entropy

MDPI

Article

Kernel Methods for Nonlinear Connectivity Detection

Lucas Massaroppe [1,*] and Luiz A. Baccalá [2]

[1] Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Department of Atmospheric Sciences, University of São Paulo, São Paulo 05508-090, Brazil

[2] Escola Politécnica, Department of Telecommunications and Control Engineering, University of São Paulo, São Paulo 05508-900, Brazil; baccala@lcs.poli.usp.br

* Correspondence: lucasmassaroppe@usp.br

Received: 1 May 2019; Accepted: 15 June 2019; Published: 20 June 2019

Abstract: In this paper, we show that the presence of nonlinear coupling between time series may be detected using kernel feature space $\mathbb{F}$ representations while dispensing with the need to go back to solve the *pre-image problem* to gauge model adequacy. This is done by showing that the kernelized auto/cross sequences in $\mathbb{F}$ can be computed from the model rather than from prediction residuals in the original data space $\mathbb{X}$. Furthermore, this allows for reducing the connectivity inference problem to that of fitting a consistent linear model in $\mathbb{F}$ that works even in the case of nonlinear interactions in the $\mathbb{X}$-space which ordinary linear models may fail to capture. We further illustrate the fact that the resulting $\mathbb{F}$-space parameter asymptotics provide reliable means of space model diagnostics in this space, and provide straightforward Granger connectivity inference tools even for relatively short time series records as opposed to other kernel based methods available in the literature.

Keywords: nonlinear time series; nonlinear-Granger causality; inference

1. Introduction

Describing 'connectivity' has become of paramount interest in many areas of investigation that involve interacting systems. Physiology, climatology, and economics are three good examples where dynamical evolution modelling is often hindered as system manipulation may be difficult or unethical. Consequently, interaction inference is frequently constrained to using time observations alone.

A number of investigation approaches have been put forward [1–5]. However, the most popular and traditional one still is the nonparametric computation of cross-correlation (CC) between pairs of time series, and variants thereof, like coherence analysis [6], even despite their many shortcomings [7].

When it comes to connectivity analysis, recent times have seen the rise of Granger Causality (GC) as a unifying concept. This is mostly due to GC's unreciprocal character [8] (as opposed to CC) which allows for establishing the direction of information flow between component subsystems.

Most GC approaches rest on fitting parametric models to time series data and, again as opposed to CC, under appropriate conceptualization, also holds for more than just pairs of time series, giving rise to the ideas of (a) Granger connectivity and (b) Granger influentiability [9].

GC inferential methodology is dominated by the use of *linear* multivariate time series models [10]. This is so because linear models have statistical properties (and shortcomings) that are well understood besides having the advantage of sufficing when the data are Gaussian. As an added advantage GC characterization allows immediate frequency domain connectivity characterization via concepts like 'directed coherence' (DC) and 'partial directed coherence' (PDC) [11].

It is often the case, however that data Gaussianity does not hold. Whereas nonparametric approaches do exist [1,4,5], parametric nonlinear modelling offers little relief from the need for long observation data sets for reliable estimation in sharp contrast to linear models that perform well under the typical scenario of fairly short datasets over which natural phenomena can be considered stable. A case in point is neural data where animal behaviour changes are associated with relatively short-lived episodic signal modifications.

The motivation for the present development is that reproducing kernel transformations applied to data, as in the support vector machine learning classification case [12], can effectively produce estimates that inherit many of the good convergence properties of linear methods. Because these properties carry over under proper kernelization, it is possible to show that nonlinear links between subsystems can be rigorously detected.

Before proceeding, it is important to have in mind that the present developments focus on addressing the connectivity detection issue exclusively, in which we clearly show that solving the so-called *pre-image* reconstruction problem is unnecessary as has been until now assumed essential. This leads to a considerably simpler approach.

In Section 2, we formulate the problem and review some background about reproducing kernel theory together with the main results which are backed up by extensive numerical Monte Carlo illustrations in Section 3. Conclusions and current problem status and challenges end the paper (Section 4).

2. Problem Formulation

The most popular approach to investigating GC connectivity is through modeling multivariate time series via linear vector autoregressive models [10], where the central idea is to compare prediction effectiveness for a time series $x_i(n)$ when the past of other time series is taken into account in addition to its own past. Namely,

$$\mathbf{x}(n) = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{x}(n-k) + \mathbf{w}(n). \tag{1}$$

Under mild conditions, Equation (1) constitutes a valid representation of a linear stationary stochastic process where the evolution of $\mathbf{x}(n) = [x_1(n), \cdots, x_D(n)]^\top$ is obtained by filtering suitable $\mathbf{w}(n) = [w_1(n), \cdots, w_D(n)]^\top$ *purely stochastic innovation* processes, i.e., where $w_i(n)$ and $w_j(m)$ are independent provided $n \neq m$ [13]. If $w_i(n)$ are jointly Gaussian, so are $x_i(n)$ and the problem of characterizing connectivity reduces to well known procedures to estimate the $\mathbf{A}_k$ parameters in Equation (1) via least squares, which is the applicable maximum likelihood procedure. Nongaussian $w_i(n)$ translate into nongaussian $x_i(n)$ even if some actual (1) linear generation mechanism holds. Linearity among nongaussian $x_i(n)$ time series may be tested with help of cross-polyspectra [14,15], which, if unrejected, still allows for a representation like (1) whose optimal estimation requires a suitable likelihood function to accommodate the observed non-Gaussianity.

If linearity is rejected, $x_i(n)$ non-Gaussianity is a sign of nonlinear mechanisms of generation modelled by

$$\mathbf{x}(n) = \mathbf{g}(\mathbf{x}(n_-), \mathbf{w}(n)), \tag{2}$$

which generalizes (1) where $\mathbf{x}(n_-)$ stands for $\mathbf{x}(n)$'s past under some suitable dynamical law $\mathbf{g}(\cdot)$.

The distinction between (a) nonlinear $x_i(n)$ that are nonetheless linearly coupled as in (1) under nongaussian $\mathbf{w}(n)$ and (b) fully nonlinearly coupled processes is often overlooked. In the former case, linear methods suffice for connectivity detection [16] but fail in the latter case [17] calling for the adoption of alternative approaches. In some cases, however, linear approximations are inadequate in so far as to preclude connectivity detection [17].

In the present context, the solution to the connectivity problem entails a suitable data driven approximation of $\mathbf{g}(\cdot)$ whilst singling out the $x_i(n)$ and $x_j(n)$ of interest. To do so, we examine the employment of *kernel* methods [18] where functional characterization is carried out with the help of a high dimensional space representation

$$\boldsymbol{\phi} : \mathbb{X} \to \mathbb{F}, \tag{3}$$

for $F = \dim(\mathbb{F}) \gg D = \dim(\mathbb{X})$, where $\boldsymbol{\phi}(\mathbf{x}(n))$ is a mapping from the input space $\mathbb{X}$ into the feature space $\mathbb{F}$ whose role is to properly unwrap the data and yet ensure that the inner product $\langle \boldsymbol{\phi}(\mathbf{x}) | \boldsymbol{\phi}(\mathbf{y}) \rangle$ can be written as a simple function of $\mathbf{x}$ and $\mathbf{y}$ dispensing with the need for computations in $\mathbb{F}$. This possibility is granted by chosing $\boldsymbol{\phi}(\mathbf{x})$ to satisfy the so-called *Mercer condition* [19].

A simple example of (3) is the mapping

$$\phi : x \mapsto \langle \phi(x) | \;=\; [c, \sqrt{2c}\, x, x^2]^\top, \tag{4}$$

for $x \in \mathbb{R}$ and $\langle \phi(x) | \in \mathbb{F}$ using Dirac's bra-ket notation. In this case, the Mercer kernel is given by

$$\kappa(x, y) = \langle \boldsymbol{\phi}(x) | \boldsymbol{\phi}(y) \rangle = (c + xy)^2, \tag{5}$$

which is the simplest example of a polynomial kernel [18].

In the multivariate time series case, we consider

$$\boldsymbol{\phi} : \mathbf{x}(n) \mapsto [\langle \phi_1(x_1(n)) |, \cdots, \langle \phi_i(x_i(n)) |, \cdots, \langle \phi_D(x_D(n)) |]^\top, \tag{6}$$

where, for simplicity, we adopt the same transformation $\phi(\cdot) = \phi_i(\cdot) = \phi_j(\cdot)$ for each $x_i(n) \in \mathbb{R}$ time series component so that the

$$\langle \boldsymbol{\phi}(\mathbf{x}(n)) | \boldsymbol{\phi}(\mathbf{x}(m)) \rangle = \mathbf{K}(\mathbf{x}(n), \mathbf{x}(m)) \tag{7}$$

is a matrix whose elements are given by $K_{ij}(n, m) = \langle \phi(x_i(n)) | \phi(x_j(m)) \rangle$. In the development below, we follow the standard practice of denoting the $\mathbf{K}(\mathbf{x}(n), \mathbf{x}(m))$ quantities as $\mathbf{K}(m - n)$ in view of the assumed stationarity of the processes under study.

Rather than go straight into the statement of the general theory, a simple example is more enlightening. In this sense, consider a bivariate stationary time series

$$x_1(n) \;=\; g_1(x_1(n-1), w_1(n)), \tag{8}$$

$$x_2(n) \;=\; g_2(x_1(n-1), x_2(n-1), w_2(n)), \tag{9}$$

where $g_i(\cdot)$ are nonlinear functions and only the previous instant is relevant in producing the present behaviour. An additional feature, thru (9), is that $x_1(n)$ is connected to (Granger causes) $x_2(n)$ but not conversely. Application of the kernel transformation leads to

$$\langle \phi(x_1(n)) | \;=\; \langle \phi(g_1(x_1(n-1), w_1(n))) |, \tag{10}$$

$$\langle \phi(x_2(n)) | \;=\; \langle \phi(g_2(x_1(n-1), x_2(n-1), w_2(n))) |. \tag{11}$$

However, if one assumes the possibility of a linear approximation in $\mathbb{F}$, one may write

$$\begin{bmatrix} \langle \phi(x_1(n)) | \\ \langle \phi(x_2(n)) | \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \langle \phi(x_1(n-1)) | \\ \langle \phi(x_2(n-1)) | \end{bmatrix} + \begin{bmatrix} \langle \widetilde{w}_1(n) | \\ \langle \widetilde{w}_2(n) | \end{bmatrix}, \tag{12}$$

where $[\langle \widetilde{w}_1(n) | \; \langle \widetilde{w}_2(n) |]^\top$ stands for approximation errors in the form of innovations. Mercer kernel theory allows for taking the external product with respect to $[| \phi(x_1(n-1)) \rangle \; | \phi(x_2(n-1)) \rangle]^\top$ on both sides of (12) leading to

$$\mathbf{K}(\mathbf{x}(n), \mathbf{x}(n-1)) = \mathbf{A}\; \mathbf{K}(\mathbf{x}(n-1), \mathbf{x}(n-1)), \tag{13}$$

after taking expectations on both sides where

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \tag{14}$$

and

$$\mathbf{K}(\mathbf{x}(n),\mathbf{x}(m)) = \begin{bmatrix} \mathbb{E}[\langle\phi(x_1(n))|\phi(x_1(m))\rangle] & \mathbb{E}[\langle\phi(x_1(n))|\phi(x_2(m))\rangle] \\ \mathbb{E}[\langle\phi(x_2(n))|\phi(x_1(m))\rangle] & \mathbb{E}[\langle\phi(x_2(n))|\phi(x_2(m))\rangle] \end{bmatrix}, \tag{15}$$

since $\mathbb{E}[\langle\widetilde{w}_i(n)|\phi(x_j(m))\rangle] = 0$ for $n > m$ given that $\langle\widetilde{w}_i(n)|$ plays a zero mean innovations role.

It is easy to obtain $\mathbf{A}$ from sample kernel estimates. Furthermore, it is clear that (8) holds if and only if $\alpha_{12} = 0$.

To (13), which plays the role of Yule–Walker equations and which can be written more simply as

$$\mathbf{K}(-1) = \mathbf{A}\mathbf{K}(0), \tag{16}$$

and one may add the following equation to compute the innovations covariance

$$\mathbf{\Sigma}_{\langle\widetilde{\mathbf{w}}(n)|} = \mathbf{K}(\mathbf{x}(n),\mathbf{x}(n)) - \mathbf{A}\mathbf{K}(\mathbf{x}(n),\mathbf{x}(n))\mathbf{A}^\top, \tag{17}$$

where only reference to the $m-n$ difference is explicitly denoted assuming signal stationarity so that (17) simplifies to

$$\mathbf{\Sigma}_{\langle\widetilde{\mathbf{w}}(n)|} = \mathbf{K}(0) - \mathbf{A}\mathbf{K}(0)\mathbf{A}^\top = \mathbf{K}(0) - \mathbf{K}(-1)\mathbf{A}^\top. \tag{18}$$

This formulation is easy to generalize to model orders $p > 1$ and to more time series via

$$\langle\boldsymbol{\phi}(\mathbf{x}(n))| = \sum_{k=1}^{p} \mathbf{A}_k\langle\boldsymbol{\phi}(\mathbf{x}(n-k))| + \langle\widetilde{\mathbf{w}}(n)|, \tag{19}$$

where

$$\langle\boldsymbol{\phi}(\mathbf{x}(n))| = [\langle\phi(x_1(n))|,\cdots,\langle\phi(x_D(n))|]^\top, \tag{20}$$

which is assumed as due to filtering appropriately modelled innovations $\langle\widetilde{\mathbf{w}}(n)|$. For the present formulation, one must also consider the associated 'ket'-vector

$$|\boldsymbol{\phi}(\mathbf{x}(m))\rangle = [|\phi(x_1(m))\rangle,\cdots,|\phi(x_D(m))\rangle]^\top, \tag{21}$$

that when applied to (19) for $n > m$ after taking expectations $\mathbb{E}[\cdot]$ under the zero mean innovations nature of $\langle\mathbf{w}^\phi(n)|$ leads to

$$\mathbf{K}_{\mathbf{x}}^{\phi}(l) = \sum_{k=1}^{p} \mathbf{A}_k\mathbf{K}_{\mathbf{x}}^{\phi}(l+k), \tag{22}$$

where $l = m-n$ and $\mathbf{K}_{\mathbf{x}}^{\phi}(m)$'s elements are given by $\mathbb{E}[\langle\phi(x_i(l-m))|\phi(x_j(l))\rangle]$ so that (22) constitutes a generalization of the Yule–Walker equations. By making $l = m-n = -1,\cdots,-p$ one may reframe (22) in matrix form as

$$\bar{\boldsymbol{\kappa}}_p = \begin{bmatrix} \mathbf{K}_{\mathbf{x}}^{\phi}(-1) \\ \vdots \\ \mathbf{K}_{\mathbf{x}}^{\phi}(-p) \end{bmatrix} = \begin{bmatrix}\mathbf{A}_1 \cdots \mathbf{A}_p\end{bmatrix} \begin{bmatrix} \mathbf{K}_{\mathbf{x}}^{\phi}(0) & \mathbf{K}_{\mathbf{x}}^{\phi}(-1) & \cdots & \mathbf{K}_{\mathbf{x}}^{\phi}(-p+1) \\ \mathbf{K}_{\mathbf{x}}^{\phi}(1) & \mathbf{K}_{\mathbf{x}}^{\phi}(0) & \cdots & \mathbf{K}_{\mathbf{x}}^{\phi}(-p+2) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{K}_{\mathbf{x}}^{\phi}(p-1) & \mathbf{K}_{\mathbf{x}}^{\phi}(p-2) & \cdots & \mathbf{K}_{\mathbf{x}}^{\phi}(0) \end{bmatrix} = \mathcal{A}\mathcal{K}_p(0), \tag{23}$$

where $\mathcal{K}_p(0)$ is block Toeplitz matrix containing p Toeplitz blocks. Equation (23) provides pD^2 equations for the same number of unknown parameters in $\mathcal{A}$.

The high model order counterpart to (17) is given by

$$\mathbf{\Sigma}_{\langle\widetilde{\mathbf{w}}(n)|} = \mathbf{K}_{\mathbf{x}}^{\phi}(0) - \sum_{k=1}^{p}\sum_{l=1}^{p} \mathbf{A}_k\mathbf{K}_{\mathbf{x}}^{\phi}(k-l)\mathbf{A}_l^\top = \mathbf{K}_{\mathbf{x}}^{\phi}(0) - \mathcal{A}\mathcal{K}_p(0)\mathcal{A}^\top. \tag{24}$$

It is not difficult to see that the more usual Yule–Walker complete equation form becomes

$$[\mathbf{I}\ -\mathcal{A}]\ \mathcal{K}_{p+1}(0) = \begin{bmatrix} \boldsymbol{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|} \\ \mathbf{0} \end{bmatrix}. \tag{25}$$

There are a variety of ways for solving for the parameters. A simple one is to define $\mathbf{a} = \text{vec}(\mathcal{A})$ leading to

$$\text{vec}(\bar{\boldsymbol{\kappa}}_p) = (\mathcal{K}_p^\top(0) \otimes \mathbf{I})\ \mathbf{a}. \tag{26}$$

Even though one may employ least-squares solution methods to solve either (26) or (23), a Total-Least-Squares (TLS) approach [20] has proven a better solution since both members of the equations are affected by estimation inaccuracies that are better dealt with using TLS.

Likewise, (24) can be used in conjunction with generalizations of model order criteria of Akaike's AIC type

$$_{\text{g}}\text{AIC}(k) = \ln(\det(\boldsymbol{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|})) + \frac{c_{n_s}}{n_s} kD^2, \tag{27}$$

where n_s stands for the number of available time observations. In generalizing Akaike's criterion to the multivariate case $c_{n_s} = 2$, whereas $c_{n_s} = \ln(\ln(n_s))$ for the Hannan–Quinn criterion, our choice in this paper.

Thus far, we have described procedures for choosing model order in the $\mathbb{F}$ space. In ordinary time series analysis, in addition to model order identification, one must also perform proper model diagnostics. This entails checking for residual whiteness among other things. This is usually done by checking the residual auto/crosscorrelation functions for their conformity to a white noise hypothesis.

In the present formulation because, we do not explicitly compute the $\mathbb{F}$ space series, we must resort to means other than computing the latter correlation functions from the residual data as usual. However, using the same ideas for computing (24), one may obtain estimates of the innovation cross-correlation in the feature space at various lags as

$$\boldsymbol{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|\tilde{\mathbf{w}}(m)\rangle} = \boldsymbol{\Sigma}_{\tilde{\mathbf{w}}}(m-n) = \mathbf{K}_{\mathbf{x}}^{\phi}(m-n) - \sum_{k=1}^{p}\sum_{l=1}^{p} \mathbf{A}_k \mathbf{K}_{\mathbf{x}}^{\phi}(m-n+k-l)\mathbf{A}_l^\top, \tag{28}$$

by replacing $\mathbf{K}_{\mathbf{x}}^{\phi}(m-n+k-l)$ by their estimates and using $\mathbf{A}_k$ obtained by solving (22) for $m-n$ between a minimum $-L$ to a $+L$ maximum lag. The usefulness of (28) is to provide means to test model accuracy and quality as a function of ϕ choice under the best model order provided by the model order criterion.

By defining a suitable normalized estimated lagged *kernel correlation function (KCF)*

$$\text{KCF}_{ij}(\tau) = \frac{K_{ij}(\tau)}{\sqrt{K_{ii}(0)K_{jj}(0)}}, \tag{29}$$

which, given the inner product nature of kernel definition, satisfies the condition

$$|\text{KCF}_{ij}(\tau)| \leq 1, \tag{30}$$

as easily proved using the Cauchy–Schwarz inequality.

The notion of $\text{KFC}(\tau)$ applies not only to the original kernels but also in connection with the residual kernel values given by (28) where, for explicitness, we write it as

$$\text{KCF}_{ij}^{(r)}(\tau) = \frac{\Sigma_{ij}(\tau)}{\sqrt{\Sigma_{ii}(0)\Sigma_{jj}(0)}}, \tag{31}$$

where $\Sigma_{ij}(\tau)$ are the matrix entries in (28).

In the numerical illustrations that follow, we have assumed that $\mathrm{KCF}_{ij}^{(r)}(\tau) \sim \mathcal{N}(0, 1/n_s)$ asymptotically under the white residual hypothesis

$$\mathcal{H}_0 : \mathrm{KCF}_{ij}^{(r)}(\tau) = 0. \tag{32}$$

This choice turned out to be reasonably consistent in practice. Along the same line of reasoning, other familiar tests over residuals, such as the Portmanteau test [10] were also carried out and consistently allowed verifying residual nonwhiteness.

One may say that the present theory follows closely the developments of ordinary second order moment theory with the added advantage that now nonlinear connections can be effectively captured by replacing second order moments by their respective lagged kernel estimates.

2.1. Estimation and Asymptotic Considerations

The essential problem then becomes that of estimating the entries of $\mathbf{K}_{\mathbf{x}}^{\phi}(n, m)$, entries. They can be obtained by averaging kernel values computed over the available data

$$K_{ij}(n, m) = \frac{1}{n_s} \sum_s \langle \phi(x_i(n-s)) | \phi(x_j(m-s)) \rangle, \tag{33}$$

for nonzero terms in the $s \in [1, n_s]$ range.

Under these conditions, for an appropriately defined kernel function, the feature space becomes linearized and, following [21], it is fair to assume that the estimated vector stacked representation of the model coefficient matrices

$$\mathbf{a} = \mathrm{vec}([\mathbf{A}_1 \cdots \mathbf{A}_p]) \tag{34}$$

is asymptotically Gaussian, i.e.,

$$\sqrt{n_s}(\hat{\mathbf{a}} - \mathbf{a}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}^{-1} \otimes \mathbf{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|}), \tag{35}$$

where $\mathbf{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|}$ is the feature space residual matrix given by (28) and where

$$\mathbf{\Gamma} = \mathbb{E}[\mathbf{y}_L \mathbf{y}_R^\top] \tag{36}$$

for the 'bra'-vector

$$\mathbf{y}_L^\top = [\langle \boldsymbol{\phi}(\mathbf{x}(n))|, \cdots, \langle \boldsymbol{\phi}(\mathbf{x}(n-p+1))|]^\top \tag{37}$$

and the 'ket'-vector

$$\mathbf{y}_R^\top = [|\boldsymbol{\phi}(\mathbf{x}(n))\rangle, \cdots, |\boldsymbol{\phi}(\mathbf{x}(n-p+1))\rangle]^\top, \tag{38}$$

which are used to construct the kernel scalar products. It is immediate to note that (36) is a Toeplitz matrix composed of suitably displaced $\mathbf{K}_{\mathbf{x}}^{\phi}(\cdot)$ blocks.

An immediate consequence of (35) is that one may test for model coefficient nullity and thereby provide a kernel Granger Causality test. This is equivalent to testing for $a_{ij}(k) = 0$ so that the statistic

$$_g\lambda_W = \hat{\mathbf{a}}^\top \mathbf{C}^\top \left[\mathbf{C} \left(\mathbf{\Gamma}^{-1} \otimes \mathbf{\Sigma}_{\langle \tilde{\mathbf{w}}(n)|} \right) \mathbf{C}^\top \right]^{-1} \mathbf{C} \hat{\mathbf{a}}, \tag{39}$$

where $\mathbf{C}$ is a contrast matrix (or structure selection matrix) so that the null hypothesis becomes

$$\mathcal{H}_0 : \mathbf{C}\mathbf{a} = \mathbf{0}. \tag{40}$$

Hence, under (35),

$$_g\lambda_W \xrightarrow{d} \chi^2_\nu, \tag{41}$$

where $\nu = \mathrm{rank}(\mathbf{C})$ corresponds to the number of the explicitly imposed constraints on $a_{ij}(k)$.

Data Workflow

Given $x_i(n)$, analysis proceeds by

1. Computing the kernel values (33) to obtain the kernel Yule–Walker equations (25) or equivalently (26) for a given value of p (starting from $p = 1$);
2. After solving the latter for the parameters in $\mathbf{a}$ via Total-Least-Squares (TLS), one computes (18) wherefrom the generalized model order choice criterion (27) can be computed;
3. With the help of the computed (33) values, one can obtain the residual $\text{KCF}_{ij}^{(r)}(\tau)$ functions in (31) which can be used to check model adequacy via (32). Additionally, Portmanteau tests may be also used;
4. If $\text{KCF}_{ij}^{(r)}(\tau)$ analysis does not suggest feature space model residual whiteness, p is increased by 1, and the procedure from step 1 is repeated until feature space model residual whiteness is obtained and $_g\text{AIC}(k)$ attains its first local minimum meaning that the ideal model order has been reached;
5. Once the best model is attained, one employs the (39) to infer connectivity.

These steps closely mirror those of ordinary time series model fitting and analysis.

3. Numerical Illustrations

The following examples consist of nonlinearly coupled systems that are simulated with the help of zero mean unit variance normal uncorrelated innovations $w_i(n)$. All simulations (10,000 realizations each) were preceded by an initial a burn-in period of 10,000 data points to avoid transient phenomena. Estimation results are examined as a function of $n_s = \{32, 64, 128, 256, 512, 1024, 2048\}$ with $\alpha = 1\%$ significance.

For brevity, Example 1 is carried out in full detail, whereas approach performance for the other ones is gauged mostly through the computation of observed detection rates except for Examples 4 and 5 which also portray model order choice criterion behaviour.

Simulation results are displayed in terms of how true and false detection rates depend on realization length n_s.

3.1. Example 1

Consider the simplest possible system whose connectivity cannot be captured by linear methods [17] as there is a unidirectional quadratic coupling from $x_2(n)$ to $x_1(n)$

$$\begin{cases} x_1(n) = a x_1(n-1) + c x_2^2(n-1) + w_1(n), \\ x_2(n) = b x_2(n-1) + w_2(n), \end{cases} \tag{42}$$

with $a = 0.2$, $b = 0.6$ and $c = 0.7$.

An interesting aspect of this simple system is the possibility of easily relating its coefficients a, b and c to those in (14) that describe its $\mathbb{F}$ space evolution. This may be carried out explicitly after substituting (42) into the computed kernels of Equation (13). After a little algebra, this leads to

$$\left[\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} - \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \right] = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_{11} & \theta_{12} \\ 0 & 0 \end{bmatrix}, \tag{43}$$

where θ_{11} and θ_{12} depend on the computed kernel values. From (43), it immediately follows for example that $b = \alpha_{22}$ and more importantly that $\alpha_{21} = 0$ as expected. Vindication of the observation of these theoretically determined values also gives the means for testing estimation accuracy.

For illustrations' sake, we write the kernel Yule–Walker Equations (22) with their respective solutions ($n_s = 512$) for one given (typical) realization

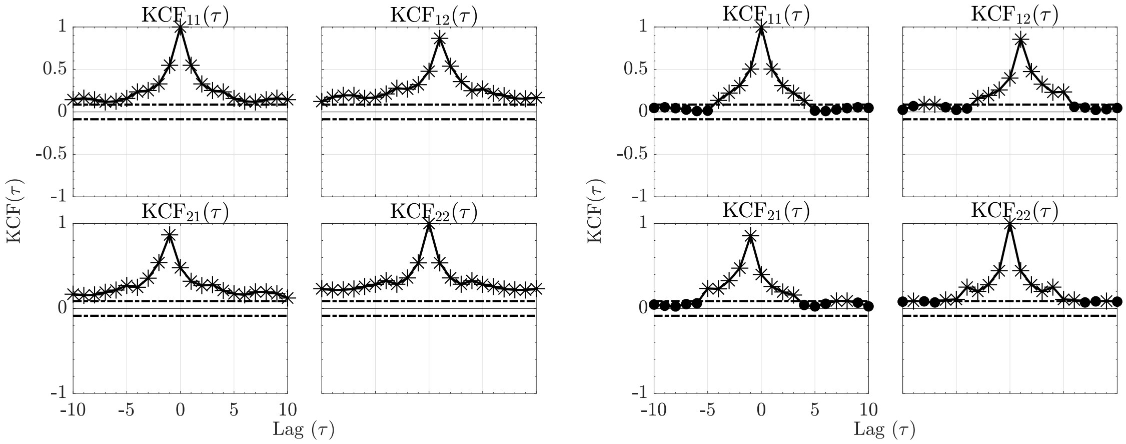
$$\begin{bmatrix} 210.7583 & 23.5416 \\ 23.5416 & 8.6450 \end{bmatrix} \mathbf{A}^{(2)} = \begin{bmatrix} 125.7501 & 37.7803 \\ 17.7389 & 5.3788 \end{bmatrix} \rightarrow \mathbf{A}^{(2)} = \begin{bmatrix} 0.1559 & 3.9456 \\ 0.0211 & 0.5648 \end{bmatrix}, \tag{44}$$

for the quadratic kernels ($\kappa(x,y) = (xy)^2$) and

$$10^5 \times \begin{bmatrix} 8.0302 & 0.0868 \\ 0.0868 & 0.0052 \end{bmatrix} \mathbf{A}^{(4)} = 10^5 \times \begin{bmatrix} 4.1597 & 0.1755 \\ 0.0594 & 0.0025 \end{bmatrix} \rightarrow \mathbf{A}^{(4)} = \begin{bmatrix} 0.1843 & 30.8922 \\ 0.0027 & 0.4386 \end{bmatrix}, \tag{45}$$

for the quartic kernels ($\kappa(x,y) = (xy)^4$). Superscripts point to kernel order. One may readily notice approximate compliance to the expected α_{ij} coefficients.

Further appreciation of this example may be obtained via a plot of the normalized estimated KCF(τ) (29) shown in Figure 1.



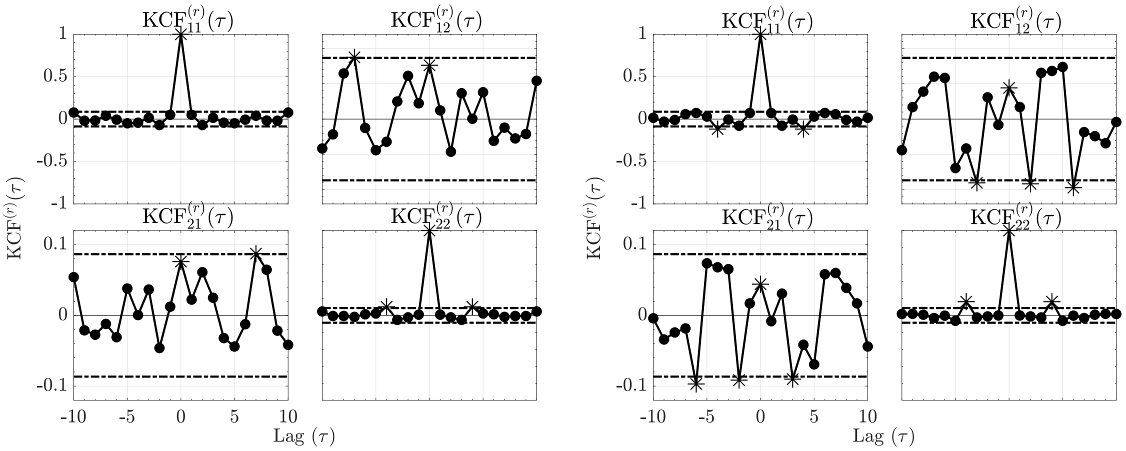
(**a**) $\kappa(x,y) = (xy)^2$ (**b**) $\kappa(x,y) = (xy)^4$.

Figure 1. The sequence kernel correlation functions (KCF(τ)) respectively for the quadratic and quartic kernels are contained in Figure 1a,b for Example 1. Horizontal dashed lines represent 95% significance threshold interval out of which the null hypothesis $\mathcal{H}_0$ of no correlation is rejected. Asterisks ($*$) further stress signficant values.

The residual normalized kernel sequences (31) computed using (28) are depicted in Figure 2 for each kernel and show effective decrease below the null hypothesis decision threshold line vindicating adequate modelling.

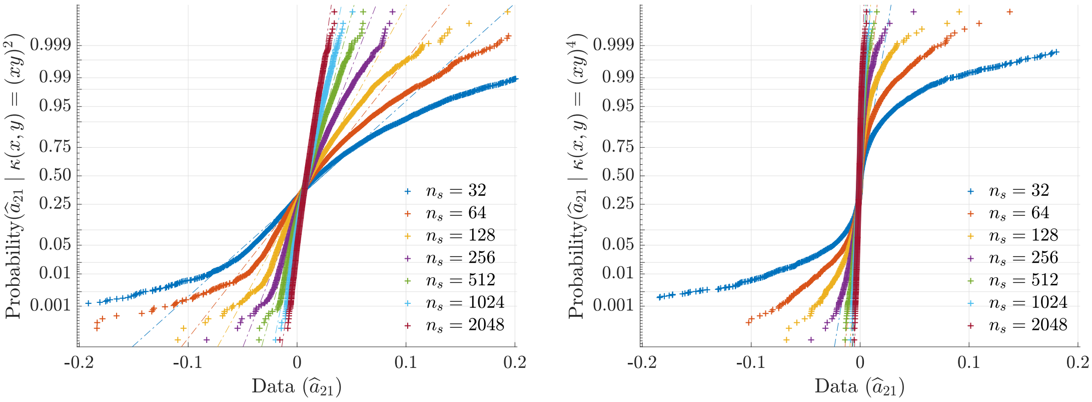
Moreover, for this realization, one may show that the Hannan–Quinn Information Criterion (27) points to the correct order of $p = 1$. In addition, Portmanteau tests do not reject whiteness in the $\mathbb{F}$ space for either kernel further confirming successful modelling in both cases.

To illustrate and confirm the Gaussian asymptotic behaviour discussed in Section 2.1, normal probability plots for $\widehat{a}_{21}$ are presented in Figure 3. Further objective quantification of the convergence speed towards normality is provided by the evolution towards 1 of the *Filliben* squared-correlation coefficient [22–24] as a function of n_s (Figure 4).



(**a**) $\kappa(x, y) = (xy)^2$ (**b**) $\kappa(x, y) = (xy)^4$.

Figure 2. The residue kernel correlation functions ($\mathrm{KCF}^{(r)}(\tau)$) respectively for the quadratic and quartic kernels are shown in Figure 2a,b for Example 1. Comparing them to Figure 1, it is clear that the kernel correlations are reduced after modelling as it is now impossible to reject $\mathrm{KCF}^{(r)}(\tau)$ nullity at 95% as no more than 5% of the values lie outside the dashed interval around zero. Asterisks ($*$) further stress significant values.



(**a**) $\kappa(x, y) = (xy)^2$ (**b**) $\kappa(x, y) = (xy)^4$.

Figure 3. Ensemble normal probability plots for $\widehat{a}_{21}$, respectively for 3a quadratic and 3b quartic kernels, illustrate and confirm asymptotic normality.

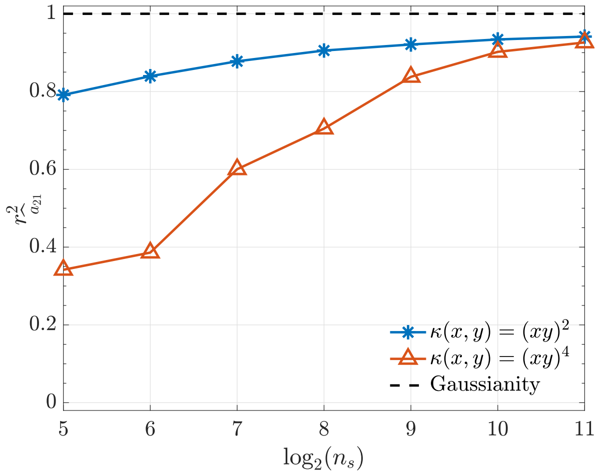


Figure 4. *Filliben* squared-correlation coefficient convergence to Gaussianity as a function of n_s for both kernels used in Example 1.

Convergence to normality justifies using (39) to test for null connectivity hypotheses. Test perfomance is depicted in Figure 5.

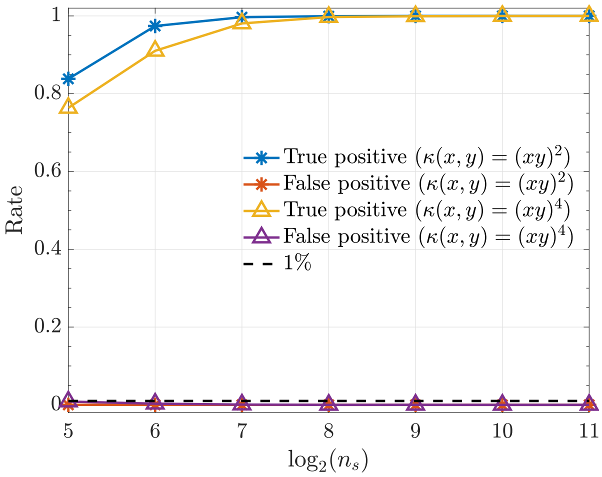


Figure 5. True positive and false positive rates from the kernelized Granger causality test for various samples sizes (n_s) for $\alpha = 1\%$. Note that the false-positive-rates for both kernels overlap.

3.2. Example 2

Consider $x_1(n)$, a highly resonant ($R = 0.99$) linear oscillator (at a normalized frequency of $f = 0.1$) to be unidirectionally coupled to a low pass system $x_2(n)$ through a delayed squared term

$$\begin{cases} x_1(n) = 2R\cos(2\pi f)x_1(n-1) - R^2 x_1(n-2) + w_1(n), \\ x_2(n) = -0.9x_2(n-1) + cx_1^2(n-1) + w_2(n), \end{cases} \tag{46}$$

where $c = 0.1$ [17].

This system was already investigated elsewhere [17,25,26] under a different estimation algorithm and with fewer Monte Carlo replications. The null hypothesis connectivity results are presented in Figure 6 showing adequate asymptotic decision success. A quadratic kernel was used in all cases.

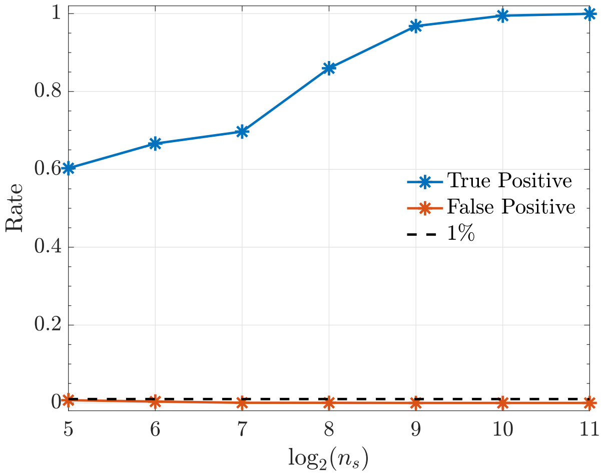


Figure 6. True positive ($x_1 \to x_2$) and false positive rates ($x_2 \to x_1$) from the kernelized Granger causality test under a quadratic kernel as a function n_s in Example 2.

3.3. Example 3

The present example comes from a model in [27],

$$\begin{cases} x_1(n) = 3.4x_1(n-1)[1-x_1^2(n-1)]\mathrm{e}^{-x_1^2(n-1)} + w_1(n), \\ x_2(n) = 3.4x_2(n-1)[1-x_2^2(n-1)]\mathrm{e}^{-x_2^2(n-1)} + c_1x_1^2(n-1) + w_2(n), \\ x_3(n) = 3.4x_3(n-1)[1-x_3^2(n-1)]\mathrm{e}^{-x_3^2(n-1)} + c_2x_2^4(n-1) + w_3(n). \end{cases} \tag{47}$$

This choice was dictated by the nonlinear wideband character of its signals. The values $c_1 = 0.7$ and $c_2 = 0.9$ were adopted.

Figure 7 shows that connection detectability improves as signal duration n_s increases except for the nonexisting $x_3(n) \leftarrow x_1(n)$ connection whose performance stays more or less constant with a false positive rate slightly above $\alpha = 1\%$. All computations used quadratic kernels.

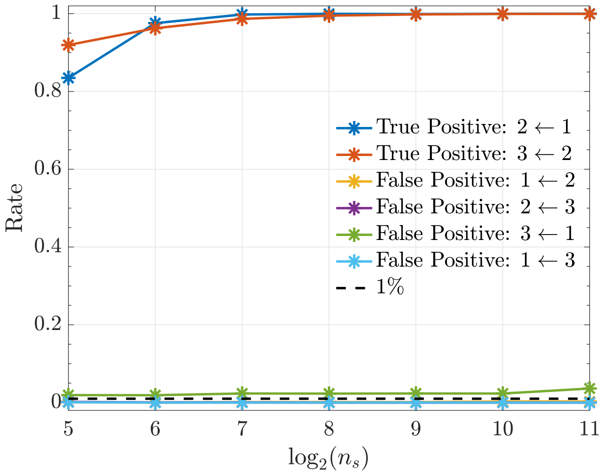


Figure 7. True positive and false positive rates (Example 3) from the kernelized Granger causality test using a quadratic kernel as a function of n_s. Note that the false-positive-rate for the connections 1 ← 2, 2 ← 3 and 1 ← 3 overlap over the investigated n_s range.

3.4. Example 4

For this numerical illustration, consider the model presented in [28]

$$\begin{cases} x_1(n) = 3.4x_1(n-1)[1-x_1^2(n-1)]\mathrm{e}^{-x_1^2(n-1)} + 0.8x_1(n-2) + w_1(n), \\ x_2(n) = 3.4x_2(n-1)[1-x_2^2(n-1)]\mathrm{e}^{-x_2^2(n-1)} + 0.5x_2(n-2) + cx_1^2(n-2) + w_2(n). \end{cases} \tag{48}$$

System (48) produces nonlinear wideband signals with a quadratic (1 → 2) coupling factor whose intensity is given by c taken here as 0.5.

It is worth noting that, kernelized Granger causality true positive rate improves as sample size (n_s) increases (Figure 8) and using the generalized Hannan–Quinn criterion, the order of kernelized autoregressive vector models identified for a typical realization was correctly identified and equals 2 as expected (see Figure 9).

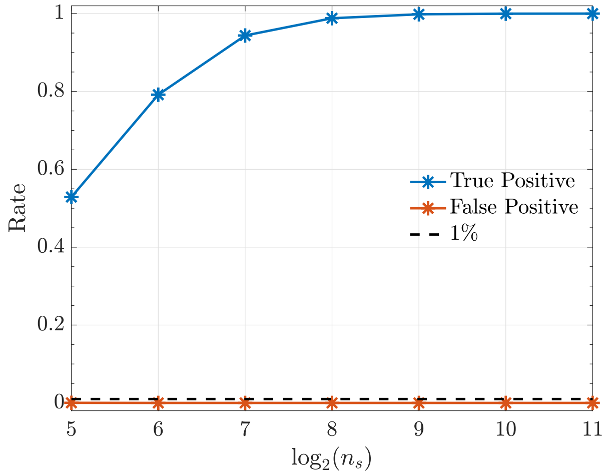


Figure 8. True positive and false positive rates from the kernelized Granger causality test under a quadratic kernel as function of record length n_s in Example 4.

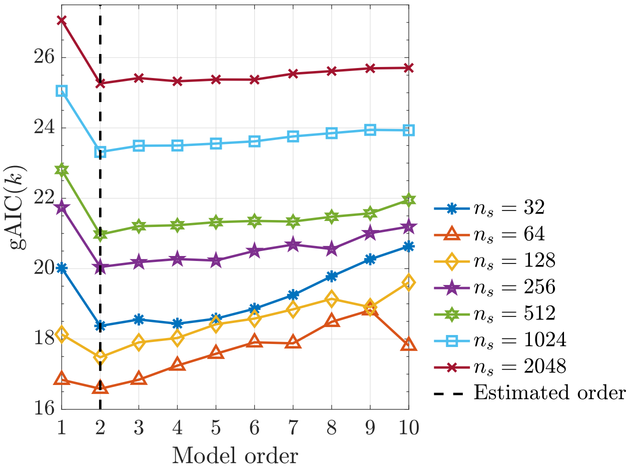


Figure 9. Generalized Hannan–Quinn criterion (gAIC(k)) with $c_{n_s} = \ln(\ln(n_s))$ as a function of model order for various observed record lengths n_s using a typical realization from (48).

3.5. Example 5

As a last numerical illustration, consider data generated by

$$\begin{cases} x_1(n) = 3.4x_1(n-3)[1 - x_1^2(n-3)]\mathrm{e}^{-x_1^2(n-3)} + 0.4x_1(n-4) + w_1(n), \\ x_2(n) = 3.4x_2(n-1)[1 - x_2^2(n-1)]\mathrm{e}^{-x_2^2(n-1)} + c_1 x_1^2(n-2) + w_2(n), \\ x_3(n) = 3.4x_3(n-2)[1 - x_3^2(n-2)]\mathrm{e}^{-x_3^2(n-2)} + c_2 x_2^2(n-3) + w_3(n), \end{cases} \tag{49}$$

with $c_1 = 0.9$ and $c_2 = 0.4$.

Under the quadratic kernel and employing kernelized Hannan–Quinn information criterion (27) (see Figure 10), one can see that the estimated model order is $p = 3$ as expected judging from the $x_2^2(n-3)$ term in (49). In addition, kernelized Granger causality detectability improves with record length n_s increase (Figure 11).

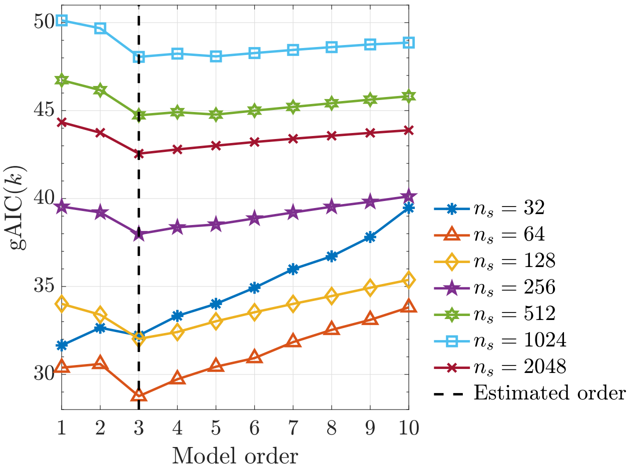


Figure 10. Generalized Hannan–Quinn criterion ($_g$AIC(k)) as a function of model order for the various data lengths n_s from a typical realization from (49).

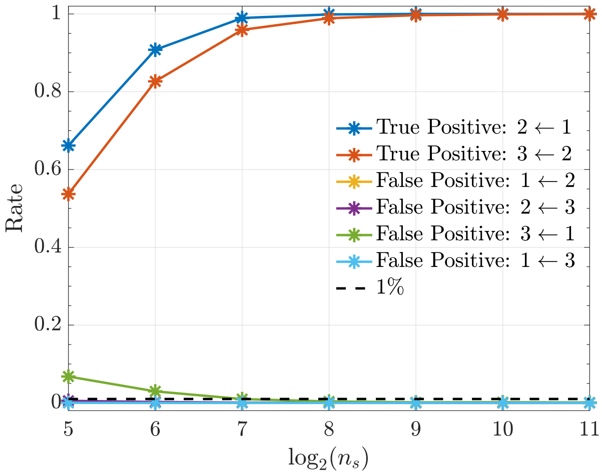


Figure 11. Observed true positive and false positive rates from the kernelized Granger causality test under a quadratic kernel for various record lengths n_s in Example 5. Note that the false-positive-rate for the connections 2 ← 3, 3 ← 1 and 1 ← 3 overlap over the n_s range, except for 1 ← 2, which, however, attains the same level as the others after $n_s = 128$.

4. Conclusions and Future Work

After a brief theoretical presentation (Section 2), we have shown that canonical model fitting procedures that involve (a) model specification with order determination and (b) explicit model diagnostic testing can be successfully carried out in the feature space $\mathbb{F}$ to detect connectivity via reproducing kernels. In dealing with Granger causality detection using kernels as in [29,30], this stands in sharp contrast as the latter depend on solving the *reconstruction/pre-image* problem to provide prediction error estimates in the original data space $\mathbb{X}$. In fact, part of the challenge in pre-image determination lies in its frequently associated numerical ill-condition [31].

The key result behind doing model diagnostics and inference in $\mathbb{F}$ is (28) by realizing that kernel quantities may be normalized much as correlation coefficients. It should be noted that (28) holds even in the case of (nonkernel) linear modelling by replacing the **K** matrices by auto/crosscorrelation matrices, something that, in practice, is never adopted in classical linear time series modelling because the necessary auto/crosscorrelations are more efficiently computed from model residuals that are easy to obtain as no pre-imaging problem is involved there.

Thus, what importantly sets the present approach apart from previous work is the lack of need for returning to the original input space $\mathbb{X}$ to gauge model quality as the *reconstruction/pre-image problem* can be fully circumvented bypassing unnecessary uncertainties.

As such, we showed that, because model adequacy testing can be performed *directly* in the feature space $\mathbb{F}$, directional Granger type connectivity can be detected for a variety of multivariate nonlinear coupling scenarios, thereby totally dispensing with the need for detailed 'a priori' model knowledge.

We observed that successful connectivity detection is achievable at the expense of a relatively short time series. A systematic comparison with other approaches [4,5,32–35] is planned for future work, but, at least for the cases we tested so far, savings of at least one order magnitude in record lengths are feasible.

One of the basic tenants of the present work is that model coefficients in the feature space are asymptotically normal, something whose consistency was successfully illustrated though the need for a more formal proof remains, especially in connection to explicit kernel estimates under the total-least-squares solution to (23). Our choice of TLS was dictated by its apparent superiority when compared to the 'kernel trick' [32] whose multivariate version we employed in [26,36,37].

In this context, it is important to note that, contrary to other methods that require time-consuming resampling procedures for adequate inference, the present approach relies on asymptotic statistics and is thus less susceptible to eventual problems derived from data shuffling.

One of the advantages of the present development is that the procedure allows for determining how far in the past to look via the model order criteria we employed (27).

Even though order estimation and model testing were successful upon borrowing from the usual linear modelling practices, further systematic examination is still needed and is under way.

One may rightfully argue that the kernels we chose for illustrating the present work are equivalent to modelling the original time series after the application of a suitable $\phi(\cdot)$ transformation to the data and that they look for the causality evidence present in higher order momenta. This, in fact, explains why quadratic kernels converge much faster than quartic ones in Example 1. The merit of framing the time series transformation discussion for connectivity detection in terms of kernels produces a simple workflow and paves the way to developing future data-driven criteria towards optimum data transformation choice for a given problem. Other kernel choices are being investigated.

The signal model used in the present development does not contemplate additive measurement noise whose impact on connectivity detection we also leave for future examination.

One thing the present type of analysis cannot do is expose details of how the nonlinearity takes place. For example, coupling may be quadratic or involve higher exponential powers or some other function. What the present approach can do, however, is to expose existing connections, so that modelling efforts can be concentrated on them, thereby avoiding modelling parameter waste on non relevant links.

Finally, the present systematic empirical investigation sets the proposal of using feature space-frequency domain descriptions of connectivity like *kernel partial directed coherence* [26,36], and *kernel directed transfer function* [37] on sound footing, especially with respect to their asymptotic connectivity behaviour.

Author Contributions: All authors conceptualized, cured the data, did the formal analysis, acquired the funding, did the investigation, obtained the resources, analyzed and interpreted the data, created the software used in the work, did the supervision, validation, visualization, drafted the work, and revised it.

Funding: L.M. was funded by in part by an institutional Ph.D. CAPES Grant at Escola Politécnica, University of São Paulo, São Paulo, Brazil, and by CAPES, grant number 88887.161474/2017 − 00 (PALEOCEANO Project) and by FAPESP, grant number 2015/50686-1 (PACMEDY Project), both at Instituto de Astronomia, Geofísica e Ciências Atmosféricas, University of São Paulo, São Paulo, Brazil. L.A.B. was funded by CNPq, grant number 308073/2017-7.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schumacker, R.E.; Lomax, R.G. *A Beginner'S Guide to Structural Equation Modeling*, 4th ed.; Taylor & Francis: New York, NY, USA, 2016.

2. Applebaum, D. *Probability and Information: An Integrated Approach*, 2nd ed.; Cambridge University Press: Cambridge: New York, NY, USA, 2008; p. 273. [CrossRef]
3. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley Series in Telecommunications; Wiley: New York, NY, USA, 2006; p. 774.
4. Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep. Rev. Sect. Phys. Lett.* **2007**, *441*, 1–46. [CrossRef]
5. Schreiber, T. Measuring Information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef] [PubMed]
6. Bendat, J.S.; Piersol, A.G. *Engineering Applications of Correlation and Spectral Analysis*; Wiley: New York, NY, USA; Chichester, UK, 1980.
7. Baccalá, L.A.; Sameshima, K. Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. *Prog. Brain Res.* **2001**, *130*, 33–47. [CrossRef] [PubMed]
8. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438. [CrossRef]
9. Baccalá, L.A.; Sameshima, K. Multivariate time series brain connectivity: A sum up. *Methods in Brain Connectivity Inference Through Multivariate Time Series Analysis*; CRC Press: Boca Raton, FL, USA, 2014; pp. 245–251.
10. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin, Germany, 2005.
11. Baccalá, L.A.; Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.* **2001**, *84*, 463–474. [CrossRef] [PubMed]
12. Vapnik, V.N. *Statistical Learning Theory*, 1st ed.; John Wiley and Sons: Hoboken, NJ, USA, 1998; p. 736.
13. Priestley, M.B. *Spectral Analysis and Time Series*; Probability and Mathematical Statistics; Academic Press London: New York, NY, USA, 1981; p. 890.
14. Nikias, C.; Petropulu, A.P. *Higher Order Spectra Analysis: A Non-linear Signal Processing Framework*; Prentice Hall Signal Processing Series; Prentice Hall: Upper Saddle River, NJ, USA, 1993; p. 528.
15. Subba-Rao, T.; Gabr, M.M. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*; Lecture Notes in Statistics; Springer: New York, NY, USA, 1984. [CrossRef] [PubMed]
16. Schelter, B.; Winterhalder, M.; Eichler, M.; Peifer, M.; Hellwig, B.; Guschlbauer, B.; Lücking, C.H.; Dahlhaus, R.; Timmer, J. Testing for directed influences among neural signals using partial directed coherence. *J. Neurosci. Methods* **2006**, *152*, 210–219. [CrossRef] [PubMed]
17. Massaroppe, L.; Baccalá, L.A.; Sameshima, K. Semiparametric detection of nonlinear causal coupling using partial directed coherence. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 5927–5930. [CrossRef]
18. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
19. Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **1909**, *A*, 415–446. [CrossRef]
20. Golub, G.H.; van Loan, C.F. *Matrix Computations*, 4th ed.; Number 3 in Johns Hopkins Studies in the Mathematical Sciences; Johns Hopkins University Press: Baltimore, MD, USA, 2013; p. 784.
21. Hable, R. Asymptotic normality of support vector machine variants and other regularized kernel methods. *J. Multivar. Anal.* **2012**, *106*, 92–117. [CrossRef]
22. Filliben, J.J. The probability plot correlation coefficient test for normality. *Technometrics* **1975**, *17*, 111–117. [CrossRef]
23. Vogel, R.M. The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses. *Water Resour. Res.* **1986**, *22*, 587–590. [CrossRef]
24. Vogel, R.M. Correction to "The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses". *Water Resour. Res.* **1987**, *23*, 2013–2013. [CrossRef]
25. Massaroppe, L.; Baccalá, L.A. Método semi-paramétrico para inferência de conectividade não-linear entre séries temporais. In Proceedings of the Anais do I Congresso de Matemática e Computacional da Região Sudeste, I CMAC Sudeste, Uberlândia, Brazil, 20–23 September 2011; pp. 293–296.
26. Massaroppe, L.; Baccalá, L.A. Detecting nonlinear Granger causality via the kernelization of partial directed coherence. In Proceedings of the 60th World Statistics Congress of the International Statistical Institute, ISI2015, Rio de Janeiro, RJ, USA, 26–31 July 2015; pp. 2036–2041.

27. Gourévitch, B.; Bouquin-Jeannès, R.L.; Faucon, G. Linear and nonlinear causality between signals: Methods, examples and neurophysiological applications. *Biol. Cybern.* **2006**, *95*, 349–369. [CrossRef] [PubMed]
28. Chen, Y.; Rangarajan, G.; Feng, J.; Ding, M. Analyzing multiple nonlinear time series with extended Granger causality. *Phys. Lett. A* **2004**, *324*, 26–35. [CrossRef]
29. Amblard, P.O.; Vincent, R.; Michel, O.J.J.; Richard, C. Kernelizing Geweke's measures of Granger causality. In Proceedings of the 2012 IEEE International Workshop on Machine Learning for Signal Processing, Santander, Spain, 23–26 September 2012; pp. 1–6. [CrossRef]
30. Kallas, M.; Honeine, P.; Francis, C.; Amoud, H. Kernel autoregressive models using Yule-Walker equations. *Signal Process.* **2013**, *93*, 3053–3061. [CrossRef]
31. Honeine, P.; Richard, C. Preimage problem in kernel-based machine learning. *IEEE Signal Process. Mag.* **2011**, *28*, 77–88. [CrossRef]
32. Kumar, R.; Jawahar, C.V. Kernel approach to autoregressive modeling. In Proceedings of the Thirteenth National Conference on Communications (NCC 2007), Kanpur, India, 26–28 January 2007; pp. 99–102.
33. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103. [CrossRef] [PubMed]
34. Park, I.; Príncipe, J.C. Correntropy based Granger causality. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 3605–3608. [CrossRef]
35. Príncipe, J.C. *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*, 1st ed.; Number XIV in Information Science and Statistics; Springer Publishing Company, Incorporated: New York, NY, USA, 2010; p. 448. [CrossRef] [PubMed]
36. Massaroppe, L.; Baccalá, L.A. Kernel-nonlinear-PDC extends Partial Directed Coherence to detecting nonlinear causal coupling. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 2864–2867. [CrossRef]
37. Massaroppe, L.; Baccalá, L.A. Causal connectivity via kernel methods: Advances and challenges. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016.