

The Assumption of Non-Gaussianity in Natural and Social Sciences and Its Influence on Detection of Causal Relationships

Kateřina Hlaváčková-Schindler
*Commission for Scientific Visualization, Austrian Academy of Sciences
Austria*

1. Introduction

The concept of causality is changing as human knowledge changes. Causality as an abstract notion has been traditionally studied in the field of metaphysics in philosophy. The Greek philosophers understood the time causality as explanation in general (Aristotle, 350 B.C.). The search for causes was a search for "first principles", which were meant to be explanatory. In the more recent philosophy introduced by (Newton, 1687) was causality connected with determinism. The current experimental science reveals the non-deterministic notion of cause, which has to be also taken into consideration. The introduction of probability theory into all scientific disciplines allows to formalize and mathematize the wider conceived notion of cause.

This paper does not deal with the philosophical approach to causality, to this we refer the reader for example to the works (Mackie, 1988), (Hume, 1896), (Russo, 2009) or (Pearl, 2000). Here we deal exclusively with the formal mathematical approaches to detect cause-effect relationships, namely with Granger causality and transfer entropy and their application in sciences.

The generally non-deterministic approaches to causality apply various probability distributions to model the real-world phenomena. The selection of an appropriate or inappropriate model to fit the real world data has obviously an important influence on the credibility of the achieved conclusions.

In the present paper we discuss the influence of the selection of a data model for detection of causal relationships between two or more time series. We focus here on cases when the Gaussianity of the investigated process can be assumed and when not. The causality detection methods considered here are Granger causality (Granger, 1969) and transfer entropy (Schreiber, 2000). We investigated time series with a wider class probability distributions than Gaussian, the generalized Gaussian probability distributions. These distributions are given parametrically. We set conditions on their parameters so that one can from their values decide whether the relationships between the involved time series are unidirectional causal or whether no causality is present.

Being aware of outstanding philosophical papers on causality in the sciences, for example (Illari et al., 2011), we are though not aware of any similar publication on mathematically

conceived causality and their application in natural and social sciences, neither of any analysis of the probabilistic assumptions about the investigated time series and their influence on causality detection by Granger causality or transfer entropy.

The paper is organized as follows. The application of various probabilistic models in natural and social sciences is discussed in Section 2. Granger causality and its application in natural and social sciences is treated in Section 3. Section 4 deals with transfer entropy, directionality index and other information-theoretical measures, inclusive their applications. Section 5 devotes a special attention to causal relationships among Gaussian time series and generalized Gaussian time series. In Section 6 we presented original and simple criteria for phenomena having generalized Gaussian distributions given parametrically which decide about a presence or absence of causality between these phenomena given by concrete time series. Section 7 discussed the criteria deciding about the applied causality detection method and concludes with the importance of the achieved results.

We presented original and simple criteria for phenomena having generalized Gaussian distributions given parametrically which decide about a presence or absence of causality between these phenomena given by concrete time series.

2. Probabilistic distributions and their application in natural and social sciences

Gaussian, more frequently called "normal" distribution has a special position among all probability distributions used in data modeling and is the most popular. It has been known for a relatively long time, is simple and analytically tractable. Its symmetry about its mean value is one of the basic principles realized in nature as well as in human culture. The bell shape of its graph makes the normal distribution attractive for modeling of real world data in many scientific or social disciplines.

Indeed, many common natural or social phenomena show to have normal distribution. For example, such phenomena as women's height, Brownian motion of particles, milk production by cows and random deviations from target values in industrial processes fit a normal distribution (Limpert et al., 2001). However, many phenomena which fit normal distribution, have been shown that they fit also log-normal distribution or generalized normal distribution or, more precisely, fit it even better. What is the difference between normal and log-normal distribution? Both forms of variability are based on a variety of forces (causes) acting independently of one another. A major difference is however that the effects can be additive or multiplicative, thus leading to normal or log-normal distributions, respectively (Limpert et al., 2001).

The length of spoken words in phone conversation (Herdan, 1958), the length of sentences (Williams, 1940) have been shown to have log-normal distribution, as well as the age of marriage (Preston, 1981) or income (Statistical yearbook in Switzerland, 1997). Prices, incomes or populations, i.e. phenomena which grow exponentially, are often skewed to the right, and hence may be better modeled by other distributions than by the normal one, such as the log-normal distribution, Pareto distribution or skewed generalized normal distribution. Statistical inference using a normal distribution is not robust to the presence of outliers. When outliers are expected, data may be better described using a heavy-tailed distribution such as the Student's t-distribution.

Generalized normal distribution (generalized Gaussian distribution), first time mathematically defined in (Nadarajah, 1995) can model for example Brownian motion of particles or fractional Brownian motion with a better precision than the normal distribution (Zinde-Walsh & Phillips, 2003). Other experiments have shown a better approximation precision of the generalized Gaussian distributions than of the Gaussian distributions, for example (Sharifi & Leon-Garcia, 1995), (Moulin & Liu, 1999) for in image processing and video analysis, (Bicego et al., 2008) for EEG time series modeling.

The modeling in linguistics applies mostly Gaussian mixtures. Mixtures of generalized Gaussian distributions have been recently used in text independent speaker identification (Sailaja et al., 2010) and showed that it outperforms the earlier existing text independent speaker identification models. This model was applied for speaker identification like voice dialing, banking by telephone, telephone shopping information services etc.

Exponential distribution has been frequently used in modeling in astrophysics, for example the Weinman exponential distribution has been shown to be a good model for dusty galactic discs (Misiriotis et al., 2000).

To summarize, other probability distributions than the normal one have an important role in modeling both in natural and social sciences. We will call them non-Gaussian distributions in the following. The selection of a correct distributions for modeling natural or social phenomena is of great importance, especially when mutual interactions among these phenomena are investigated. A crucial question is whether there are causal relationships among the studied phenomena. This leads to a formal definition of causality and causal measures.

In the following chapters we define formally two causality detection measures, namely the Granger causality and transfer entropy.

3. Granger causality

The introduction of the concept of causality into the experimental science, namely into analyses of data observed in consecutive time instants (time series), is due to C.W.J. Granger in (Granger, 1969), the 2003 Nobel prize winner in economy. In his Nobel lecture (Granger, 2003) he recalled the inspiration by the Wiener's work and identified two components of the statement about causality:

1. The cause occurs before the effect; and
2. The cause contains information about the effect that is unique, and is in no other variable.

As Granger put it, a consequence of these statements is that the causal variable can help to forecast the effect variable after other data has been first used (Granger, 2003). This restricted sense of causality, referred to as *Granger causality*, G-causality thereafter, characterizes the extent to which a process X_t is leading another process Y_t , and builds upon the notion of incremental predictability. It is said that the *process X_t Granger causes another process Y_t* if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t . The standard test of G-causality developed in (Granger, 1969) is based on a linear regression model.

In the following we will define Granger causality by using the notation from (Barnett, 2009). Let \oplus denotes concatenation of vectors, so that for $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_m)$

$x \oplus y$ is the $1 \times (d + m)$ vector $(x_1, \dots, x_d, y_1, \dots, y_m)$. Given jointly distributed multivariate random variables \mathbf{X} and \mathbf{Y} i.e. random vectors in R^d , we denote by $\Sigma(\mathbf{X})$ the $d \times d$ matrix of covariances $cov(X_i, Y_j)$ and by $\Sigma(\mathbf{X}, \mathbf{Y})$ the $d \times m$ matrix of cross-covariances $cov(X_i, Y_\alpha)$. Let $\Sigma(\mathbf{X}|\mathbf{Y})$ denotes the $d \times d$ matrix

$$\Sigma(\mathbf{X}|\mathbf{Y}) = \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y})\Sigma(\mathbf{Y})^{-1}\Sigma(\mathbf{X}, \mathbf{Y})^T \quad (1)$$

define when $\Sigma(\mathbf{Y})$ is invertible.

Suppose we have a stationary multivariate stochastic process \mathbf{X}_t in discrete time (i.e. marginal distributions are jointly distributed). Denote $\mathbf{X}_t^{(p)} = \mathbf{X}_t \oplus \mathbf{X}_{t-1} \oplus \dots \oplus \mathbf{X}_{t-p+1}$ for \mathbf{X} along with $p - 1$ lags so that $\mathbf{X}_t^{(p)}$ is a $1 \times pd$ random vector for each t . Given the lag p , we use the shorthand notation $\mathbf{X}_t^- = \mathbf{X}_{t-1}^{(p)}$ for the lagged variable.

Suppose we have three jointly distributed stationary multivariate stochastic processes X_t, Y_t, Z_t . Consider the regression models

$$\mathbf{X}_t = \alpha_t + (\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Z}_{t-1}^r).A + \epsilon_t \quad (2)$$

$$\mathbf{X}_t = \alpha'_t + (\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Y}_{t-1}^{(q)} \oplus \mathbf{Z}_{t-1}^r).A' + \epsilon'_t \quad (3)$$

where A and A' are the matrices of regression coefficients, α_t and α'_t are the constant terms and the random vectors ϵ and ϵ' comprise the residuals, so that so that the predictee variable \mathbf{X} is regressed firstly on the previous p lags of itself plus r lags of the conditioning variable \mathbf{Z} and secondly, in addition, on q lags of the predictor variable \mathbf{Y} . By stationarity this expression does not depend on time t , so we omit t from the notation. The G-causality of \mathbf{Y} to \mathbf{X} given \mathbf{Z} is a measure of the extent to which inclusion of \mathbf{Y} in the second model (3) reduces the prediction error of the first model (2). The standard measure of G-causality in the literature is defined for univariate predictor and predictee variables \mathbf{Y} and \mathbf{X} , and is given by the natural logarithm of the ratio of the residual variance in the restricted regression (2) to that of the unrestricted regression (3).

(Barnett, 2009) have shown that G-causality can be expressed as

$$\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} = \ln\left(\frac{\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Z}^-)}{\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-)}\right) \quad (4)$$

where \ln denotes the natural logarithm.

3.1 Extensions of Granger causality

The linear framework of Granger causality given by equations 2 and 3 has been widely applied not only in economy and finance (for a comprehensive survey of the literature see i.e. (Geweke, 1984)), but also in diverse fields of natural sciences, i.e. climatology (see (Triacca, 2005) and references therein) or neurophysiology, where specific problems of multichannel electroencephalogram recordings were solved by generalizing the Granger causality concept to multivariate case (Blinowska et al., 2004; Kamiński et. al., 2001). Nevertheless, the limitation of the present concept to linear relations required further generalizations.

Recent development in nonlinear dynamics (Abarbanel, 1993) evoked lively interactions between statistics and economy (econometrics) on one side, and physics and other natural

sciences on the other side. In the field of economy, (Baek & Brock, 1992) and (Hiemstra & Jones, 1994) proposed a nonlinear extension of the Granger causality concept. Their non-parametric dependence estimator is based on so-called correlation integral, a probability distribution and entropy estimator, developed by physicists Grassberger and Procaccia in the field of nonlinear dynamics and deterministic chaos as a characterization tool of chaotic attractors (Grassberger & Procaccia, 1983).

Another non-linear extension of Granger causality is so called correntropy (Park & Principe, 2008).

A non-parametric approach to non-linear causality testing, based on non-parametric regression, was proposed in (Bell et al., 1996). Following (Hiemstra & Jones, 1994), (Aparicio & Escribano, 1998) succinctly suggested an information-theoretic definition of causality which include both linear and nonlinear dependence.

Another nonlinear extension of the Granger causality approach was proposed by Chen et al. (Chen et al., 2004) using local linear predictors. An important class of nonlinear predictors are based on so-called radial basis functions (Broomhead & Lowe, 1988) which were used for nonlinear parametric extension of the Granger causality concept (Ancona et al., 2004; Marinazzo, 2006).

In physics and nonlinear dynamics, a considerable interest recently emerged in studying cooperative behavior of coupled complex systems (Boccaletti et al., 2002; Pikovsky et al., 2001). Synchronization and related phenomena were observed not only in physical, but also in many biological systems, i.e. (Schäfer et al., 1998; 1999) or in (Paluš et al., 2001a;b; Quyen et al., 1999; Schiff et al., 1996; Tass et al., 1998). In such physiological systems it is not only important to detect synchronized states, but also to identify drive-response relationships and thus the causality in evolution of the interacting (sub)systems. (Schiff et al., 1996) and (Quyen et al., 1999) used ideas similar to those of Granger, however, their cross-prediction models utilize zero-order nonlinear predictors based on mutual nearest neighbors. A careful comparison of these two papers (Quyen et al., 1999; Schiff et al., 1996) reveals how complex is the problem of inferring causality in nonlinear systems. While the latter two papers use the method of mutual nearest neighbors for mutual prediction, (Arnhold, 1999) proposed asymmetric dependence measures based on averaged relative distances of the (mutual) nearest neighbors.

(Ge et al, 2009) presented a novel approach which is an extension of Granger causal model and also shares the features of the bilinear approximation of dynamic causal model (David et al., 2006). The authors demonstrated face discrimination learning-induced changes in inter- and intra-hemispheric connectivity and in the hemispheric predominance of theta and gamma frequency oscillations in sheep infero-temporal cortex. The results provide the first evidence for connectivity changes between and within left and right infero-temporal cortices as a result of face recognition learning.

3.2 Application of Granger causality in natural and social sciences

As already said, the Granger causality was introduced by its author in econometry and applied by him and his followers mainly in econometry, finance and market analysis, for example in (Granger, 1969), (Poon & Granger, 2003). Other applications in humanities and social sciences are in linguistics and psychology (Gilbert & Karahalios, 2009) or demography (Feridun, 2007).

Granger causality has also been extensively applied in natural sciences, for example in medicine, especially to neuroscience (Smith et al., 2011) for the functional magnetic resonance method and (Hesse et al., 2003), analysis of EEG signals or causal interaction in neural populations (Seth & Edelman, 2007) and in many other papers. Granger causality was applied as well as in climatology (Kufmann & Stern, 1997), in cognitive and systematical biology, (Kim et al., 2011), (Fujita et al., 2010) etc.

The main drawback of Granger causality and its extensions as a model dependent method are their instability which can cause a high variability in the final estimation of errors in (2) and (3). As an alternative, we will present in the following model-free methods whose formal definitions apply information-theoretic functionals.

4. Information-theoretical causality measures

Using distributions of random processes and their definitions, introduce the information-theoretic causality measures determinism into the notion of causality.

(Paluš et al., 2001b) proposed to study synchronization phenomena in experimental time series by using the tools of information theory. Mutual information, an information-theoretic functional of probability distribution functions, is a measure of general statistical dependence. For inferring causal relation, conditional mutual information or so called transfer entropy can be used.

4.1 Transfer entropy

Transfer entropy as a non-linear causality measure was introduced in (Schreiber, 2000). It is an information-theoretic measure of time-directed information transfer between jointly dependent processes.

Let us first remind some basic definitions. The differential entropy of a (continuous) random vector \mathbf{X} taking its values in R^d with the probability density function $p(\mathbf{x})$ is defined by

$$h(\mathbf{X}) = - \int_{R^d} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}. \quad (5)$$

If \mathbf{X} is a discrete (multivariate) random variable given by a set of possible values $\{x_1, \dots, x_n\}$ then the entropy can explicitly be written as

$$H(\mathbf{X}) = - \sum_{i=1}^n p(x_i) \ln p(x_i) \quad (6)$$

where p denotes the probability mass function of X . With $\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t$ defined as before, the transfer entropy of \mathbf{Y} to \mathbf{X} given \mathbf{Z} is defined as the difference between the entropy of X conditioned on its own past and the past of \mathbf{Z} , and its entropy conditioned, in addition, on the past of \mathbf{Y} :

$$\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} = H(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Z}^-) - H(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-) \quad (7)$$

where $H(\cdot | \cdot)$ is the conditional entropy. For stationary variables, similarly as for Granger causality, the transfer entropy does not depend on t , so we omitted it from labeling. Transfer entropy is a Kullback-Leibler distance of transition probabilities.

It was shown in (Hlaváčková-Schindler et. al., 2007) that with proper conditioning, the transfer entropy is equivalent to the conditional mutual information (Paluš et al., 2001b). The latter, however, is a standard measure of information theory (Cover & Thomas, 1991). More details on the information-theoretic methods for causality detection can be found in our review paper (Hlaváčková-Schindler et. al., 2007). Marschinski and Kantz in 2002 suggested so called effective transfer entropy to reduce the bias of transfer entropy on small data sets (Marschinski & Kantz, 2002).

4.2 Transfer entropy, other information-theoretical measures and their application in natural and social sciences

Turning our attention back to econometrics, we can follow further development due to (Diks & DeGoede, 2001). They applied a nonparametric approach to nonlinear Granger causality using the concept of correlation integrals (Grassberger & Procaccia, 1983) and pointed out the connection between the correlation integrals and information theory. (Diks & Panchenko, 2005) critically discussed the previous tests of (Hiemstra & Jones, 1994). As the most recent development in economics, (Baghli, 2006) proposes information-theoretic statistics for a model-free characterization of causality, based on an evaluation of conditional entropy.

The information-theoretical approaches to causality detection are model free and can detect non-linear causal relationships, which are their advantages with respect to the approach of the linear Granger causality.

The nonlinear extension of the Granger causality based the information-theoretic formulation has found numerous applications in various fields of natural and social sciences. Let us mention just a few examples.

The Schreiber's transfer entropy has been used in climatology, i.e. (Mokhov & Smirnov, 2006; Verdes, 2005), in physiology, i.e. (Verdes, 2005), in neurophysiology, i.e. (Chávez et al., 2003) and also in analysis of financial data, i.e.(Marschinski & Kantz, 2002).

(Paluš et al., 2001a;b) applied their conditional mutual information based measures in analysis of electroencephalograms of patients suffering from epilepsy.

Other applications of the conditional mutual information in neurophysiology are due to (Hinrichs et. al., 2006) and (Pflieger & Greenblatt, 2005).

Causality or coupling directions in multimode laser dynamics is another diverse field where the conditional mutual information was applied (Otsuka et al., 2004). (Paluš & A. Stefanovska, 2003) adapted the conditional mutual information approach (Paluš et al., 2001b) to analysis of instantaneous phases of interacting oscillators and demonstrated suitability of this approach for analyzing causality in cardio-respiratory interaction (Paluš et al., 2001b). The later approach has also been applied in neurophysiology (Brea et al., 2006).

More recent applications of information-theoretical functionals in natural sciences (medicine) are for example in (Van Dijck et al., 2007), inferring and quantifying causality in neuronal networks (Chicharro et al., 2011), (Vicente et al., 2010), in the computer simulation of human-robot interaction in (Sumioka et al., 1997) or in the relationship of predator-prey in etiology (Bochmann, 2007).

The information-theoretical functionals applied in social sciences are mostly in financial applications: i.e. application of transfer entropy to the information flow between various

financial time series (Dimpfl et al., 2011) or analysis of the Korean stock market by transfer entropy (Baek et al., 2006). Applications of transfer entropy in linguistics can be found in the book (Baeyer, 2005).

Social media (for example Twitter or Facebook) serves to researchers as an important source for studying social interactions. One important problem is the characterization and identification of influentials, which can be defined as users who influence the behavior of large number of other users. To characterize influence (in other words a causal relationship) in Twitter, researchers have suggested number of followers, mentions, and retweets (Cha et al., 2010), and Pagerank of follower network (Kwak et al., 2010). (Ver Steeg et al., 2011) however argue that the purely structural measures of influence (causality) can be misleading (Ghosh & Lerman, 2010) and high popularity does not necessarily imply high influence (Romero et al., 2010; Ver Steeg et al., 2011).

More recent work has used the size of the cascade trees (Bakshy et al., 2011) and influence-passivity score (Romero et al., 2010). One serious drawback of existing methods is that they are based on explicit causal knowledge (i.e., A responds to B), whereas for many data sets such knowledge is not available. (Ver Steeg et al., 2011) suggest a model-free transfer entropy approach to detect causal relationships and identifying influential users based on their capacity to predict the behavior of other users.

Having reviewed the relevant literature and also after extensive practical experience, we can state that the information-theoretic approach to the Granger causality plays an important, if not a dominant role in analyses of causal relationships in nonlinear systems.

In the following we define a practical criterium for detection of causal relationships among time series by means of transfer entropy.

4.3 Directionality index

To measure causal structure on small data sets and to allow conclusions about the dominant direction of the information flow, the (*causal*) *directionality index* was defined for transfer entropy or conditional mutual information by Paluš in (Paluš & A. Stefanovska, 2003) and analogically in (Rosenblum & Pikovsky, 2001). It is given by

$$DI(\mathbf{Y} \rightarrow \mathbf{X}|\mathbf{Z}) = \frac{\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}|\mathbf{Z}} - \mathcal{T}_{\mathbf{X} \rightarrow \mathbf{Y}|\mathbf{Z}}}{\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}|\mathbf{Z}} + \mathcal{T}_{\mathbf{X} \rightarrow \mathbf{Y}|\mathbf{Z}}}, \quad (8)$$

where X, Y, Z are time series. Paluš et al. in Paluš & A. Stefanovska (2003) consider special cases for Z , the so called phase increments of X and Y :

$$DI(\mathbf{Y} \rightarrow \mathbf{X}) = \frac{\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}} - \mathcal{T}_{\mathbf{X} \rightarrow \mathbf{Y}}}{\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}} + \mathcal{T}_{\mathbf{X} \rightarrow \mathbf{Y}}}, \quad (9)$$

where $\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}} = H(\mathbf{X}|\mathbf{X}^- \oplus \Delta\mathbf{X}^-) - H(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Y}^- \oplus \Delta\mathbf{X}^-)$ and $\Delta\mathbf{X} = \mathbf{X}(n+k) - \mathbf{X}(n)$ and similarly $\mathcal{T}_{\mathbf{X} \rightarrow \mathbf{Y}} = H(\mathbf{Y}|\mathbf{Y}^- \oplus \Delta\mathbf{Y}^-) - H(\mathbf{Y}|\mathbf{Y}^- \oplus \mathbf{X}^- \oplus \Delta\mathbf{Y}^-)$ and $\Delta\mathbf{Y} = \mathbf{Y}(n+k) - \mathbf{Y}(n)$.

The index varies between -1 and 1 , where negative values imply that the information flow from X to Y dominates and positive values indicate a large information flow from Y to X .

The definitions in the literature on the concrete subintervals are unfortunately not united. As well as there exist other modifications of the causal directionality index in the literature. We

will use the following definitions. The case when the index equals to -1 or 1 respectively we call *explicit unidirectional causality* from Y to X or X to Y respectively. The case when the directional index is in interval $< -1, 0$ we call it a *prevailing unidirectional causality* from X to Y (and analogically when the directional index is in interval $(0, 1 >$ we call it a *prevailing unidirectional causality* from Y to X). In case the directionality index equals to 0 , we call it an *absence of unidirectional causality*.

5. Gaussianity and causal relationships

To avoid any misunderstanding, we assume that the investigated random processes are given by time series with a finite number of data (a discrete case). We assume that one knows the probability (density) distribution of the processes which the time series represent, and these we define for a continuous case by explicit analytical formulas. By terms a probability density function or probability distributions which will be used in the text we mean a probability density distribution function.

(Barnett, 2009) recently proved that if all processes (time series) X, Y, Z defined by (4) are jointly Gaussian then Granger causality and transfer entropy are equivalent (up to a multiplication constant of 2). This result provides for the first time a unified framework for data-driven causal inference that bridges information-theoretic and autoregressive methods. For practice it means that in the complexity sense cheaper linear test can be applied for detection of causality, when one knows the time series are Gaussian.

We investigated in our paper (Hlaváčková-Schindler, 2011) the question, to which other multivariate probability distributions of the time series can be the equivalence (up to a multiplicative constant) of the two causality measures extended. In the same paper we extended the equivalence of these two measures to the generalized normal distribution, to the log-normal distribution and Weibull exponential distribution. Since a lot of phenomena in nature and social areas have these distributions, have our results practical implications for studying causal relationships among those phenomena.

In the following we will further investigate generalized Gaussian distributions given parametrically and causal relationships among them. Let us recall their definition.

5.1 Generalized Gaussian distributions

Generalized Gaussian distributions which was defined in 1995 by Nadarajah in (Nadarajah, 1995) is a parametrical class of distributions containing all Gaussian and Laplacian distributions as special cases.

Generalized Gaussian density (GGD) is defined as

$$p(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta} \quad (10)$$

where $\Gamma(\cdot)$ is the Gamma function $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$, $z > 0$. The parameter α , the scale parameter, models the width of the pdf peak (standard deviation) and β , the shape parameter, is inversely proportional to the decreasing rate of the peak.

The generalized Gaussian distribution for $\beta = 2$ is the Gaussian distribution and for $\beta = 1$ the Laplacian distribution.

5.1.1 Special subclasses of generalized Gaussians and their Kullback-Leibler divergence

Let us recall that the Kullback -Liebler distance (KLD) of two parametrical (general) probability density functions $p(x, \theta_p)$ and $q(x, \theta_q)$ is defined as

$$D(p(x, \theta_p) || q(x, \theta_q)) = \sum_x p(x, \theta_p) \log \frac{p(x, \theta_p)}{q(x, \theta_q)} \quad (11)$$

where θ_p and θ_q are parameters of probability distributions p and q respectively and $x \in R$.

It can be easily shown that for two probability density functions which are generalized Gaussians $p(x, \alpha_1, \beta_1), p(x, \alpha_2, \beta_2)$, the Kullback-Leibler divergence can be expressed as

$$D(p(x, \alpha_1, \beta_1) || p(x, \alpha_2, \beta_2)) = \log \left(\frac{\beta_1 \alpha_2 \Gamma(1/\beta_2)}{\beta_2 \alpha_1 \Gamma(1/\beta_1)} \right) + \left(\frac{\alpha_1}{\alpha_2} \right)^{\beta_2} \frac{\Gamma((\beta_2 + 1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}. \quad (12)$$

A subclass of generalized Gaussians where the shape parameter β is fixed is defined as

$$\mathcal{P}_\beta(\alpha) = \{p_\beta(x; \alpha), \alpha \in R^+ | p_\beta(x; \alpha) = \frac{\beta}{2\alpha \Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta}\}. \quad (13)$$

The Kullback-Leibler divergence between two pdfs from $\mathcal{P}_\beta(\alpha)$ is

$$D(p_\beta(x, \alpha_1) || p_\beta(x, \alpha_2)) = \log \left(\frac{\alpha_2}{\alpha_1} \right) + \left(\frac{\alpha_2}{\alpha_1} \right)^\beta \frac{1}{\beta} - \frac{1}{\beta}. \quad (14)$$

For $\beta = 1$ it is a Laplacian distribution. In this case is the Kullback -Liebler distance between two Laplacian distributions

$$D(p_1(x, \alpha_1) || p_1(x, \alpha_2)) = \log \left(\frac{\alpha_2}{\alpha_1} \right) + \left(\frac{\alpha_1}{\alpha_2} \right) - 1. \quad (15)$$

For two Gaussian functions with parameter α_1 and α_2 and $\beta_1 = \beta_2 = 2$ holds

$$D(p_2(x, \alpha_1) || p_2(x, \alpha_2)) = \log \left(\frac{\alpha_2}{\alpha_1} \right) + \frac{1}{2} \left(\frac{\alpha_1}{\alpha_2} \right) - \frac{1}{2}. \quad (16)$$

5.2 Transfer entropy as a conditional Kullback-Leibler divergence

Define $p(X) := \{p(x) | x \in X\}$ and similarly $p(X, Y) := \{p(x, y) | x \in X, y \in Y\}$ for joint probability distributions. Kullback-Leibler divergence D between two joint probability density functions $p(X, Y)$ and $q(X, Y)$ can be expressed as

$$D(p(X, Y) || q(X, Y)) = D(p(X) || q(X)) + D(p(Y|X) || q(Y|X)) \quad (17)$$

(the chain rule from Cover (Cover & Thomas, 1991).) Similarly, the conditional Kullback-Leibler divergence can be expressed as

$$D(p(X, Y|Z) || q(X, Y|Z)) = D(p(X|Z) || q(X|Z)) + D(p(Y|X, Z) || q(Y|X, Z)) \quad (18)$$

where

$$D(p(X|Z) || q(X|Z)) = \sum_{x \in X, z \in Z} p(x, z) \log \frac{p(x|z)}{q(x|z)}. \quad (19)$$

Transfer entropy can be in terms of Kullback-Leibler divergence rewritten as

$$\mathcal{T}_{x \rightarrow y|z} = D(p(x_{n+1}|x_n^{(k)}, y_n^{(l)}, z_n^{(k)}) || p(x_{n+1}|z_n^{(k)}), \quad (20)$$

where $x_n^{(k)} = (x_n^1, \dots, x_n^k)$, $z_n^{(k)} = (z_n^1, \dots, z_n^k)$, and $y_n^{(l)} = (y_n^1, \dots, y_n^l)$.

We will use the statements from Sections 5.1. and 5.2 in proofs of our results in Section 6.

6. Causality in time series with generalized Gaussian distributions: our results

The time series having generalized Gaussian distributions given parametrically allow to study and to express the relationships between them by means of their parameters. In the framework of definition of the directionality index introduced in section 4.3, we investigated the cases of the prevailing unidirectional causality, the explicit unidirectional causality and the absence of causality.

In the following we consider generalized Gaussian distributions given parametrically with the same shape parameter β . We give conditions on the relationships between the parameters of the involved probability distributions so that causal relationships between them can be detected by the directionality index.

Theorem 1. *Assume that times series X, Y have generalized Gaussian distributions given parametrically by $p_X(x, \alpha_x, \beta) = \frac{\beta}{2\alpha_x \Gamma(1/\beta)} e^{-(|x|/\alpha_x)^\beta}$ and $p_Y(y, \alpha_y, \beta) = \frac{\beta}{2\alpha_y \Gamma(1/\beta)} e^{-(|y|/\alpha_y)^\beta}$, $\alpha_x, \alpha_y, \beta > 0$. Then*

(i) *there is a prevailing unidirectional causality from Y to X , if for the parameters of both distributions*

$$p_X(x, \alpha_x, \beta) \text{ and } p_Y(y, \alpha_y, \beta) \text{ holds } \frac{\beta \log(\frac{\alpha_x}{\alpha_y})^2 + \Gamma(\frac{1}{\beta}) [(\frac{\alpha_y}{\alpha_x})^\beta - (\frac{\alpha_x}{\alpha_y})^\beta]}{\Gamma(\frac{1}{\beta}) [(\frac{\alpha_y}{\alpha_x})^\beta + (\frac{\alpha_x}{\alpha_y})^\beta] - 2} \in (0, 1 >.$$

(ii) *there is a prevailing unidirectional causality from X to Y , if for the parameters of both distributions*

$$p_X(x, \alpha_x, \beta) \text{ and } p_Y(y, \alpha_y, \beta) \text{ holds } \frac{\beta \log(\frac{\alpha_x}{\alpha_y})^2 + \Gamma(\frac{1}{\beta}) [(\frac{\alpha_y}{\alpha_x})^\beta - (\frac{\alpha_x}{\alpha_y})^\beta]}{\Gamma(\frac{1}{\beta}) [(\frac{\alpha_y}{\alpha_x})^\beta + (\frac{\alpha_x}{\alpha_y})^\beta] - 2} \in < -1, 0).$$

(iii) *there is an explicit unidirectional causality from Y to X , if for the parameters of both distributions*

$$p_X(x, \alpha_x, \beta) \text{ and } p_Y(y, \alpha_y, \beta) \text{ holds } \beta \log(\frac{\alpha_x}{\alpha_y})^2 = (\frac{\alpha_x}{\alpha_y})^\beta [1 + \Gamma(\frac{1}{\beta})] - 2.$$

(iv) *there is an explicit unidirectional causality from X to Y , if for the parameters of both distributions*

$$p_X(x, \alpha_x, \beta) \text{ and } p_Y(y, \alpha_y, \beta) \text{ holds } \beta \log(\frac{\alpha_x}{\alpha_y})^2 = 2 - 2\Gamma(\frac{1}{\beta}) (\frac{\alpha_y}{\alpha_x})^\beta.$$

(v) *there is an absence of unidirectional causality between time series X and Y , if for the parameters of both distributions $p_X(x, \alpha_x, \beta)$ and $p_Y(y, \alpha_y, \beta)$ holds $\beta \log(\frac{\alpha_x}{\alpha_y}) = -\frac{1}{2}\Gamma(\frac{1}{\beta}) [(\frac{\alpha_y}{\alpha_x})^\beta - (\frac{\alpha_x}{\alpha_y})^\beta]$ and $[(\frac{\alpha_y}{\alpha_x})^\beta + (\frac{\alpha_x}{\alpha_y})^\beta] \neq 2$.*

Proof: The items (i) -(v) can be proven by a direct application of the directionality index and the expression of transfer entropy by means of Kullback-Leibler divergence for generalized

$$\text{Gaussian distributions } D(X||Y) = \frac{2\beta\alpha_x\alpha_y \log \frac{\alpha_y}{\alpha_x} + \alpha_x^2 - \alpha_y^2}{(\alpha_x - \alpha_y)^2}.$$

6.1 Causality detection between two Laplacian distributions

For the Laplacian distributions holds $\beta = 1$, which simplifies the conditions set in Theorem 1 for the remaining parameters. Theorem 1 is then for the remaining parameters reformulated as follows.

Corollary 1. Assume that times series X, Y have Laplacian distributions given parametrically by $p_X(x, \alpha_x) = \frac{1}{2\alpha_x} e^{-(|x|/\alpha_x)}$ and $p_Y(y, \alpha_y) = \frac{1}{2\alpha_y} e^{-(|y|/\alpha_y)}$, $\alpha_x, \alpha_y > 0$. Then

- (i) there is a prevailing unidirectional causality from Y to X , if for the parameters of both distributions holds $\frac{\alpha_x}{\alpha_y} - 1 \geq \log\left(\frac{\alpha_x}{\alpha_y}\right) > \frac{\alpha_x^2 - \alpha_y^2}{2\alpha_x\alpha_y}$.
- (ii) there is a prevailing unidirectional causality from X to Y , if for the parameters of both distributions holds $\frac{2\log\left(\frac{\alpha_y}{\alpha_x}\right) + \frac{\alpha_y}{\alpha_x} - \frac{\alpha_x}{\alpha_y}}{\frac{\alpha_y}{\alpha_x} + \frac{\alpha_x}{\alpha_y} - 2} \in \langle -1, 0 \rangle$.
- (iii) there is an explicit unidirectional causality from Y to X , if for the parameters of both distributions holds $\log\left(\frac{\alpha_x}{\alpha_y}\right) = \frac{\alpha_x}{\alpha_y} - 1$.
- (iv) there is an explicit unidirectional causality from X to Y , if for the parameters of both distributions holds $\log\left(\frac{\alpha_y}{\alpha_x}\right) = \frac{\alpha_y}{\alpha_x} - 1$.
- (v) there is an absence of unidirectional causality between time series X and Y , if for the parameters of both distributions $p_X(x, \alpha_x)$ and $p_Y(y, \alpha_y)$ holds $\log\left(\frac{\alpha_x}{\alpha_y}\right) = -\frac{1}{2}\left[\frac{\alpha_y}{\alpha_x} - \frac{\alpha_x}{\alpha_y}\right]$ and $\frac{\alpha_y^2 + \alpha_x^2}{\alpha_x\alpha_y} \neq 2\alpha_x\alpha_y$.

Proof: The items (i) - (v) can be proven by a direct application of the directionality index and the expression of transfer entropy by means of Kullback-Leibler divergence for Laplacian distributions.

6.2 Causality detection between two Gaussian distributions

For the Gaussian distributions holds $\beta = 2$ and Theorem 1 for the remaining parameters simplifies into the following corollary.

Corollary 2. Assume that times series X, Y have Gaussian distributions given parametrically by $p_X(x, \alpha_x) = \frac{1}{\alpha_x\sqrt{\pi}} e^{-(x/\alpha_x)^2}$ and $p_Y(y, \alpha_y) = \frac{1}{\alpha_y\sqrt{\pi}} e^{-(y/\alpha_y)^2}$, $\alpha_x, \alpha_y > 0$. Then

- (i) there is a prevailing unidirectional causality from Y to X , if for the parameters of both distributions $p_X(x, \alpha_x)$ and $p_Y(y, \alpha_y)$ holds $\frac{4\log\left(\frac{\alpha_x}{\alpha_y}\right) + \sqrt{\pi}\left[\left(\frac{\alpha_y}{\alpha_x}\right)^2 - \left(\frac{\alpha_x}{\alpha_y}\right)^2\right]}{\sqrt{\pi}\left[\left(\frac{\alpha_y}{\alpha_x}\right)^2 + \left(\frac{\alpha_x}{\alpha_y}\right)^2\right] - 2} \in (0, 1 >$.
- (ii) there is a prevailing unidirectional causality from X to Y , if for the parameters of both distributions holds $\frac{4\log\left(\frac{\alpha_y}{\alpha_x}\right) + \sqrt{\pi}\left[\left(\frac{\alpha_x}{\alpha_y}\right)^2 - \left(\frac{\alpha_y}{\alpha_x}\right)^2\right]}{\sqrt{\pi}\left[\left(\frac{\alpha_x}{\alpha_y}\right)^2 + \left(\frac{\alpha_y}{\alpha_x}\right)^2\right] - 2} \in \langle -1, 0 \rangle$.
- (iii) there is an explicit unidirectional causality from Y to X , if for the parameters of both distributions holds $\log\left(\frac{\alpha_x}{\alpha_y}\right) = \frac{\sqrt{\pi}}{4}\left[\left(\frac{\alpha_x}{\alpha_y}\right)^2 + \left(\frac{\alpha_y}{\alpha_x}\right)^2\right] - \frac{1}{2}$.
- (iv) there is an explicit unidirectional causality from X to Y , if for the parameters of both distributions holds $\log\left(\frac{\alpha_y}{\alpha_x}\right) = 1 - \sqrt{\pi}\left(\frac{\alpha_y}{\alpha_x}\right)^2$.

(v) there is an absence of unidirectional causality between time series X and Y , if for the parameters of both distributions $p_X(x, \alpha_x)$ and $p_Y(y, \alpha_y)$ holds $\log \frac{\alpha_x}{\alpha_y} = -\frac{\sqrt{\pi}}{4} [(\frac{\alpha_y}{\alpha_x})^2 - (\frac{\alpha_x}{\alpha_y})^2]$ and $(\frac{\alpha_y}{\alpha_x})^2 + (\frac{\alpha_x}{\alpha_y})^2 \neq \frac{2}{\sqrt{\pi}}$.

Proof: The items (i) -(v) can be proven by a direct application of the directionality index and the expression of transfer entropy by means of Kullback-Leibler divergence for Gaussian distributions.

In praxis, for time series given by finite number of observations, one can apply methods for finding parameters of corresponding generalized Gaussian distributions, which these data sets interpolate. Knowing these parameters, the statements in our theorem and corollaries provide simple decision criteria about a presence or absence of causality between two time series.

7. Conclusion

In this paper we presented linear and non-linear methods on causality detection among time series, namely Granger causality and transfer entropy, respectively. We discussed their applications both in natural and social sciences. For the purpose of selecting a method for causality detection among time series, the approach of using statistical hypothesis testing techniques, applied in G-causality has several difficulties. The issue of multiple testing which is being done sequentially in general does not provide an optimal solution. In addition, there is no objective guideline for the choice of the input of the individual test (i.e. the size of the time series interval) and it is unclear how would such a choice influence the detection of causality. The main drawback of model dependent methods for causality detection is their instability. As expected, with a small or moderate number of observations, models close to each other are often hard to distinguish and the values of the model selection criterion are usually close to each other. A small change on the data may result in a choice of a different hypothesis. As a consequence, the causality detection based on the selected hypothesis may have a high variability (ill-posedness of the problem). The main conceptual advantage of the Granger causality model over the information-theoretical approaches to causality detection is its linearity and its straightforward generalization for multivariate time series testing. The main conceptual drawback of Granger causality is its inability to detect eventual non-linear causal relationships.

For the reasons stated above, in this paper we focused on causality detection by transfer entropy. We dealt with the Gaussian and some non-Gaussian distributions of natural phenomena as well of phenomena occurring in social sciences. We presented original and simple criteria for finite time series having generalized Gaussian distributions given parametrically representing natural or social phenomena which decide about a presence or absence of causality between them.

Our results can considerably simplify current more computationally demanding data processing methods applied in natural and social sciences for detection of causal relationships.

8. References

Abarbanel, H.D.I. (1993) Introduction to Nonlinear Dynamics for Physicists. *Lecture Notes in Physics*, World Scientific, Singapore.

- Ancona, N., Marinazzo, D. & Stramaglia, S. (2004) Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E* 70(5): 6221.
- Aparicio F.M. & Escribano, A. (1998) Information-theoretic analysis of serial dependence and cointegration. *Studies in Nonlinear Dynamics and Econometrics* 3: 119-140.
- Aristotle (350 B.C.) Physics. *The Internet Classics Archive*. Translated by R. P. Hardie and R. K. Gaye, <http://classics.mit.edu//Aristotle/physics.html>
- Arnhold, J., Grassberger, P., Lehnertz, K. & Elger, C.E. (1999) A robust method for detecting interdependences: application to intracranially recorded EEG. *Physica D* 134: 419-430.
- Baek, E.G. & Brock, W.A. (1992) A general test for nonlinear Granger causality: Bivariate model, Working paper, *Iowa State University and University of Wisconsin*, Madison.
- Baek, S.K., Jung, W.S., Kwon, O. & Hie-Tae Moon, H.T. (2006) Transfer entropy analysis of the Korean stock market. *Physica A: Statistical Mechanics and its Applications*, Elsevier Science.
- Baghli, M. (2006) A model-free characterization of causality *Economics Letters* 91: 380-388.
- Bakshy, E., Hofman, J.M., Mason, W.A. & Watts, D.J. (2011) Everyone's an influencer: quantifying influence on twitter. In *Proc. fourth ACM international conference on Web search and data mining, WSDN'11*, pages 65-74, New York, NY USA, ACM.
- Barnett, L., Barrett, A.B. & Seth, A.K. (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* 103, 238701.
- Bayer, H.C. (2005) Information: The New Language of Science. *Harvard University Press*, Harvard.
- Bell, D., Kay, J. & Malley, J. (1996) A non-parametric approach to non-linear causality testing. *Economical Letters* 51: 7-18.
- Bicego, M., Gonzalez-Jimenez, D., Grosso, E. & Alba Castro, J.L. (2008) Generalized Gaussian distributions for sequential data classification. *978-1-4244-2175-6/08/2008 IEEE*.
- Blinowska, K.J., Kuś, R., and Kamiński (2004) M. Granger causality and information flow in multivariate processes. *Phys.Rev. E* 70 050902(R).
- Boccaletti, S., Kurths, J., Osipov, G., Valladares, D.L., & Zhou, C.S. (2002) The synchronization of chaotic systems. *Physics Reports* 366: 1-101.
- Bochmann, O., Lizier, J.T., Mahoney, J., Obernosterer, G. & Pahle, J. (2007) Computational mechanics and information measures in food webs. *Workshop on Computational Biology*, University of Santa Fe.
- Brea, J., Russell, D.F. & Neiman A.B. (2006) Measuring direction in the coupling of biological oscillators: A case study for electroreceptors of paddlefish. *Chaos* 16: 026111.
- Broomhead D.S. & Lowe, D. (1988) Multivariate functional interpolation and adaptive networks *Complex Systems* 2: 321- 355.
- Cha, M., Haddadi, H., Benevenuto, F. & Gummadi, K.P. (2010) Measuring user influence in twitter: The milion follower fallacy. ICMSM'10: *In Proceedings of international AAAI Conference on Weblogs and Social*.
- Chávez, M., Martinerie, J. & Le Van Quyen, M. (2003) Statistical assessment of nonlinear causality: Application to epileptic EEG signals. *Journal of Neuroscience Methods* 124: 113-128.
- Chen, Y., Rangarajan, G., Feng J. & Ding, M. (2004) Analyzing multiple nonlinear time series with extended Granger causality. *Phys. Lett. A* 324: 26-35.
- Chicharro, D., Andrzejak, R.G. & Anders Ledberg (2011) Inferring and quantifying causality in neuronal networks. *BMC Neuroscience* 12(Suppl 1): 192.

- Cover T. & Thomas, J. (1991) Elements of Information Theory. *John Wiley and Sons, New York, NY*, Chapter 9.
- David O., Kiebel S., Harrison L., Mattout J., Kilner J., et al. (2006) Dynamic causal modelling of evoked responses in EEG and MEG. *NeuroImage* 30: 1255–1272.
- Diks, C. & DeGoede, J. (2001) A general nonparametric bootstrap test for Granger causality. *Global Analysis of Dynamical Systems*. Chapter 16, eds Broer, Krauskopf and Vegter, 391-403.
- Diks C. & Panchenko, V. (2005) A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics and Econometrics* 9 (4):1-7.
- Dimpfl, T., Huergo, L. & Peter, F.J. (2011) Using transfer entropy to measure information flows from and to the CDS market. *Proc. of the European Economic Association and Econometric Society*, 25 - 29 August 2011, Oslo, Norway, 2011.
- Feridun, M. (2007) Immigration, income and unemployment: An application of the bounds testing approach to cointegration. *Journal of Developing Areas* 41 (1): 37-49.
- Fujita, A., Severino, P., Sato, J.R. & Miyano, S. (2010) Granger causality in systems biology: Modeling gene networks in time series microarray data using vector autoregressive models. *Advances in Bioinformatics and Computational Biology*, LNCS 2010, Volume 6268/2010: 13-24.
- Ge T., Kendrick K.M., Feng J. (2009) A Novel Extended Granger Causal Model Approach Demonstrates Brain Hemispheric Differences during Face Recognition Learning. *PLoS Comput Biol* 5(11):e1000570.
- Geweke, J. (1984) Inference and causality in economic time series models. *Handbook of Econometrics*, eds Z. Griliches and M.D. Intriligator, North-Holland, 2: 1101–1144.
- Ghosh, R. & Lerman, K. (2010) Predicting influential users in online social networks. In *Proc. KDD workshop on Social Network Analysis (SNAKDD)*, May 2010.
- Gilbert, E. & Karahalios, K. (2009) Widespread worry and the stock market. *Association for the Advancement of Artificial Intelligence*.
- Granger, C.W.J. (2003) Time series analysis, cointegration, and applications. Nobel Lecture. In: *Les Prix Nobel. The Nobel Prizes 2003*, ed. Tore Frängsmyr, Nobel Foundation (Stockholm, 2004): 360–366.
- Granger, C.W.J. (1969) Investigating causal relations by econometric and cross spectral methods. *Econometrica* 37: 424-438.
- Grassberger, P. & Procaccia, I. (1983) Measuring of strangeness of strange attractors. *Physica D*: 189-208.
- Herdan, G. (1958) The relations between the dictionary distribution and the occurrence distribution of word length and its importance for the study of quantitative linguistics. *Biometrika* 45: 222-228.
- Hesse, W., Moller, E., Arnold, M. & Schack, B. (2003) The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *J Neurosci Methods* 124: 27-44.
- Hiemstra, C. & Jones, J.D. (1994) Testing for linear and nonlinear Granger causality in the stock price–volume relation. *Journal of Finance* 49: 1639-1664.
- Hinrichs, H., Heinze, H.J. & Schoenfeld, M.A. (2006) Causal visual interactions as revealed by an information theoretic measure and fMRI. *NeuroImage* 31 1051–1060.
- Hlaváčková-Schindler, K., Paluš, M. Vejmelka, M. & Bhattacharya, J. (2007) Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441: 1-46.

- Hlaváčková-Schindler, K. (2011) Equivalence of Granger causality and transfer entropy: A generalization. *Applied Mathematical Sciences* 5(73) 3637-3648.
- Hume, D. (1896) *A Treatise of Human Nature*. Selby-Bigge, ed., Clarendon Press.
- Illari, P.M., Russo, F. & Williamson, J. (2011) Causality in the Sciences. *Oxford University Press*.
- Kamiński, M., Ding, M., W.A. Truccolo, W.A. & Bressler, S.L. (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance *Biol. Cybern.* 85: 145–157.
- Kim, S., Putrino, D., Ghosh, S. & Brown, E.N. (2011) A Granger causality measure for point process models of ensemble neural spiking activity. *Computational Biology*7 (3) Open Access Publication.
- Kufmann, R.K. & Stern, D. (1997) Evidence for human influence on climate from hemispheric temperature relations. *Nature* 388, 3 July 1997, p.39.
- Kwak, H., Lee, C., Park, H. & Moon, S. (2010) What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web, WWW'10*, pages 591-600, New York, NY, USA, ACM.
- Limpert, E., Stahel, W.A. & Abbt, M. (2001) Log-normal distributions across the sciences: Keys and clues. *BioScience* 51 (5) May 2001.
- Mackie, J.L. (1988) *The Cement of the Universe: A study in Causation*. Clarendon Press, Oxford, England.
- Marinazzo, D., Pellicoro M. & Stramaglia, S. (2006) Nonlinear parametric model for Granger causality of time series. *Physical Review E* 73: 066216.
- Marschinski, R. & Kantz, H. (2002) Analysing the information flow between financial time series - An improved estimator for transfer entropy. *European Physical Journal B* 30: 275-281.
- Misiriotis, A., Kylafis, N.D., Papamastorakis, J. & Xilouris, E.M. (2000) Is the exponential distribution a good approximation of dusty galactic disks? *Astron. Astrophys.* 353: 117-123.
- Mokhov, I.I. & Smirnov, D.A. (2006) El Nino-Southern oscillation drives North Atlantic oscillation as revealed with nonlinear techniques from climatic indices. *Geophysical Research Letters* 33 L03708.
- Moulin, P. & Liu, J. (1999) Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Trans. Inform. Theory* vol. 45: 909-919.
- Nadarajah, S. (1995) A generalized normal distribution. *Journal of Applied Statistics* 32 (7): 685-694.
- Newton, I. , (1687) *Philosophiae Naturalis Principia Mathematica A new translation by I.B. Cohen and A. Whitman, University of California press, Berkeley* 1999.
- Otsuka, K., Miyasaka, Y., & Kubota, T. (2004) Formation of an information network in a self-pulsating multimode laser. *Phys. Rev. E* 69: 046201.
- Paluš, M. & A. Stefanovska (2003) Direction of coupling from phases of interacting oscillators: An information-theoretic approach. *Phys. Rev. E* 67: 055201(R).
- Paluš, M., Komárek, V., Procházka, T., Hrnčíř, Z. & Štěrbová, K. (2001) Synchronization and information flow in EEG of epileptic patients. *IEEE Engineering in Medicine and Biology Magazine* 20: 65-71.
- Paluš, M., Komárek, V., Hrnčíř, Z. & Štěrbová, K. (2001) Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E* 63: 046211.

- Park, I. & Principe, J.C. (2008) Correntropy based Granger causality. *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008*.
- Pearl, J. Causality (2000) Models, reasoning and inference. *Cambridge University Press*.
- Pflieger, M.E. & Greenblatt, R.E. (2005) Using conditional mutual information to approximate causality for multivariate physiological time series. *Int. J. Bioelectromagnetism* 7: 285-288.
- Pikovsky, A., Rosenblum M. & Kurths, J. (2001) Synchronization. A Universal Concept in Nonlinear Sciences. *Cambridge University Press, Cambridge*.
- Poon, S.H. & Granger, C.W.J. (2003) Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature* 41: 478-539.
- Preston, F.W. (1981) Pseudo-log-normal distributions. *Ecology* 62:355-364.
- Romero, D.M., Galuba, W., Asur, S. & Huberman, B.A. (2010) Influence and passivity in social media. *Social Science Research Network Working Paper Series*, August 2010.
- Quyen, M.L.V., Martinerie, J., Adam, C. & Varela, F.J. (1999) Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Physica D* 127: 250-266.
- Rosenblum, M.G. & Pikovsky, A.S. (2001) Detecting direction of coupling in interacting oscillators. *Phys Rev E* 64: 045202.
- Russo, F. (2009) Causality and causal modelling in the social sciences. Measuring variations. *Springer, New York*.
- Sailaja, V., Srinivasa Rao, K. & Reddy, K.V.V.S. (2010) Text independent speaker identification with finite multivariate generalized gaussian mixture model and hierarchical clustering algorithm. *Int. Journal of Computer Applications* 11 (11): 0975-8887.
- Schäfer, C., Rosenblum, M.G., Kurths, J. & Abel, H.H. (1998) Heartbeat synchronized with ventilation. *Nature* 392: 239-240.
- Schreiber, T. (2000) Measuring information transfer. *Phys. Rev. Lett.* 85: 461-464.
- Schäfer, C., Rosenblum, M.G., Kurths, J. & Abel, H.H. (1999) Synchronization in the human cardiorespiratory system. *Phys. Rev. E* 60: 857-870.
- Seth, A. K., Edelman, G. M. (2007) Distinguishing causal interactions in neural populations. *Neural Comput.* 19(4): 910-933.
- Sharifi, K., Leon-Garcia, A. (1995) Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Trans Circuits Syst. Video Technol.* 5: 52-56.
- Schiff, S.J., So, P., Chang, T., Burke, R.E. & Sauer, T. (1996) Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* 54: 6708-6724.
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D. & Woolrich, M.W. (2011) Network modelling methods for FMRI. *Neuroimage* 54 (2): 875-891.
- Statistical yearbook in Switzerland (1997).
- Sumioka, H., Asada, M. & Yoshikawa, Y. (2007) Causality detected by transfer entropy leads acquisition of joint attention. *IEEE 6. Intern. Conference on Development and Learning, London.* 264-269.
- Tass, P., Rosenblum, M.G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A. & Freund, H.J. (1998) Detection of n:m phase locking from noisy data: Application to Magnetoencephalography. *Phys. Rev. Lett.* 81: 3291-3294.

- Triacca, U. (2005) Is Granger causality analysis appropriate to investigate the relationships between atmospheric concentration of carbon dioxide and global surface air temperature? *Theor. Appl. Climatol.* 81: 133-135.
- Van Dijk, G., Van Vaerenbergh, J. & Van Hulle, M.M. (2007) Information theoretic derivations for causality detection: Application to human gait, LNCS, Vol. 4669: 159-168.
- Verdes, P.F. (2005) Assessing causality from multivariate time series. *Physical Review E* 72: 026222.
- Ver Steeg, G. & Galstyan, A. (2011) Information transfer in social media, October 2011, arxiv.org/PS_cache/arxiv/pdf/1110/1110.2724v1.pdf.
- Vicente R., Wibral M., Lindner M. & Pipa, G. (2010) Transfer entropy: a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience* doi:10.1007/s10827-010-0262-3: 1-23.
- Williams, C.B. (1940) A note on the statistical analysis of sentence length as a criterion of literary style. *Biometrika* 31:356-361.
- Zinde-Walsh, V. & Phillips, P. (2003) Fractional Brownian motion as a differential generalized Gaussian process. *Lecture Notes Monograph Series* Vol. 41.

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.