T_{X_{i+l},k'}} = \overline{T_{X_i,k}T_{X_i,k'}} =$

$\overline{T_{X_i,k}T_{X_i,k'}}\exp\left[\mathrm{i}\frac{2\pi}{N}l(k+k')\right]$, hence $\overline{T_{X_i,k}T_{X_i,k'}}=0$, $k'\neq -k$, i.e. all pairs of DFTs are uncorrelated except those in which one is the complex conjugate of the other, giving the average Fourier power: $\overline{T_{X_i,k}T_{X_i,-k}}=\overline{|T_{X_i,k}|^2}\equiv P_{X,k}\geq 0$. By application of the Parseval theorem, one obtains the total variance of $\boldsymbol{X}$ as $N\sigma_X^2=1/N\mathrm{var}(T_{X,0})+2/N\sum_{k=1}^{N/2-1}P_{X,k}+1/NP_{X,N/2}$.

The uncorrelatedness of the DFTs has a relevant consequence regarding the principal components (PCs) of the state vector $\boldsymbol{X}(t)$. The standardized PCs, merged in the vector $\boldsymbol{X}_\mathrm{u}(t)$, are uncorrelated between each other and have unit variance. The standardized DFTs are collected in the N-dimensional zero-average vector $\boldsymbol{T}_{\boldsymbol{X},\mathrm{u}}\equiv\mathbf{P}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{T}_{\boldsymbol{X}}-\overline{\boldsymbol{T}_{\boldsymbol{X}}})$ whose covariance matrix is $\mathbf{P}_{\boldsymbol{X}}\equiv\mathrm{diag}(P_{N/2-1},\ldots,P_1,\mathrm{var}(T_0),P_1,\ldots,P_{N/2})$. The standardized DFTs are uncorrelated and have unit variance too, leading to the covariance matrix between complex variables: $\overline{\boldsymbol{T}_{\boldsymbol{X},\mathrm{u}}\boldsymbol{T}^*_{\boldsymbol{X},\mathrm{u}}}=\mathbf{I}_N$. Therefore, since both $\boldsymbol{X}_\mathrm{u}(t)$ and $\boldsymbol{T}_{\boldsymbol{X},\mathrm{u}}$ are statistically spherized, $\boldsymbol{X}_\mathrm{u}(t)$ can be expressed in terms of an orthogonal rotation of $\boldsymbol{T}_{\boldsymbol{X},\mathrm{u}}$, i.e. $\boldsymbol{X}_\mathrm{u}(t)=\mathbf{R}\boldsymbol{T}_{\boldsymbol{X},\mathrm{u}}=\mathbf{R}\mathbf{P}_{\boldsymbol{X}}^{-1/2}\boldsymbol{\Psi}^*(\boldsymbol{X}-\overline{\boldsymbol{X}})$, where $\mathbf{R}$ is a N-dimensional complex unitary matrix, i.e. $\mathbf{R}\mathbf{R}^*=\mathbf{R}^*\mathbf{R}=\mathbf{I}_N$. Now, using the notation of Sect. 2.1, the vector of PCs is

$$\boldsymbol{X}_\mathrm{PC}=\boldsymbol{\Lambda}^{1/2}\boldsymbol{X}_\mathrm{u} \tag{34a}$$

$$=\mathbf{W}'(\boldsymbol{X}-\overline{\boldsymbol{X}}) \tag{34b}$$

$$=\boldsymbol{\Lambda}^{1/2}\mathbf{R}\mathbf{P}_{\boldsymbol{X}}^{-1/2}\boldsymbol{\Psi}^*(\boldsymbol{X}-\overline{\boldsymbol{X}}), \tag{34c}$$

where, as before, $\mathbf{W}$ is the matrix of column EOF-vectors and $\boldsymbol{\Lambda}$ is the diagonal matrix of PC variances. Eq. (34) holds for any time, i.e. for any value of $\boldsymbol{X}-\overline{\boldsymbol{X}}$, hence the matrices multiplying that in Eqs. (34b) and (34c) must be equal, i.e.

$$\mathbf{W}'=\boldsymbol{\Lambda}^{1/2}\mathbf{R}\mathbf{P}_{\boldsymbol{X}}^{-1/2}\boldsymbol{\Psi}^*. \tag{35}$$

Since the EOFs are spatially orthogonal, one has $\mathbf{W}'\mathbf{W}=\mathbf{I}_N$ leading to $\mathbf{I}_N=N\boldsymbol{\Lambda}^{1/2}\mathbf{R}\mathbf{P}_{\boldsymbol{X}}^{-1}\mathbf{R}^*\boldsymbol{\Lambda}^{1/2}$, which is only possible if the rotation matrix $\mathbf{R}$ is composed of a sum of N' unitary matrices: $\mathbf{R}=\mathbf{R}_1+\ldots+\mathbf{R}_{N'}$, with each one $\mathbf{R}_k$ rotating a different set of $d(k),k=1,\ldots,N'$ degenerated DFTs, i.e. with the same variance $p(k)$, taken from $\mathbf{P}_{\boldsymbol{X}}$, producing the same number $d(k)$ of degenerated, equally variant PCs with the same variance $\lambda(k)=(1/N)p(k)$.

Let us arrange $\boldsymbol{\Psi}=\boldsymbol{\Psi}_1+\ldots+\boldsymbol{\Psi}_{N'}$ as a sum of matrices, each one corresponding to the concatenation of the $d(k)$ rotated exponential complex vectors. Since $\mathbf{R}_k\mathbf{R}^*_{k'}=\mathbf{I}_N\delta_{k,k'}$ and $\boldsymbol{\Psi}^*_k\boldsymbol{\Psi}_{k'}=N\mathbf{I}_N\delta_{k,k'}$, one can arrange the EOF matrix as $\mathbf{W}=\mathbf{W}_1+\ldots+\mathbf{W}_{N'}$ with

$$\mathbf{W}'_k=\frac{\lambda(k)^{1/2}}{p(k)^{1/2}}\mathbf{R}_k\boldsymbol{\Psi}^*_k \tag{36a}$$

$$=\frac{1}{\sqrt{N}}\mathbf{R}_k\boldsymbol{\Psi}^*_k. \tag{36b}$$

If the Fourier spectrum is non-degenerated, the EOFs corresponding to a certain wave-number $k=1,\ldots,\frac{N}{2}-1$ have degeneracy two with EOFs being sinusoidal functions in quadrature with values in the ith point: $\mathbf{W}_{i,k}=\sqrt{2/N}\left[\cos(\mathrm{i}\frac{2\pi}{N}ik+\phi_k),\sin(\mathrm{i}\frac{2\pi}{N}ik+\phi_k)\right]$, where ϕ_k is a given phase. The PCs come simply as the cosine and sine transforms. For degenerated spectra, EOFs mix different wave-numbers.

The product of three DFTs must also be constrained by the circular symmetry, i.e. the moment $\overline{T_{X_{i+l},k}T_{X_{i+l},k'}T_{X_{i+l},k''}}$ shall be independent of l, leading to $\overline{T_{X_{i+l},k}T_{X_{i+l},k'}T_{X_{i+l},k''}}=\overline{T_{X_i,k}T_{X_i,k'}T_{X_i,k''}}=\overline{T_{X_i,k}T_{X_i,k'}T_{X_i,k''}}\exp\left[\mathrm{i}\frac{2\pi}{N}l(k+k'+k'')\right]$, from which $\overline{T_{X_i,k}T_{X_i,k'}T_{X_i,k''}}=0\,\mathrm{iff}\,k+k'+k''\neq 0$. It means that for $k,k',k''\neq 0$, the triadic correlations given by Eq. (18) and consequently the interaction information between the standardized DFTs $T_{X_i,k},T_{X_i,k'},T_{X_i,k''}$ can only be non null if the wave-numbers sum is $k+k'+k''=0$ which is equivalent to the resonance condition for wave-numbers (Bartello, 1995). The

PCs corresponding to wave-numbers under triadic resonance must also exhibit non-null triadic correlations. Similar relationships can be obtained for quartic interactions: $\overline{T_{X_i,k}T_{X_i,k'}T_{X_i,k''}T_{X_i,k'''}} = 0$ iff $k+k'+k''+k''' \neq 0$, to be identified by the quartic interaction information appearing in the expansion of Eqs. (7) and (8).

There is a vast literature on Resonance Wave Theory (RIT) on different contexts (Ziman, 1960; Ball, 1964; Hammack, 1993). Let us briefly describe triadic wave ressoance. By linearizing the evolution Eqs. (30) around the fixed point, $X_i = F$ and admitting propagating wave perturbations $C\exp\left[\mathrm{i}\frac{2\pi}{N}(kX_i - \omega t)\right]$ of small amplitude C, one obtains a dispersion relationship $\omega = \omega(k)$. Then, by superposing two waves with the (wave-number, frequency) pairs (k_1, ω_1) and (k_2, ω_2), then their weakly nonlinear interaction lead to the growing of waves (k_3, ω_3) satisfying the resonance condition $k_1 \pm k_2 \pm k_3 = 0$ and $\omega_1 \pm \omega_2 \pm \omega_3 = 0$. Here, a non-null triadic correlation and synergetic interaction has the meaning of a triadic resonance, where one wave comes as the physical result of the two others. However, the triads are not perfect because the DFTs are uncorrelated but not independent due to quartic moments $\overline{T_{X_i,k}T_{X_i,-k}T_{X_i,k'}T_{X_i,-k'}} \neq 0$.

The Fourier power can be fully explained by triadic wave interactions. In fact, let us consider the evolution equation of the kth DFT by taking the DFT of Eq. (30)

$$\frac{\mathrm{d}T_{X,k}}{\mathrm{d}t} = -T_{X,k} + FN\delta_{k,0} + \frac{1}{N}\sum_{l=-N/2+1}^{N/2}(T_{X,l}T_{X,k-l})(\eta^{2l-k} - \eta^{2k-l}); \tag{37a}$$

$$\eta \equiv \exp\left(\mathrm{i}\frac{2\pi}{N}\right). \tag{37b}$$

Then, multiplying Eq. (37a) by $2T_{X_i,-k}$ and taking the real part, one gets

$$\frac{\mathrm{d}|T_{X,k}|^2}{\mathrm{d}t} = 2\mathrm{Re}\left(T_{X,-k}\frac{\mathrm{d}T_{X,k}}{\mathrm{d}t}\right) \tag{38a}$$

$$= -2|T_{X,k}|^2 + 2FN\delta_{k,0}T_{X,0} + \frac{2}{N}\sum_{l=-N/2+1}^{N/2}\mathrm{Re}\left[\left(T_{X,l}T_{X,k-l}T_{X,-k}\right)\left(\eta^{2l-k} - \eta^{2k-l}\right)\right]. \tag{38b}$$

Finally, the time average of Eq. (38), leads to the power of DFTs

$$P_k = \delta_{k,0}N^2(\sigma_X^2 + \mu_X^2) + \frac{2}{N}\sum_{l=-N/2+1}^{N/2}\mathrm{Re}\left[\left(\overline{T_{X,l}T_{X,k-l}T_{X,-k}}\right)\left(\eta^{2l-k} - \eta^{2k-l}\right)\right], \tag{39}$$

where the energy of waves ($k \neq 0$) is totally due to triadic interactions.

4.3 Principal component analysis

The aforementioned theoretical results will be tested in practice, especially in what concerns the values of triadic correlation and interaction information. First, we have integrated Eq. (30) for a long period of 7200 time units ($\sim$ 98 years), after an attractor relaxation of 100 time units, using the fourth-order Runge–Kutta method and a time-step $\Delta t = 0.01$ ($\sim$ 1.2 h). The sampling average and variance are $\mu_X = 2.34$ and $\sigma_X^2 = 3.63^2 = 13.18$ respectively, while the one- and two- shift correlations are $c_{X,1} = 0.063$; $c_{X,2} = -0.361$ with Eqs. (32a, b) verifying quite accurately.

In order to compress data and reduce data redundancy, we have used running averages ($\tilde{X}$) of length 0.2 time units (1 day), leading to a sample of 36 000 (daily) realizations. For that period, the lag correlation (between running averages) is reduced to $\mathrm{cor}[X_i(t), X_i(t+0.2)] = 0.6$ producing a quite weak variance reduction since $\sigma_{\tilde{X}}^2 =$

12.14 which is quite close σ_X^2, hence Eqs. (32a, b) still verifies with $\sim 10\,\%$ relative error.

The equation governing running averages is identical to Eq. (30) by adding running time-covariances (equivalent Reynolds stresses), much smaller than the average terms since the decorrelation time (~ 2 time units) is much larger than the running time length. Therefore, the conclusions of Sect. 4.2 are plausible.

A PCA over this set is produced with the correspondent EOFs $W_k(x)$ and explained variances λ_k, $k = 1, \ldots, N$. As demonstrated, for non-degenerated Fourier spectra, the theoretical EOFs must be sinusoidal functions of wave numbers $k = 0, \ldots, N/2 = 20$.

The graphs of the data-derived EOFs are presented in Fig. 3 (wave-numbers 1–8) and Fig. 4 (wave-numbers 9–20). They are quite close to sinusoidal functions and appear in pairs corresponding to wave-numbers $k = 1, \ldots, N/2 = 20$, which are identified by visual inspection by counting the number of oscillations around the latitude circle. There is some wave mixing in EOF graphs due to the variance spectrum quasi-degeneracy, finite sampling and running averaging producing some spectrum convolution.

The PC variances λ_k for each pair of degenerated EOFs are shown in Fig. 5 for increasing values of wave-number k. The dominant mode is $k = 8$, quite near to the central wave-number 10. Then, the variances decrease both for increasing $k > 8$ up to $k = 20$ and for decreasing $k < 8$ down to $k = 1$. From Fig. 5, we see that pairs of smaller variances tend to be degenerated, i.e. those associated to k and $21–k$ are quite similar. In fact, from Fig. 4, EOFs with high wave number k tend to be modulated by a smaller wave-number, close to $21–k$ (e.g. $k = 19$ and $k = 20$). Differences between degenerated variances are quite small and are due to finite sampling effects.

The autocorrelation-function (ACF) of PCs $X_{\mathrm{PC},i}$, $i = 1, \ldots, N$ is oscillating with a typical period and tending to vanish for growing time lags. In Fig. 5, we present the decorrelation time (in time units), given by the lag beyond which the absolute value of autocorrelation is less than 0.05. It is higher for the most variant PCs and generally decreases with the PC variances, which also occurs with ACF periods. Since

the largest decorrelation time is ~ 7.5 units, the number of temporal degrees of freedom is $N_{\mathrm{dof}} \sim 7200/7.5 = 960$.

4.4 Triadic correlation and wave resonance

In order to assess the possibility of triadic resonance, we compute the triadic correlations from Eq. (18) between PCs associated to a triad of wave-lengths p, q, r. However, the wave phase must be the appropriate for resonance, hence we take the maximum from the 6 possible triadic correlations by considering the pairs of degenerated PCs associated to each wave-number. Let us denote by w_i the wave-number associated to the ith PC (quite clear in our case), where i grows for decreasing variance and by $X_{\mathrm{u},i}$ the standardized ith PC. That maximum is defined for a triad (p, q, r) of wave-numbers as

$$C_{\mathrm{RS}}(p,q,r) \equiv \max_{w_i=p, w_j=q, w_k=r} |\mathrm{cor}_3(X_{\mathrm{u},i}, X_{\mathrm{u},j}, X_{\mathrm{u},k})| \tag{40}$$

Moreover, in order to study the dependence of correlations on wave-numbers, we present the wave-number of the PC of maximum variance in the triad:

$$P_{\mathrm{RS}}(p,q,r) = \underset{w_i=p, w_j=q, w_k=r}{\arg\max} (\lambda_i, \lambda_j, \lambda_k) \tag{41}$$

and the wave-resonance number:

$$\delta_{\mathrm{RS}}(p,q,r) \equiv \min(|p+q-r|, |p+r-q|, |r+q-p|) \tag{42}$$

The wave resonance condition holds for $\delta_{\mathrm{RS}}(p,q,r) = 0$. The scatter-plot of $P_{\mathrm{RS}}(p,q,r)$ and $\delta_{\mathrm{RS}}(p,q,r)$ against $C_{\mathrm{RS}}(p,q,r)$ for all combinations p, q, r is shown in Fig. 6. The 95 % confidence threshold of significant correlation is $1.96/\sqrt{N_{\mathrm{dof}}} = 0.063$. From values of P_{RS}, we conclude that in the triads of largest C_{RS} values, intervene the most variant PCs, i.e. with wave-numbers $P_{\mathrm{RS}} \in \{7, 8, 9, 10\}$ and verifying resonance

condition $\delta_{RS} = 0$, consistent with results of Sect. 4.2. A few statistically significant values of C_{RS} occur for $\delta_{RS} \neq 0$ which can be due to wave resonance coming from non-dominant wave-numbers contaminating the EOFs.

The estimated triadic IT (see estimator in Appendix B) and the triadic absolute correlation $|\mathrm{cor}_3|$ in the validation period (second half the total period) of 18 000 samples ($N_{dof} \sim 480$), is shown in Fig. 7. The thresholds of 95 %-significant correlations and ITs are ~ 0.09 and ~ 0.008 respectively (see Figs. B1 and B2). The three largest values of $C_{RS}(p,q,r)$ are 0.23 for $p = q = 10, r = 20$, 0.22 for $p = q = 9, r = 18$ and 0.21 for $p = 8$, $q = 9$, $r = 17$, all satisfying the WRC. Note that when $p = q$, one uses PCs in quadrature for the same wave-number.

For $|\mathrm{cor}_3| > 0.12$, IT is generally a growing function of $|\mathrm{cor}_3|$ with a quite good matching $|\mathrm{cor}_3| \sim \mathrm{cor}_{3eq}$ given by Eq. (20), showing that the interaction information comes from the MI between one PC and the product of the remaining two.

4.5 Optimized triads

In this section we apply the algorithm of Sect. 3.2 in order to find orthogonal rotations of standardized PCs (within a given set S introduced in Sect. 2.1) that maximize $|\mathrm{cor}_3|$ and IT. As expected, much larger values can be obtained than when using non-rotated PCs. For that, we consider a sequence of 9 encapsulated sets $S_1 \subset S_2 \subset \ldots \subset S_9$ of rotated PCs, each one corresponding to a set S_{wn-i}, $i = 1, \ldots, 9$ of wave-numbers chosen from the triads (p,q,r) providing the higher C_{RS} values. Let us denote $DS_{wn-i} \equiv S_{wn-i}/S_{wn-(i-1)}$, $i = 2, \ldots, 9$. Then, after analysing triadic correlations among PCs, we obtain:$S_{wn-1} = \{10, 20\}$, $DS_{wn-2} = \{9, 18\}$, $DS_{wn-3} = \{8, 17\}$, $DS_{wn-4} = \{7\}$, $DS_{wn-5} = \{11\}$, $DS_{wn-6} = \{16\}$, $DS_{wn-7} = \{12\}$, $DS_{wn-8} = \{19\}$, $DS_{wn-9} = \{15\}$. The first and last sets have 4 and 24 PCs with the spaces of rotations spanned by 6 and 276 Euler angles respectively. In order to get the maximum value of the control function $F_{opIT}(\mathbf{R})$ (Eq. 25), proportional to $|\mathrm{cor}_3|^2$, we use a set of $N_{trial} = 40$ first guesses of the Euler angles. The number of local maxima of F_{opIT} increases in general with the size of space

of spanning PCs: S_{wn-i}. Figure 8 shows the maximum value of $|\mathrm{cor}_3|$, optimized over rotations and for each set S_{wn-i}. That is computed in the validation period by using the optimized rotation matrix $\mathbf{R}$ (see Sect. 2.1) calibrated in the training period (first half of the total period). Let us denote that maximum as $|\mathrm{cor}_3|_{max\ val,i}$. The rotation coefficients are computed in the calibration and then applied in the validation period. As expected, $|\mathrm{cor}_3|_{max\ val,i}$ grows with the rotation space dimension, reaching the value ~ 0.44. The values are also very close to those computed in the training (not shown), showing that rotation coefficients are robust and not subjected to over-fitting regarding the number of calibrated Euler angles.

Let (Y_1, Y_2, Y_3) be the optimally rotated normalized PCs for each set. Figure 8 also shows the IT values $I_3(Y_1, Y_2, Y_3)$, which are largely due to the non-Gaussian part since the Gaussian part is $\sim 10^{-5}$. IT grows with the rotation space dimension where $I_3 \sim -1/2 \log[1 - (\mathrm{cor}_3)^2]$, since the approximation $|\mathrm{cor}_3| \sim \mathrm{cor}_{3eq}$ verifies quite well (as in Fig. 7 for high correlations).

As previously discussed, the triads are not perfect due to the lack of statistical dyadic independence, between PCs and of their rotations. Therefore, in order to evaluate how triads (Y_1, Y_2, Y_3) of rotated normalized PCs approximate to asymmetric triads, we estimate the values I_{3-as2}, I_{3-as1} Eqs. (10a) and (10b) in the validation period (see Fig. 8), by measuring the maximal values of the type $I(Y_i, Y_j|Y_k)$ and $I[(Y_i, Y_j), Y_k]$, with i, j, k being a permutation of 1,2,3. The non-Gaussian dyadic MIs come from high-order cumulants which cannot be vanished by rotations alone. The total MI I_i is also evaluated, also growing with a concomitant triadic correlation.

As explained in Sect. 2.6, a high value of $|\mathrm{cor}_3(Y_1, Y_2, Y_3)|$ means a high value of the absolute correlation between the sign of any product: $\mathrm{sgn}(Y_i Y_j)$ and the sign of the remaining variable: $\mathrm{sgn}(Y_k)$. This means that the binary random variables: $Y_{1c} = \mathrm{sgn}(Y_1), Y_{2c} = \mathrm{sgn}(Y_2), Y_{3c} = \mathrm{sgn}(Y_3)$ tend also to exhibit high interaction information with the bulk of probabilities disposed in the entries of a 2-category Latin square.

In order to show that, we estimate the discrete version of the interaction information: $I_{3c} \equiv I_3(Y_{1,c}, Y_{2,c}, Y_{3,c})$, (Eq. 22) in the validation period for the different sets (see

Fig. 8). All the estimated information measures grow consistently with the size of the optimization space.

Now, we study the physical meaning of the optimized triads. Therefore, for a generic PC-set S (within the selected ones S_k, $k = 1,2,3,5,7,9$), we plot in Fig. 9 the loadings given by Eq. (1): $W_{Y_i,k} = \sum_{j\in S} \lambda_j^{-1/2} R_{k,j} W_{i,j}$; $k = 1,2,3$; $i = 1,\ldots,N$ giving the triad components $Y_k = \sum_{i=1}^{N} W_{Y_i,k}(X_i - \mu_X)$ where $W_{i,j}$ is the jth value multiplying the perturbed value of the original field at the ith longitude.

For the sets S_1, S_2, the loadings cover the whole latitude circle whereas for S_3, S_5, S_7, S_9 they are quite restricted to local segments. In all cases, the loading-fields oscillate approximately in sinusoidal form within a revelant interval (circular or local) with the number of oscillations p_k, $k = 1,2,3$ satisfying the resonance condition. For example, for sets S_5, S_7, S_9, we obtain triadic resonant local wave-trains where $p_2 = p_3 = 3, p_1 = 6$, thus satisfying the resonance condition: $3+3 = 6$. The rotations act in such a way as to practically vanish loadings outside of the wave-train location. Other local maxima of the control function $F_{\mathrm{opIT}}(\mathbf{R})$, different from the absolute one verify for other wave-trains at different locations (not shown).

In order to verify how signals of triad components behave, we plot in Fig. 10, the time-series of the standardized variables $Y_1(t), Y_2(t), Y_3(t)$ corresponding to the optimized triad for set S_9 issued from an interval of 20 time units. The time-series $Y_2(t), Y_3(t)$, corresponding to the 3-wave dominating loadings (Fig. 9), present oscillating lag-correlations with a typical period of ~ 4 time units. The time series $Y_1(t)$ has no statistically significant lag-correlations with both $Y_2(t), Y_3(t)$ and it has a de-correlation lag of ~ 5 time units. Any of the 3 variables are quite close to Gaussianity, since that all the absolute skewnesses are lower than 0.04. The kurtosis excesses ($E[()^4] - 3$) are respectively $-0.6, -0.4, -0.3$ for Y_1, Y_2, Y_3.

Despite the pair-wise uncorrelatedness, the three variables as a whole are clearly not independent. In order to highlight that, we show in Fig. 11, the normalized histogram (in logarithmic scale) of the product of mutually uncorrelated variables: $Y_1 Y_2 Y_3$. The PDF of that product is highly asymmetric and leptokurtic (higher kurtosis than for a Gaussian)

presenting negative values of the average and skewness, respectively: -0.44 and -2.66 and a kurtosis excess of 10.16. That skewness means the occurrence of quite enhanced conditions of local resonance. The triadic correlation has a consistent value of $\mathrm{cor}_3[Y_1(t), Y_2(t), Y_3(t)] \sim -0.44$.

That negative triadic correlation is compatible with the fact that the signs of (Y_1, Y_2, Y_3) lie mostly of the time in one of the 4 octets: (+,+,−), (+,−,+), (−,+,+) and (−,−,−), each one occurring with probabilities $\sim 17\,\%$ over a total of 68 %. Those states alternate randomly throughout the time-series as it is clear in Fig. 10. In about $\sim 51\,\%$ of the time, one of the three variable lies in the negative phase while the two others are positively phased. When one of the octets occurs, the Boolean exclusive disjunction $\mathrm{sgn}(Y_1) \oplus \mathrm{sgn}(Y_2) = \mathrm{sgn}(Y_3)$ holds, corresponding to a Latin square of $N_c = 2$ categories as described in Sect. 2.6. The other 4 octets are more infrequent, occurring with $\sim 8\,\%$ of probability each. The most probable octets lie near local maxima of the joint PDF ρ_{Y_1,Y_2,Y_3} of the optimized triad, as deduced from the PDF iso-surface shown in Fig. 12. The twisted PDF iso-surface presents similarities to those of a typical perfect non-Gaussian triad as in Fig. 2. Finally, Theorem 1 given by Eq. (3), about the invariance of total negentropy over the rotations space, states in this case that while the optimized triad in wave resonance (corresponding to $\boldsymbol{Y}_{\mathrm{proj}}$) maximizes NGty, the remaining PC components not included in the triad tend to be closer to Gaussians.

5 Discussion and conclusions

Multivariate non-Gaussian PDFs of climate and geofluid data allow for the possibility of nonlinear correlations between two, three or even more mutually linearly uncorrelated variables. Those non-trivial correlations, or interactions in latu sensu, make variables statistically dependent leading to positive mutual information and multi-information values, shared by two, three or more variables. The triadic MII, among three variables, can be totally explained by pairs (redundant triad) or, on the opposite situation, by tripartite synergies (perfect triad) i.e. by an emerging property across the three pair-

wise independent variables (e.g. a triangle of lovers who meet together only in pairs and in an equitable way). In that situation, the MII is reduced to the interaction information (IT), measuring the synergy level between variables. Perfect triads are impossible in the Gaussian world since pair-wise uncorrelation is sufficient for global independence.

An example of perfect triad between binary random variables (Y_1, Y_2, Y_3) occurs when any of the variables is the result of a Boolean exclusive disjunction and/or a logic equivalence of the two others. In both cases the look-up table of results is a Latin-Square (in which categories are all used but repeated, either in rows or in columns) under which IT is maximal.

For continuous uncorrelated variables, used in the paper, IT expands as function of cross cumulants of order greater-than-three (vanishing for Gaussian PDFs) and mixing the three variables together. The occurrence of perfect triads is quite extreme and difficult to observe in nature because of the constraints of total pair-wise independence, especially in systems of high dimensionality where time averages or interaction averaging tend to render variables more Gaussian due to the Central Limit Theorem. Consequently, nonlinear correlations, non-Gaussianity and triadic interactions are somehow hidden in a multi-dimensional system, hence they can only be put in evidence by variable transformations.

Therefore, in order to optimize triads and maximize their IT, we develop and test a methodology of finding highly synergetic non-Gaussian triads: (Y_1, Y_2, Y_3) in the space of orthogonal rotations of uncorrelated spherized variables (e.g. normalized PCs). Uncorrelatedness and total negentropy is kept invariant over that rotations space (Theorem 1). The direct estimation of IT for arbitrary rotated variables issued from a high-dimensional space is quite difficult to write and to estimate due to the "curse of dimensionality". Therefore, instead of directly maximizing IT, we maximize a proxy of that relying on cross-cumulants mixing the three variables, here, the square of the cumulant $E(Y_1Y_2Y_3)$. This has the advantage of being proportional to a geometrically intuitive triadic correlation: $\mathrm{cor}_3(Y_1, Y_2, Y_3)$ between uncorrelated variables and satisfying the Schwartz inequality. Maximization of the control function is

performed by optimization gradient-descent-based methods in the space of orthogonal rotations, totally spanned by generalized Euler angles of rotation and used as control variables.

In the application, we have estimated and optimized triads for the output of a minimal fluid motion model. It is a one-dimensional model with nonlinear advection, linear dissipation and constant forcing, integrated along a latitude circle. Circular symmetry of statistics leads to a set of twice-degenerated empirical orthogonal functions that are sinusoids, each one with a certain associated wave-number. Therefore PCs are Discrete Fourier Transforms (DFTs). Non-null statistically significant values of $E(Y_1Y_2Y_3)$ and of estimated IT are obtained when the integer zonal wave-numbers k_1, k_2, k_3 satisfy the wave resonance condition (WRC): $k_1 \pm k_2 \pm k_3 = 0$, thus providing a physical meaning for the non-Gaussian statistical triads in the studied case. For the studied model, high IT values are quite closely approximated by an exclusive function of $\mathrm{cor}_3(Y_1, Y_2, Y_3)$.

Enhanced triads, issued from sets of rotated PCs satisfying WRC are thus obtained using optimization methods. Triad components are obtained as the inner product of loading fields of sinusoidal type, locally confined to the segments of the latitude circle and also satisfying WRC between corresponding number of oscillations (the local wave-numbers). Consequently, triadic interaction corresponds here to a generalized criterion and diagnostic of triadic wave-trains resonance, which comes totally out of the classical approach of the Resonance Interaction Theory (RIT) (Hammack, 1993).

The method of triads' optimization is quite general and applies to any multivariate random vector. In particular for climatic fields over the Globe triads generalize the classical concept of teleconnections relying on far-distant correlations to the new concept of non-Gaussian triadic teleconnections where three far distant uncorrelated variables play, which is deserved for works in preparation.

Appendix A: Proof of Theorem 1

For any N-dimensional RVecs $\boldsymbol{A}, \boldsymbol{B} = \mathbf{R}\boldsymbol{A}$, both with identity correlation matrices and $\mathbf{R}$ an orthogonal matrix, the following equality holds

$$J(\boldsymbol{A}) = J(\boldsymbol{B}) \tag{A1a}$$
$$= I_i(\boldsymbol{A}) + J_i(\boldsymbol{A}) \tag{A1b}$$
$$= I_i(\boldsymbol{B}) + J_i(\boldsymbol{B}). \tag{A1c}$$

It comes from immediate application of definitions. For the concatenation $\boldsymbol{A} = \boldsymbol{A}_1 \| \boldsymbol{A}_2$ we also get

$$I_i(\boldsymbol{A}) = I_i(\boldsymbol{A}_1) + I_i(\boldsymbol{A}_2) + I_i(\boldsymbol{A}_1, \boldsymbol{A}_2); \tag{A2a}$$
$$J_i(\boldsymbol{A}) = J_i(\boldsymbol{A}_1) + J_i(\boldsymbol{A}_2). \tag{A2b}$$

Theorem 1 (2), follows from application of Eq. (A2a) to $\boldsymbol{Y}_{\text{rot}} = \boldsymbol{Y}_{\text{proj}} \| \boldsymbol{Y}_{\text{comp}}$ and Eq. (A2b) to $J(\boldsymbol{Y}_{\text{rot}})$. Then, Eq. (A1) is applied to $\boldsymbol{Y}_{\text{rot}} = \mathbf{R}\boldsymbol{X}_{\text{u, signal}}$ to expand $I_i(\boldsymbol{Y}_{\text{rot}})$.

Appendix B: Multi-information and interaction information estimators

Many MII estimators have been tested in literature (e.g. histogram-based, kernel-based, adaptative grids, next nearest neighbors, maximum-entropy-based) (Walters-Williams and Li, 2009 for a thorough review). Here, we use a kernel-based PDF estimator where single variables of the original working vector $\boldsymbol{U} \in \mathbb{R}^D$ (e.g. the $\boldsymbol{Y}_{\text{proj}}$ vector introduced in Sect. 2 and $D = 3$ for triads) have been subjected to Gaussian anamorphoses, converting them into standard Gaussians N (0,1), collected in the transformed vector $\boldsymbol{U}_{\text{g}} \equiv \boldsymbol{G}(\boldsymbol{U}) \in \mathbb{R}^D$. In practice and for a multivariate sample of N_{dof} iid realizations and a generic component U of $\boldsymbol{U}$, we take the discrete anamorphosis of

the lth increasing sorted term $U_{(l)}$ as:

$$U_{g,(l)} = \Phi^{-1}[l/(N_{\text{dof}} + 1)]; \quad l = 1, \ldots, N_{\text{dof}} \tag{B1a}$$
$$= G(U_{(l)}); \quad l = 1, \ldots, N_{\text{dof}}, \tag{B1b}$$

where Φ^{-1} is the inverse of the mass probability function of the standard Gaussian. Then, thanks to the MII equality $I_i(\boldsymbol{U}) = I_i(\boldsymbol{U}_{\text{g}})$, we estimate it directly from Gaussianized data. This procedure regularizes the effect of marginal outliers in the MII estimation. Since marginal PDFs are Gaussian, i.e. $\rho_{Y_{\text{g}}}(U_{\text{g}}) = \varphi(U_{\text{g}}) \equiv (2\pi)^{-1/2} \exp(-U_{\text{g}}^2/2)$, the MII formula (1) becomes:

$$I_i(\boldsymbol{U}) = I_i(\boldsymbol{U}_{\text{g}}) \tag{B2a}$$
$$= \int_{\mathbb{R}^D} \rho_{\boldsymbol{U}_{\text{g}}}(\boldsymbol{U}_{\text{g}}) \ln[\rho_{\boldsymbol{U}_{\text{g}}}(\boldsymbol{U}_{\text{g}})] \mathrm{d}\boldsymbol{U}_{\text{g}} + (D/2)[\ln(2\pi) + 1]. \tag{B2b}$$

The above formula has the advantage of uniquely using the joint PDF, while marginals are kept fixed, thus reducing the MII estimator's variance.

Now the PDF $\rho_{\boldsymbol{U}_{\text{g}}}(\boldsymbol{U}_{\text{g}})$ is estimated using a Kernel-based method (Silverman, 1986). Then the integral of MII is computed by Gauss quadrature rule of integration with Hermite weights. The joint PDF of Gaussianized variables is written as an average of localized Gaussian Kernel functions $\rho_{\boldsymbol{U}_{\text{g}}}(\boldsymbol{U}_{\text{g}}) = N_{\text{dof}}^{-1} \sum_{l=1}^{N_{\text{dof}}} K(\boldsymbol{U}_{\text{g}}, \boldsymbol{U}_{\text{g},l})$ with

$$K(\boldsymbol{U}_{\text{g}}, \boldsymbol{U}_{\text{g},l}) = \frac{1}{(2\pi)^{D/2} h^D [\det(\mathbf{S}_{\text{g}})]^{1/2}} \exp\left[-\frac{(\boldsymbol{U}_{\text{g}} - \boldsymbol{U}_{\text{g},l})' \mathbf{S}_{\text{g}}^{-1} (\boldsymbol{U}_{\text{g}} - \boldsymbol{U}_{\text{g},l})}{2h^2} \right], \tag{B3}$$

satisfying the normalization: $\int_{\mathbb{R}^D} K(\boldsymbol{U}_{\text{g}}, \boldsymbol{U}_{\text{g},l}) \mathrm{d}\boldsymbol{U}_{\text{g},l} = 1$, where $\boldsymbol{U}_{\text{g},l} = (U_{\text{g},1l}, U_{\text{g},2l}, \ldots, U_{\text{g},Dl})'$, $l = 1, \ldots, N_{\text{dof}}$ are the N_{dof} realizations of Gaussianized vectors expanded in terms of the D components, $\mathbf{S}_{\text{g}}$ is the $D \times D$ estimated covariance

matrix of $\boldsymbol{U}_g$ and $h = \{4/[N_{dof}(D+2)]\}^{[1/(D+4)]}$ is the "optimal" Gaussian bandwidth. The MII is a D-dimensional multiple integral which is evaluated here through the Gauss quadrature integration formula using the Hermite weight function. This is written for a generic integrable function $F(\boldsymbol{U}_g) \equiv F(U_{g,1}, \ldots, U_{g,D})$ as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{+\infty} F(U_{g,1}, \ldots, U_{g,D}) \mathrm{d}\boldsymbol{U}_g \approx \sum_{i_1=1}^{Q} \cdots \sum_{i_D=1}^{Q} F(Q_{i_1}, \ldots, Q_{i_D}) \exp\left(Q_{i_1}^2 + \ldots + Q_{i_D}^2\right) W_{i_1} \ldots W_{i_D} \tag{B4}$$

where Q is the number of quadrature points: $Q_1, \ldots, Q_Q \in \mathbb{R}$ and $W_1, \ldots, W_Q$ are the weights. The formula is exact for $F(\boldsymbol{U}_g)\exp(-\|\boldsymbol{U}_g\|^2)$ being a multivariate polynomial of degree less than $2Q$. For the MII estimation, $F(\boldsymbol{U}_g) = \rho_{U_g}(\boldsymbol{U}_g)\ln[\rho_{U_g}(\boldsymbol{U}_g)]$ is used in the quadrature formula where the PDF was previously estimated. Thus the used estimator writes as:

$$I(\boldsymbol{U}_g) \approx I_{\mathrm{Quad}}(\boldsymbol{U}_g) \tag{B5a}$$

$$\equiv \sum_{i_1=1}^{Q} \cdots \sum_{i_D=1}^{Q} \rho_{U_g}(\boldsymbol{Q}) \log\left[\rho_{U_g}(\boldsymbol{Q})\right] \exp(\|\boldsymbol{Q}\|^2) W_{i_1} \ldots W_{i_D} + (D/2)\,[\ln(2\pi) + 1], \tag{B5b}$$

where $\boldsymbol{Q} \equiv (Q_{i_1}, \ldots, Q_{i_D})'$ (Eq. B5) is a D-dimensional sum indexed by D indices $i_1, \ldots, i_D$. For the kind of integrand discrimination, we have taken $Q = 20$.

In the paper the dimension is $D = 2$ or 3. This estimator has shown fairly good convergence speed for PDFs being close to multivariate Gaussians or not having sharp gradients or discontinuities. Furthermore, the estimator's performance is better under small values of compactness (i.e. data concentration along a manifold), since it leads to smoother PDFs. The biases of I_{Quad} are evaluated here for the case where the standardized rotated vector (of uncorrelated components) $\boldsymbol{U}$ follows a multivariate

isotropic standard Gaussian $N(0, \mathbf{I}_D)$ leading to the vanishing of MII. Biases shall not differ much when the true PDF is close to Gaussian. For that purpose, we have used the Monte-Carlo methodology by generating an ensemble of 1000 samples of N_{dof} iid realizations of $\boldsymbol{U}$. Then statistics (95 %-quantile Q_{95} and the average or bias B) of I_{Quad} at dimension $D = 2$ and 3, were computed from the ensemble. Those statistics are obtained for the multiplicative sequence $N_{dof} = 25$, 50, 100, 200, 400, 800, 1600, 3200, 6400 and 12 800, collected altogether in the log-log graphic of Fig. B1. The 95 % quantile gives the 5 % significance-level uni-directional threshold of rejection of the null hypothesis H_0 supposing that the standardized rotated vector (of uncorrelated components) $\boldsymbol{U}$ have composed of totally independent components. Then, only the values above that are considered statistically significant. Apparently there is a good linear logarithmic fit both for B and Q_{95} of the type C/N_{dof}^m where C and m are positive constants. The bias B of the interaction information (IT) $I_3(\boldsymbol{U}_3)$, $\boldsymbol{U}_3 \in \mathbb{R}^3$ is also computed, along with the quantile Q_{95} appearing in Fig. B2. It is interesting to note that for small (large) N_{dof}, the bias of IT is negative (positive), overestimating redundancy (synergy).

Acknowledgements. This research was developed at IDL with the support of the Portuguese Foundation for Science and Technology (FCT) through project PTDC/GEOMET/3476/2012 – Predictability assessment and hybridization of seasonal drought forecasts in Western Europe (PHDROUGHT). Rui Perdigão also acknowledges support from the European Research Council under the ERC Advanced Grant "FloodChange", project no. 291152. Thanks are also due to Pedro Miranda and Ricardo Trigo for some discussions, to Andreia Ribeiro and our families for the omnipresent support.

References

Abramov, R., Majda, A. J., and Kleeman, R.: Information theory and predictability for low-frequency variability, J. Atmos. Sci., 62, 65–87, doi:10.1175/JAS-3373.1, 2005.

Almeida, L.: MISEP – linear and nonlinear ICA based on mutual information, J. Mach. Learn. Res., 4, 1297–1318, 2003.

Almeida, L. B.: Nonlinear Source Separation, Synthesis Lectures on Signal Processing, Morgan and Claypool Publishers, San Rafael, California (USA), 114 pp., doi:10.2200/S00016ED1V01Y200602SPR002, 2006.
Bailey, R. A.: Orthogonal partitions for designed experiments, Design Code Cryptogr., 8, 45–77, 1996.
Ball, F. K.: Energy transfer between external and internal gravity waves, J. Fluid Mech., 19, 465–478, 1964.
Bartello, P.: Geostrophic adjustment and inverse cascades in rotating stratified turbulence, J. Atmos. Sci., 52, 4410–4428, doi:10.1175/1520-0469(1995)052<4410:GAAICI>2.0.CO;2, 1995.
Batchelor, G. K.: The Theory of Homogeneous Turbulence, Cambridge University Press, UK, 197 pp., 1953
Bellman, R. E., Rand Corporation: Dynamic Programming, Princeton University Press, Princeton, USA, 1957.
Bocquet, M., Pires, C. A., and Lin, W.: Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, doi:10.1175/2010MWR3164.1, 2010.
Comon, P.: Independent component analysis, a new concept?, Signal Process., 36, 287–314, 1994.
Cover, T. M. and Thomas, J. A.: Elements of Information Theory, John Wiley & Sons Inc., New York, USA, 1991.
DelSole, T.: Predictability and information theory. Part I: Measures of predictability, J. Atmos. Sci., 61, 2425–2440, 2004.
Dénes, J. H. and Keedwell, A. D.: Latin Squares and their Applications, Academic Press, New York, London, 547 pp., 1974.
Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, Eur. Phys. J.-Spec. Top., 174, 157–179, 2009.
Evenson, G. and Fario, N.: Solving for the generalized inverse of the Lorenz model, J. Meteorol. Soc. Jpn., 75, 229–243, 1997.
Franzke, C., Majda, A. J., and Branstator, G.: 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, 2007.
Friedman, J. H. and Tukey, J. W.: A projection pursuit algorithm for exploratory data analysis, IEEE T. Comput., 23, 881–890, 1975.

Gilbert, J. C. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, Math. Program., 45, 407–435, 1989.
Globerson, A., Stark, E., Vaadia, E., and Tishby, N.: The minimum information principle and application to neural code analysis, P. Natl. Accd. Sci. USA, 106, 3490–3495, 2009.
Granger, C. W. J.: Investigating causal relations by econometric models and cross-spectral methods, Econometrica, 37, 424–438, 1969.
Hannachi, A. and O'Neill, A.: Atmospheric multiple equilibria and non-Gaussian behaviour in model simulations, Q. J. Roy. Meteor. Soc., 127, 939–958, doi:10.1002/qj.49712757312, 2001.
Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: a review, Int. J. Climatol., 27, 1119–2115, 2007.
Hastie, T. and Stuetzle, W.: Principal curves, J. Am. Stat. Assoc., 84, 502–516, 1989.
Hastie, T., Tibshirani, R., and Friedman, J.: Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer-Verlag, New York, USA, 2001.
Hlinka, J., Hartman, D., Vejmelka, M., Novotná, D., and Palus, M.: Nonlinear dependence and teleconnections in climate data: sources, relevance, nonstationarity, Clim. Dynam., 42, 1873–1886, 2014.
Hoffman, D., Raffenetti, R. C., and Ruedenberg, K.: Generalization of Euler angles to *N*-dimensional orthogonal matrices, J. Math. Phys., 13, 528, doi:10.1063/1.1666011, 1972.
Hsieh, W. W., Wu, A., and Shabbar, A.: Nonlinear atmospheric teleconnections, Geophys. Res. Lett., 33, L07714, doi:10.1029/2005GL025471, 2006.
Huber, P. J.: Projection pursuit, Ann. Stat., 13, 435–475, doi:10.1214/aos/1176349519, 1985.
Hyvärinen, A. and Oja, E.: Independent component analysis: algorithms and application, Neural Networks, 13, 411–430, 2000.
Jakulin, A. and Bratko, I.: Quantifying and Visualizing Attribute Interactions: an Approach Based on Entropy, arXiv:cs/0308002v3[cs.AI], 308002, 3 pp., 2004.
Jennrich, R. I.: A simple general procedure for orthogonal rotation, Psychometrika, 66, 289–306, 2001.
Liang, X. S. and Kleeman, R.: A rigorous formalism of information transfer between dynamical system components. I. Discrete mapping, Physica D, 231, 1–9, 2007.
Lorenz, E. N.: Predictability: a problem partly solved, in: Seminar on Predictability, vol. I, ECMWF, Reading, UK, 1–18, available at: http://old.ecmwf.int/publications/library/

ecpublications/_pdf/seminar/1995/predictability_lorenz.pdf, last access: 23 September 2014, 1995.
Lorenz, E. N. and Emanuel, K. A.: Optimal sites for supplementary weather observations: simulation with a small model, J. Atmos. Sci., 55, 399–414, 1998.
Majda, A. J., Abramov, R., and Grote, M.: Information Theory and Stochastics for Multiscale Nonlinear Systems, vol. 25 of CRM Monograph Series, American Mathematical Society, Centre de Recherches Mathématiques, Université de Montréal, 141 pp., 2005.
McGill, W. J.: Multivariate information transmission, Psychometrika, 19, 97–116, 1954.
Monahan, A. H.: Nonlinear principal component analysis: tropical Indo–Pacific sea surface temperature and sea level pressure, J. Climate, 14, 219–233, 2001.
Monahan, A. H. and DelSole, T.: Information theoretic measures of dependence, compactness, and non-gaussianity for multivariate probability distributions, Nonlin. Processes Geophys., 16, 57–64, doi:10.5194/npg-16-57-2009, 2009.
Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, USA, 1988.
Pires, C. A.: Source separation by non-Gaussian dyads of the low-frequency atmospheric variability, in preparation, 2014.
Pires, C. A. and Perdigão, R. A. P.: Non-Gaussianity and asymmetry of the winter monthly precipitation estimation from the NAO, Mon. Weather Rev., 135, 430–448, 2007.
Pires, C. A. and Perdigão, R. A. P.: Minimum mutual information and non-Gaussianity through the maximum entropy method: theory and properties, Entropy, 14, 1103–1126, doi:10.3390/e14061103, 2012.
Pires, C. A. and Perdigão, R. A. P.: Minimum mutual information and non-Gaussianity through the maximum entropy method: estimation from finite samples, Entropy, 15, 721–752, doi:10.3390/e15030721, 2013.
Parametrization of an orthogonal matrix in terms of generalized eulerian angles, Int. J. Quantum Chem., 4, 625–634, doi:10.1002/qua.560040725, 1969.
Roulston, M. and Smith, L.: Evaluating probabilistic forecasts using information theory, Mon. Weather Rev., 130, 1653–1660, 2002.
Schneidman, E., Still, S., Berry, M. J., and Bialek, W.: Network information and connected correlations, Phys. Rev. Lett, 91, 238701-1–238701-4, 2003.
Shannon, C. E.: A mathematical theory of communication, Bell Syst. Tech. J., 27, 379–423, 623–656, 1948.

Sura, P., Newman, M., Penland, C., and Sardeshmuck, P.: Multiplicative noise and non-Gaussianity: a paradigm for atmospheric regimes?, J. Atmos. Sci., 62, 1391–1406, 2005.
Stephenson, D. B., Hannachi, A., and O'Neill, A.: On the existence of multiple climate regimes, Q. J. Roy. Meteor. Soc., 130, 583–605, 2004.
Theis, F. J.: Multidimensional independent component analysis using characteristic functions, in: Proceedings of 13th European Signal Processing Conference (EUSIPCO), Antalya, Turkey, 4–8 September, 1692–1695, 2005.
Theis, F. J.: Towards a general independent subspace analysis, in: Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 4–7 December 2006
Theis, F. J. and Kawanabe, M.: Uniqueness of non-Gaussian subspace analysis, in: Independent Component Analysis and Blind Signal Separation, edited by: Rosca, J., Erdogmus, D., Príncipe, J. C., and Haykin, S., 6th International Conference, ICA 2006, Charleston, SC, USA, 5–8 March 2006, Lecture Notes in Computer Science, Springer, Berlin, 3889, 2006.
Timme, N., Alford, W., Flecker, B., and Beggs, J. M.: Synergy, redundancy, and multivariate information measures: an experimentalist's perspective, J. Comput. Neurosci., 36, 119–140, doi:10.1007/s10827-013-0458-4, 2013.
Trendafilov, N. T.: Dynamical system approach to multivariate data analysis, J. Comput. Graph. Stat., 15, 628–650, 2006.
Tsujishita, T.: On triple mutual information, Adv. Appl. Math., 16, 269–274, 1995.
van den Dool, H. M. and Saha, S.: Empirical orthogonal teleconnections, J. Climate, 18, 1421–1435, doi:10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2, 2000.
van Hulle, M. M.: Edgeworth approximation of multivariate differential entropy, Neural Comput., 17, 1903–1910, 2005.
van Leeuwen, P. J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. Roy. Meteor. Soc., 136, 1991–1999, 2010.
Wackernagel, H.: Multivariate Geostatistics – an Introduction with Applications, Springer Verlag, Berlin, Germany, 1995.
Wallace, J. M. and David, S. G.: Teleconnections in the geopotential height field during the Northern Hemisphere winter, Mon. Weather Rev., 109, 784–812, doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2, 1981.

Walters-Williams, J. and Li, Y.: Estimation of mutual information: a survey, Lect. Notes Comp. Sci., 5589, 389–396, 2009.

Wolter, K. and Timlin, M. S.: Measuring the strength of ENSO events – how does 1997/98 rank?, Weather, 53, 315–324, 1998.

Xu, Q.: Measuring information content from observations for data assimilation: relative entropy versus shannon entropy difference, Tellus, 59, 198–209, doi:10.1111/j.1600-0870.2006.00222.x, 2007.

Ziman, J. M.: Electrons and Phonons, Clarendon Press, Oxford, 1960.

Zupanski, D. A. Y., Hou, S. Q., Zhang, M., Zupanski, C., Kummerow, D., and Cheung, S. H.: Application of information theory in ensemble data assimilation, Q. J. Roy. Meteor. Soc., 133, 1533–1545, 2007.

Table 1. Lookup tables of three Boolean variables maximizing the interaction information. The left and right (2 × 2) logic tables correspond respectively to the exclusive disjunction and equivalency Boolean relationships.

1	1	2	2	1
2	2	1	1	2
Y_{1c}, Y_{2c}	1	2	1	2

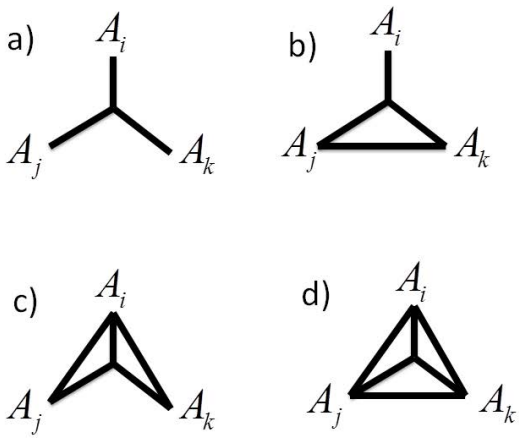

Figure 1. Schemes of triad types and appropriate interactivity measures **(a)** perfect, $I_3(A_i, A_j, A_k)$; **(b)** asymmetric, $I(A_j, A_k|A_i)$; **(c)** asymmetric $I[(A_j, A_k), A_i]$; **(d)** redundant $I(A_j, A_k, A_i)$. The triangle edges represent dyadic dependencies and the central star the IT.

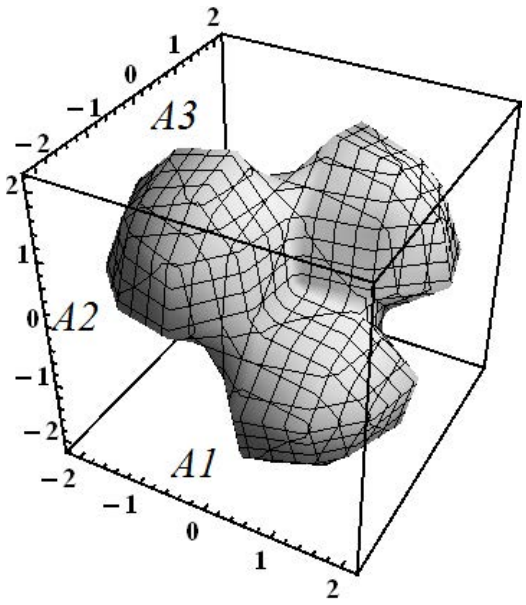

Figure 2. Iso-surface 0.01 of the PDF (13) of a perfect triad with 1-D and 2-D Gaussian projections.

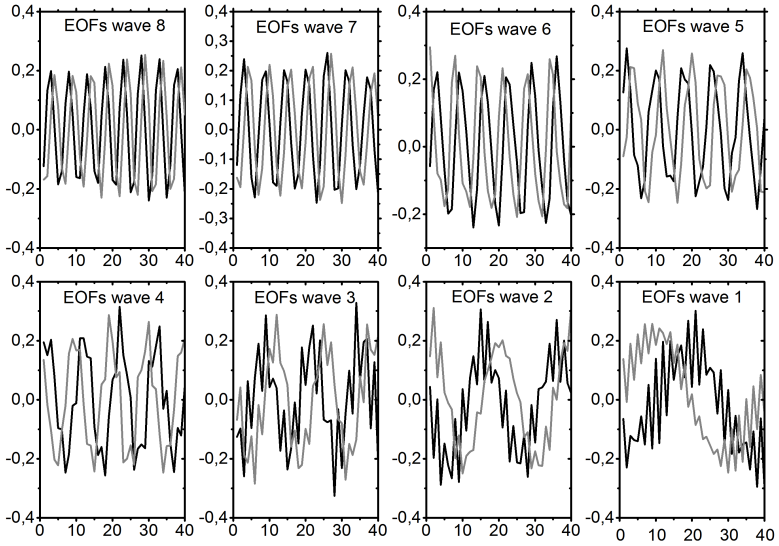

Figure 3. EOFs of Lorenz-95 model corresponding to wave-numbers 1–8.

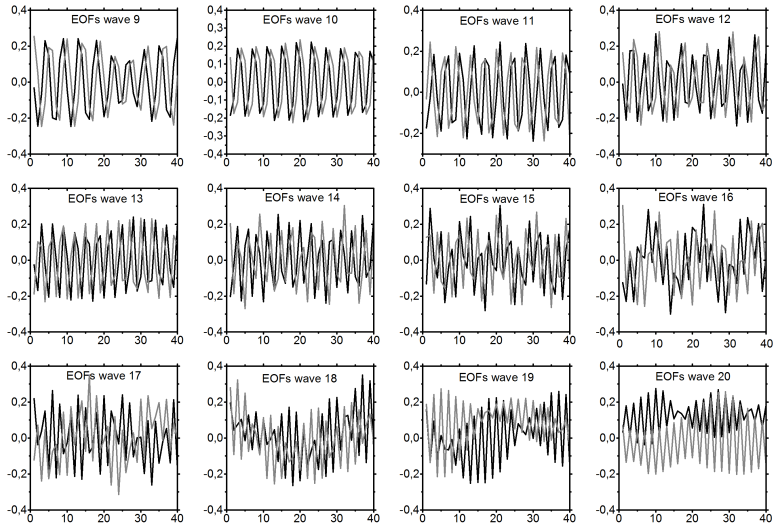

Figure 4. EOFs of Lorenz-95 model corresponding to wave-numbers 9–20.

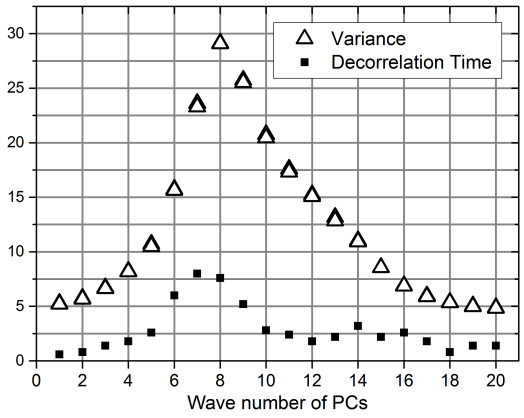

Figure 5. PC variances and decorrelation times as a function of corresponding wave-number.

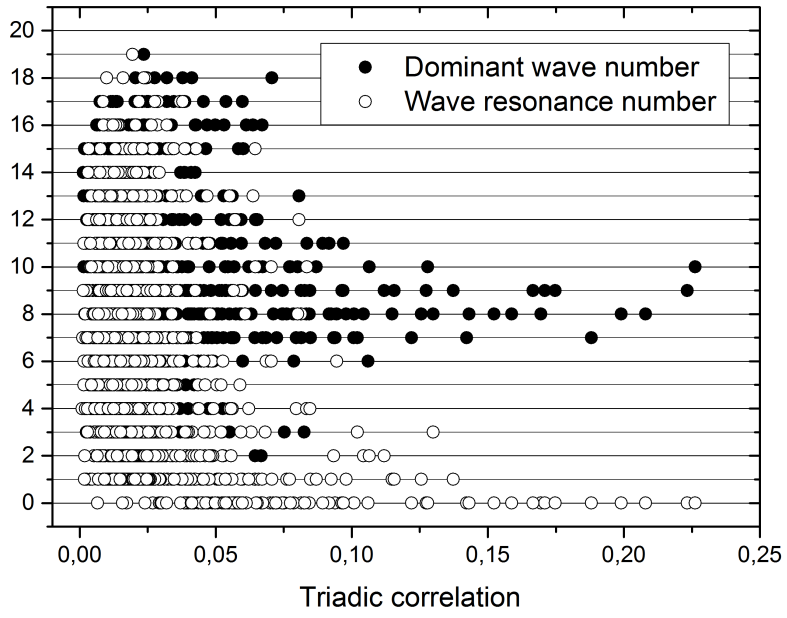

Figure 6. Scatter-plots of the dominant wave-number $P_{\mathrm{RS}}(p,q,r)$ and of the wave resonance number $\delta_{\mathrm{RS}}(p,q,r)$ as function of the triadic correlation $C_{\mathrm{RS}}(p,q,r)$ for each triad of PCs.

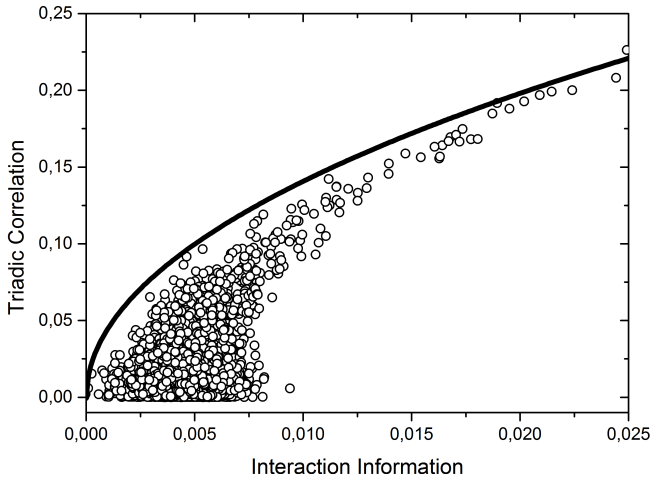

Figure 7. Triadic correlation (open circles) and equivalent triadic correlation (solid line) for each triad of PCs.

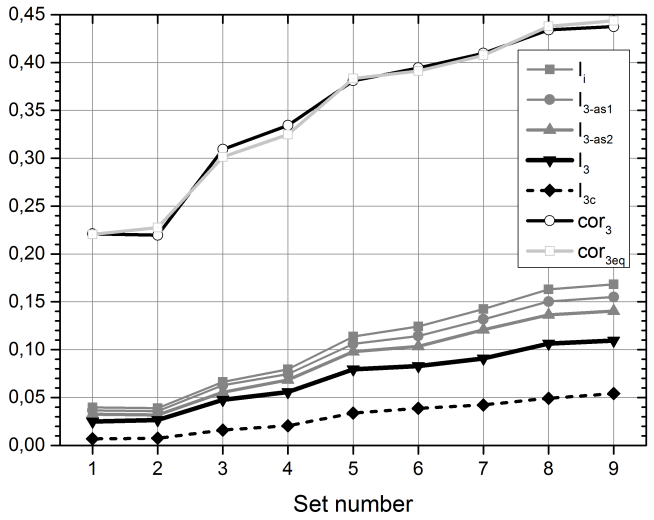

Figure 8. Absolute triadic correlation (cor_3), equivalent triadic correlation (cor_{3eq}), interaction information (I_3), multi-information (I_i), upper bounds of I_3 due to triadic asymmetry ($I_{3\text{-}as1}$, $I_{3\text{-}as2}$) (Eq. 10a–b) and categorical interaction information (I_{3c}) (Eq. 22a–b) for the optimized triad obtained for each of the 9 chosen sets of PCs.

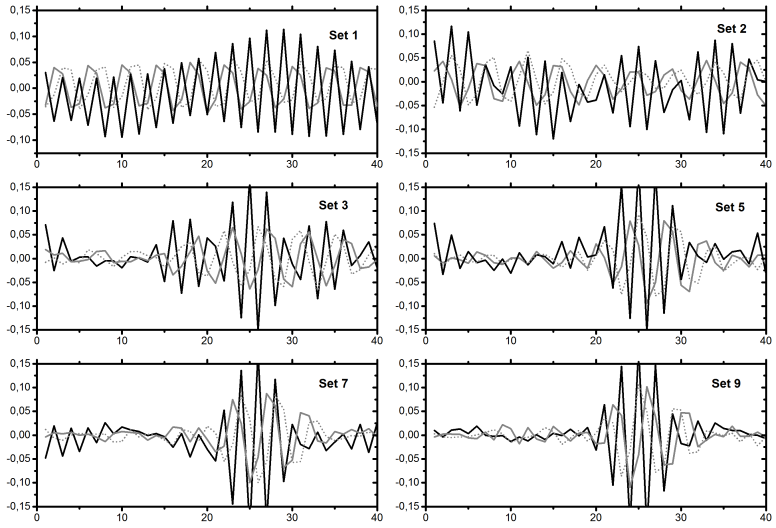

Figure 9. Loading fields of the optimized triads for sets 1, 2, 3, 5, 7, 9 of rotated normalized PCs.

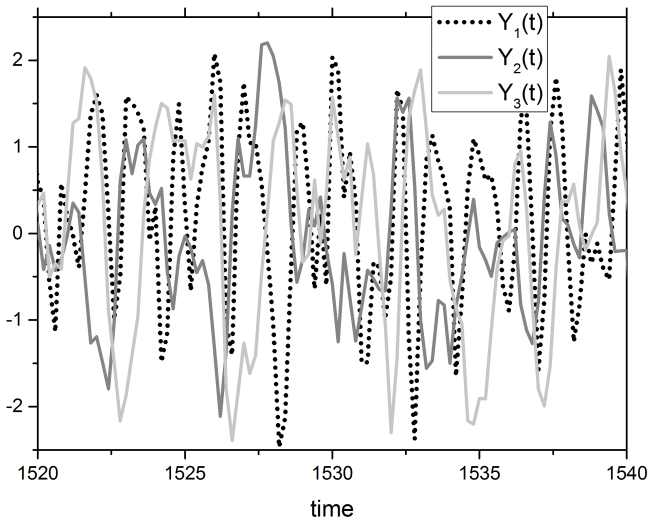

Figure 10. Time-series of the triad components during an interval of 20 time units for the triad optimized with set 9 of PCs.

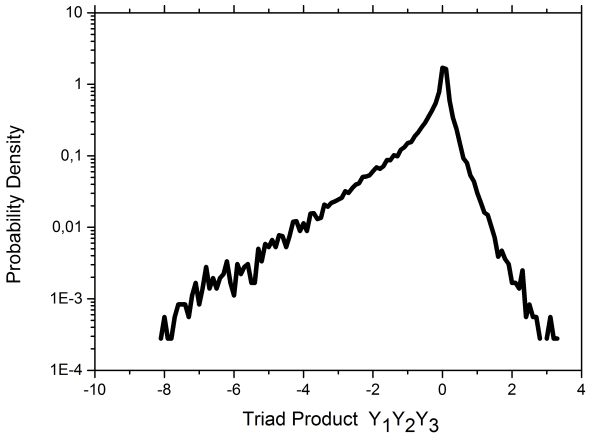

Figure 11. Normalized histogram (in logarithmic scale) of the product of optimized triad components for the triad optimized with set 9 of PCs.

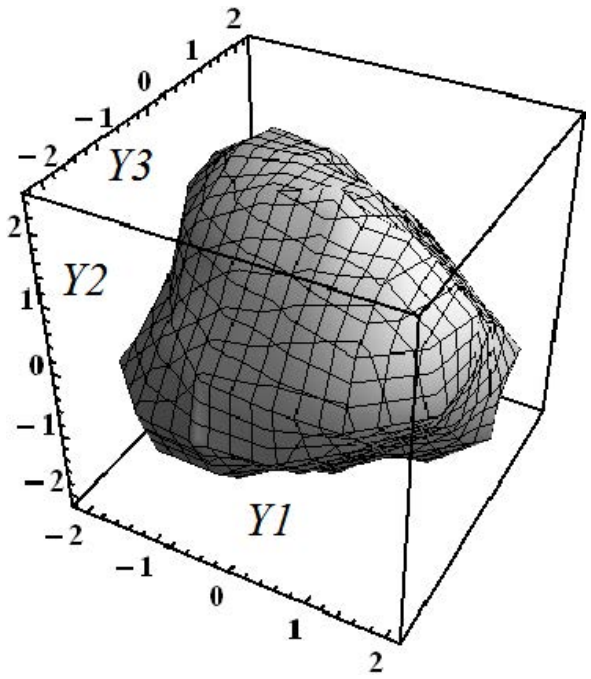

Figure 12. Iso-surface 0.001 of the PDF of triad components for the optimized triad with set 9 of PCs. Marginals are Gaussianized.

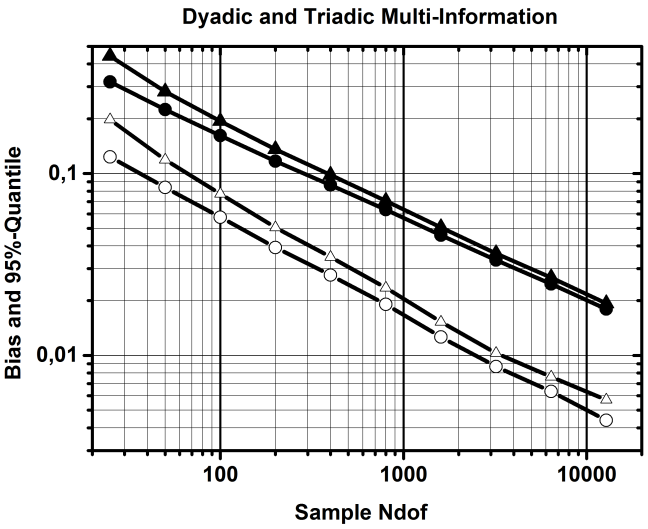

Figure B1. Bias (circles) and 95 % quantile (triangles) of the estimators of the dyadic MII (open symbols) and of the triadic MII (solid symbols).

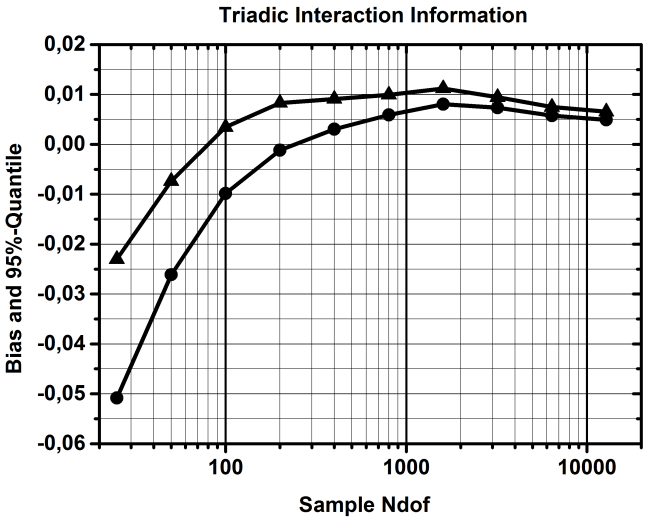

Figure B2. Bias (circles) and 95 % Quantile (squares) of the estimator of the triadic IT.