WILEY | Hindawi

Research Article

Tracking Nonlinear Correlation for Complex Dynamic Systems Using a Windowed Error Reduction Ratio Method

Yifan Zhao,[1] Edward Hanna,[2] Grant R. Bigg,[3] and Yitian Zhao[4,5]

[1]School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, UK
[2]School of Geography, University of Lincoln, Lincoln, UK
[3]Department of Geography, University of Sheffield, Sheffield, UK
[4]Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo, China
[5]School of Optics and Electronics, Beijing Institute of Technology, Beijing, China

Correspondence should be addressed to Yifan Zhao; yifan.zhao@cranfield.ac.uk and Yitian Zhao; yitian.zhao@bit.edu.cn

Studying complex dynamic systems is usually very challenging due to limited prior knowledge and high complexity of relationships between interconnected components. Current methods either are like a "black box" that is difficult to understand and relate back to the underlying system or have limited universality and applicability due to too many assumptions. This paper proposes a time-varying Nonlinear Finite Impulse Response model to estimate the multiple features of correlation among measurements including direction, strength, significance, latency, correlation type, and nonlinearity. The dynamic behaviours of correlation are tracked through a sliding window approach based on the Blackman window rather than the simple truncation by a Rectangular window. This method is particularly useful for a system that has very little prior knowledge and the interaction between measurements is nonlinear, time-varying, rapidly changing, or of short duration. Simulation results suggest that the proposed tracking approach significantly reduces the sensitivity of correlation estimation against the window size. Such a method will improve the applicability and robustness of correlation analysis for complex systems. A real application to environmental changing data demonstrates the potential of the proposed method by revealing and characterising hidden information contained within measurements, which is usually "invisible" for conventional methods.

1. Introduction

Complex systems feature a large number of measurable components interacting simultaneously and nonlinearly with each other and their environments on multiple levels [1]. Social systems consisting of people, the weather formed out of air flows, the brain formed out of neurones, and even a Computer Numeric Control (CNC) machine formed out of components are all examples of complex systems. In fact, most real systems are truly complex [2]. Complex systems are typically studied based on modelling of the observed measurements of systems. The conversion of these observed measurements into a model about a physical object or system has been one of the most important Inverse Problems because it starts with the results and then calculates the causes.

Some advanced parametric or nonparametric modelling approaches, such as Neural Networks, tend to be complicated to provide excellent prediction or approximations to the given dataset. But the disadvantage is that the outcomes are often opaque, are difficult to relate back to the underlying system, and are difficult to analyse and therefore lack transparency or interpretability [3].

Alternatively, some research focuses on revealing the hidden interdependencies between measurements, which enables a simplified and more feasible way to understand complex systems. There are several ways to tackle this problem either based on using an explicit generative model that embraces the known causal architecture [4] or by simply establishing quantitative dependencies between two signals using cross-correlation, coherence, phase synchronisation, or

the Granger causality test. The most commonly used methods are cross-correlation and coherence, which are easy-to-use and computational inexpensive but they usually assume that the system is linear and stationary [5, 6]. They, therefore, cannot sufficiently reveal and characterise hidden correlation between complex signals that is unusually nonlinear and dynamical. Fast time variation is associated with applications where rapid changes have been observed, which is usually nonlinear, for example, abrupt climate change [7], emergence of infectious diseases [8], epileptic patients moving from a normal state to a seizure [9], or a severe fault suddenly arising in a machine [10]. These methods cannot be applied in any of these scenarios.

A well-established and fundamental approach to detect causal influence between two coupled signals is the Granger causality test [11]. To calculate the Granger causality of the input X to the output Y, an unbiased model has to be preestablished which defines the relationship between the present Y and past information Y^- and past information of the input X^-, expressed as $Y = f(Y^-, X^-)$. This model can be nonlinear autoregressive and nonlinear autoregressive exogenous models, the theory of reproducing kernel Hilbert spaces, and so on. Based on the sampled data the parameters in the model $f(Y^-, X^-)$ are estimated and then the predictions of Y based on Y^- alone and on Y^- and X^- together are generated. The Granger causality is defined as the ratio between the variances of both prediction errors, which basically quantifies the prediction improvement. One advantage of this method is that if the model structure is chosen appropriately, the approach can be used for both linear and nonlinear systems. However, for a complex system, this condition cannot be always satisfied. Another important class method to study interactions in bivariate time series is cross prediction based. The main difference between this method and Granger causality is that Granger causality must consider the past information of the output Y in the relationship between Y and X while the cross prediction methods are not necessary. Cross prediction is valuable to quantify the coupling strength and predictability improvement to elicit directionality of the interactions in short and noisy bivariate time series. Recently some nonparametric methods have been developed. One well-established method is based on transfer entropy that allows nonlinear and model-free estimation [12]. However, in general, nonparametric methods tend to require larger datasets or averaging over many realisations to mitigate the effects of noise. The noise in the signals is usually unknown prior to analysis but simple averaging methods will not work well if the noise is highly correlated and nonlinear, which may be expected if the relationships are also nonlinear.

This paper introduces a new parametric nonlinear correlation analysis approach, called Windowed Error Reduction Ratio (WERR), which directly measures the correlation between input signals and output signals without producing a full model. Similar to the cross prediction methods, in this method the past information of the output is not considered. Using this method, complex systems can become analysable, especially systems (a) without or with very little prior knowledge of model structure, (b) with nonlinear and nonstationary relationships between the variables, (c) where

observed data length is too limited to identify a full unbiased model, or (d) with rapid change of correlations. The paper is organised as follows. The methodology of the WERR method is introduced in Section 2. Three simulation examples are presented in Section 3 to investigate the influence of the selection of the window function on the accuracy and precision of the onset detection of correlation, the detectability of a short-duration correlation between signals, and the sensitivity of a specific parameter on the results. A real application to environmental change data is then presented in Section 4 to demonstrate how to use the proposed method to solve real problems. Conclusions are given in Section 5.

2. Method

Considering a complex system with multiple measurable variables, the aim of this paper is to quantitatively measure and track over time the correlation between these variables. The properties of correlation that this paper primarily focuses on include (a) direction, the casual direction among variables (defining inputs and outputs); (b) strength, a value ranging from 0% to 100% to quantify the correlation level between an input and an output; (c) significance, a significance level above which the estimated correlation is significant; (d) latency, the time delay between the input and the output; (e) correlation type, whether the input makes a positive or negative impact on the output; and (f) nonlinearity, the percentage of nonlinear correlation strength in the total correlation. For most systems, the definition of inputs and outputs is known a priori (e.g., a multiple inputs and multiple outputs system (MIMO)), so the direction of correlation is therefore known. This paper primarily focuses on such a MIMO system.

2.1. Time-Varying NFIR Model. Considering a dynamic MIMO system with input time series $\{u_1, u_2, \ldots, u_A\}$ and time-varying output time series $\{y_1, y_2, \ldots, y_B\}$, where A and B denote the number of inputs and outputs respectively. To quantify and track the correlation of each input to a considered output y_j, this paper uses a Nonlinear Finite Impulse Response (NFIR) model, also known as the Volterra Nonlinear Regressive with eXogenous (VNRX) Inputs model, to represent a MIMO system. It can be expressed as

$$y_j(t) = f\left(u_1^{[t-n_1]}, u_2^{[t-n_2]}, \ldots, u_A^{[t-n_A]}\right) + \varepsilon_j(t), \quad (1)$$

where t ($t = 1, 2, \ldots$) is a time index, f is an unknown linear or nonlinear mapping which links the system output y_j to the system inputs; $\varepsilon_j(t)$ denotes the model residual. The symbol $u_i^{[t-n_i]}$ ($i = 1, 2, \ldots, A$) denotes the current and past information of the input u_i, which can be expanded as

$$u_i^{[t-n_i]} = \bigcup_{j=0}^{n_i} u_i(t-j), \quad (2)$$

where n_i is the maximum temporal lag to be considered for the input u_i.

If the system is time-invariant, a commonly employed implementation to specify the function f in (1) is a polynomial function [13], which can be expressed as

$$y_j = \theta_0 + \sum_{m=1}^{N} \theta_m p_m + \varepsilon_j, \qquad (3)$$

where p_m is the mth model term selected from a candidate terms' set constructed from all input vectors. Note that p_m, in general, can be linear or nonlinear. The constant θ_m is the coefficient of each term; N is the total number of model terms.

If the system is time-variant, (3) can be extended to

$$y_j(t) = \theta_0(t) + \sum_{m=1}^{N(t)} \theta_m(t) p_m(t) + \varepsilon_j(t), \qquad (4)$$

where the number of terms, the model structure, and the coefficients are all time-varying. This paper focuses on this type of system.

If the model order is set as q, the candidate term set, denoted by C, can be expressed as

$$C = C_1 \cup C_2 \cup \cdots \cup C_l \cup \cdots \cup C_q, \qquad (5)$$

where C_1 is the linear term set, expressed as

$$C_1 = \bigcup_{a=1}^{A} u_a^{[t-n_a]}, \qquad (6)$$

C_2 is the 2nd-order nonlinear term set, expressed as

$$C_2 = \bigcup_{a_1=1}^{A} \bigcup_{a_2=a_1}^{A} u_{a_1}^{[t-n_{a_1}]} u_{a_2}^{[t-n_{a_2}]}, \qquad (7)$$

and C_l is the 2nd-order nonlinear term set, expressed as

$$C_l = \bigcup_{a_1=1}^{A} \bigcup_{a_2=a_1}^{A} \cdots \bigcup_{a_l=a_{l-1}}^{A} \prod_{i=1}^{l} u_{a_i}^{[t-n_{a_i}]}, \qquad (8)$$

If the maximum of time lag for each input is the same, denoted by n, the number of total candidate terms is

$$M = \frac{(n+q+1)!}{q!\,(n+1)!}. \qquad (9)$$

If the inputs and output of a system are observable, the model represented by (4) can then be identified based on the principle of least square errors. In this paper, the orthogonal least squares (OLS) algorithm [14] is used to determine the model structure from the observations and estimate the unknown parameters θ_m. The OLS algorithm searches through the candidate model term set C to select the most significant model terms which are then included to build the model term by term. The significance of each selected model term is measured by an index, called the Error Reduction Ratio (ERR), which indicates how much of the variance change in the system response is caused by the considered term, in a percentage form.

Because the system is time-varying, the model structure and coefficients vary with time. Before applying the OLS method, the way to address this time variability must be determined. This paper proposes to apply a window function $W = w(t)$, $(n = 1, 2, \ldots, H)$ on the inputs and the considered output, where H is the window size. The process of model identification and correlation quantification is applied on each window, by which the correlation can be tracked through sliding the window. The time resolution of tracking is determined by the sliding step. The simplest window function is the Rectangle window, where the weighting of the input at different times is the same. A challenge for all sliding-window-based methods is the selection of window size. A selection of a small window size indicates a fast response to the change of correlation over time, but it may produce inaccurate correlation measurement due to insufficient sampling. On the other hand, a selection of a large window size can improve the accuracy of correlation measurement, but it may slow down the model response to the change of correlation over time, so the time resolution is reduced. Under this condition, there will be a higher chance that the model misses a short-duration correlation. This paper proposes to use a unique nonrectangle window function to reduce the influence of the window size. Details of this will be discussed in Section 3.

2.2. Tracking Correlation Strength. Given the windowed output y_j within the window $[t - H/2 + 1, t + H/2]$, denoted as $y_{j,w}[t - H/2 + 1, t + H/2]$, and the windowed inputs within the same window, denoted as $z_i[t - H/2 + 1, t + H/2]$ ($i = 1, 2, \ldots, A$), where

$$z_i\left[t - \frac{H}{2} + 1, t + \frac{H}{2}\right] = u_i\left[t - \frac{H}{2} + 1, t + \frac{H}{2}\right] \cdot W, \qquad (10)$$

$$y_{j,w}\left[t - \frac{H}{2} + 1, t + \frac{H}{2}\right] = y\left[t - \frac{H}{2} + 1, t + \frac{H}{2}\right] \cdot W, \qquad (11)$$

(4) can be rewritten as

$$Y_j = P\Theta, \qquad (12)$$

where

$$Y_j = \begin{bmatrix} y_{j,w}\left(t - \dfrac{H}{2} + 1\right) \\ y_{j,w}\left(t - \dfrac{H}{2} + 2\right) \\ \vdots \\ y_{j,w}\left(t + \dfrac{H}{2}\right) \end{bmatrix},$$

$$P = \begin{bmatrix} P^T(1) \\ P^T(2) \\ \vdots \\ P^T(H) \end{bmatrix},$$

$$\Theta = \begin{bmatrix} \theta(1) \\ \theta(2) \\ \vdots \\ \theta(M) \end{bmatrix}, \tag{13}$$

where

$$P^T(k) = (p_1(k), p_2(k), \ldots, p_M(k)). \tag{14}$$

The symbol $p_i(k)$ is a candidate term selected from the term set C constructed by the windowed inputs Z_i. The symbol M denotes the number of all candidate terms.

Matrix P can be decomposed as $P = X \times V$ where

$$X = \begin{bmatrix} x_1(1) & x_2(1) & \ldots & x_M(1) \\ x_1(2) & x_2(2) & \ldots & x_M(2) \\ \vdots & \ddots & \ddots & \vdots \\ x_1(H) & x_2(H) & \ldots & x_M(H) \end{bmatrix} \tag{15}$$

and V is an upper triangular matrix with unity diagonal elements

$$V = \begin{bmatrix} 1 & e_{12} & e_{13} & \ldots & e_{1M} \\ & 1 & e_{23} & \ldots & e_{2M} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & e_{M-1M} \\ & & & & 1 \end{bmatrix}. \tag{16}$$

Therefore, (12) can be rewritten as

$$Y_j = XG, \tag{17}$$

where $G = V\Theta = \begin{bmatrix} g_1 & g_2 & \cdots & g_M \end{bmatrix}^T$.

The next step is to estimate the contribution of each term to the variation of the system output. Initially, the algorithm sets values $e_{11} = 1$, $x_1(k) = p_1(k)$ and estimates

$$\widehat{g}_1 = \frac{\sum_{k=1}^{H} x_1(k) y_{j,w}(k)}{\sum_{k=1}^{H} x_1^2(k)}. \tag{18}$$

For $b = 2, 3, \ldots, H$, the algorithm sets $e_{bb} = 1$ and then calculates

$$b_{ib} = \frac{\sum_{k=1}^{H} x_i(k) p_b(k)}{\sum_{k=1}^{H} x_i^2(k)}, \tag{19}$$

where $i = 1, 2, \ldots, b-1$. Next calculate

$$x_b(k) = p_b(k) - \sum_{i=1}^{b-1} e_{ib} x_i(k) \tag{20}$$

and estimate

$$\widehat{g}_b = \frac{\sum_{k=1}^{H} x_b(k) y_{j,w}(k)}{\sum_{k=1}^{H} x_b^2(k)}. \tag{21}$$

The ERR value for each term p_i, as a criterion to measure the contribution of each term to the variation of the system output y_j, is defined as

$$\text{ERR}_{j,i}(t) = \frac{\widehat{g}_i^2 \sum_{k=1}^{H} x_i^2(k)}{\sum_{k=1}^{H} y_{j,w}^2(k)}. \tag{22}$$

Values of ERR range always from 0% to 100%. The larger the value of ERR, the higher the dependence between this term and the output. To stop the search procedure and determine the number of significant terms N, a criterion called Penalised Error-to-Signal Ratio (PESR) is used [15]. It can be written as

$$\text{PESR}_{j,m}(t) = \frac{1}{(1 - \lambda m/H)^2} \left(1 - \sum_{i=1}^{m} \text{ERR}_{j,i}(t) \right). \tag{23}$$

This criterion is introduced to monitor the search procedure, where m denotes the index of the selected terms. The search procedure stops when $\text{PESR}_{j,m}(t)$ achieves a local minimum. PESR has been successfully used to monitor the search of model structure for various applications [16, 17].

The effect of the adjustable parameter λ on the results is not much sensitive. Wei and Billings suggested that λ should be chosen between 5 and 10 [15]. Different values of λ have been tested in this study and there is no significant difference to the results. A sensitivity analysis of this parameter is presented in Section 3.4. In all examples in this paper, the value of λ is chosen as 6.

The estimation of the coefficient of each selected term can be computed from

$$\widehat{\theta}_N = \widehat{g}_N$$

$$\widehat{\theta}_i = \widehat{g}_i - \sum_{k=i+1}^{N} a_{ik}\theta_k, \quad i = N-1, \ldots, 1. \tag{24}$$

To calculate the contribution of each input to the output y_j, the sum of ERR of all selected terms, denoted by SERR_j, is calculated by

$$\text{SERR}_j(t) = \sum_{i=1}^{N} \text{ERR}_{j,i}(t). \tag{25}$$

Note that N is the number of the selected terms, not the number of total candidate terms, M. The value of SERR describes the percentage explained by the identified model to the system output. If the considered inputs can fully explain

the variation of the system output, the value of SERR is equal to 100%. The contribution of the ith input, u_i, to the variation of the system output y_j, denoted by $\text{WERR}_s(y_j, u_i, t)$, is defined as the sum of ERR values of the selected terms that include the windowed input variable z_i. Because some selected nonlinear terms may involve more than one input, the sum of WERR_s for all input can be greater than SERR. To overcome this problem, the value of WERR_s is normalised and written as

$$
\begin{aligned}
&\text{WERR}_s\left(y_j, u_i, t\right) \\
&= \frac{\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_i \in p_b\right)}{\sum_{a=1}^{A}\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_a \in p_b\right)} \times \text{SERR}_j(t).
\end{aligned}
\tag{26}
$$

The value of $\text{WERR}_s(y_j, u_i, t)$ is then always between 0% and 100%.

2.3. Significance Test. The calculated WERR_s value represents the strength of correlation between the input and the output, including both linear and nonlinear interactions. However, the reliability of the identified NFIR model depends on how many observed data points are sampled. Even for the same number of sampling data, the reliability is also affected by the model structure. Therefore, a hypothesis test is required to decide whether the correlation in the sampled data is strong enough to be used to model the correlation of the whole population. The hypothesis is that "the considered input has significant correlation with the considered output." The decision can be made by either using the p value or using a threshold. This paper uses the latter solution by introducing a significance threshold calculated by the Bootstrap Hypothesis Test. Bootstrapping operation builds the probability distribution of a measure/estimator by randomly resampling the data with replacement and recalculating the estimator value [18]. Repeating this process multiple times will lead to the description of an empirical distribution, from which mean, variance, and confidence intervals can be extracted. Here is the procedure.

Step 1. Resample. The inputs and output are chosen as white noise sequences. For the considered number of inputs A, the window size H, the window function W, the model order q, and the maximal time lag n, the value of $\text{WERR}_s(y, u_i, t)$ is calculated to establish a reference. This step is repeated hundreds of times. In this study, there were 100 repeats.

Step 2. Calculate the bootstrap distribution. If the sampled data length is sufficient, the WERR_s value of each input should have the same distribution. Therefore, the average contribution from all inputs is then calculated.

Step 3. Use the bootstrap distribution. The 95% quantile of the averaged distribution is determined as the threshold, expressed as $\tau(A, H, W, q, n)$. If the calculated WERR_s value of observed data is higher than this threshold, the correlation is determined as significant.

Such an approach has been applied in other research [19].

It should be noted that the calculated threshold of this approach is independent of the observed data, and a search table can therefore be established to reduce computational time. A more complex and accurate method is surrogate data based approach that resamples the observed time series randomly or resamples the phase of the data while preserving the amplitude [20]. This approach requires reapplying the significance test whenever the observed data are different, and usually, it requires high-computational time.

2.4. Estimation of Other Properties. To study the latency of correlation between an input and the output, this paper introduces a strength map in time-latency domain, written as $\text{WERR}_m(y_j, u_i, l, t)$, where u_i denotes the considered input, y_j denotes the considered output, l $(0 \le l \le n)$ denotes the time lag, and t is the time. The value of $\text{WERR}_m(y_j, u_i, l, t)$ can be computed by

$$
\begin{aligned}
&\text{WERR}_m\left(y_j, u_i, l, t\right) \\
&= \frac{\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_i(t-l) \in p_b\right)}{\sum_{a=1}^{A}\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_a \in p_b\right)} \times \text{SERR}_j(t).
\end{aligned}
\tag{27}
$$

The linearity of correlation is represented by the sum of ERR of the terms that are linear, and it can be computed by

$$
\begin{aligned}
&\text{WERR}_L\left(y_j, u_i, t\right) \\
&= \frac{\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_i \in p_b \in C_1\right)}{\sum_{a=1}^{A}\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_a \in p_b\right)} \times \text{SERR}_j(t).
\end{aligned}
\tag{28}
$$

The nonlinearity of correlation is represented by the sum of ERR of the terms that are nonlinear, and it can be computed by

$$
\begin{aligned}
&\text{WERR}_N\left(y_j, u_i, t\right) \\
&= \frac{\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_i \in p_b \notin C_1\right)}{\sum_{a=1}^{A}\sum_{b=1}^{N}\left(\text{ERR}_{j,b}(t) \mid z_a \in p_b\right)} \times \text{SERR}_j(t).
\end{aligned}
\tag{29}
$$

The correlation type of each term is determined by the sign of the estimated corresponding coefficients. If the coefficient is positive, the impact of this term on the output is positive; otherwise, the impact is negative. Determining the correlation type of the input variable to the output can be difficult if the nonlinear term is coupled by multiple inputs. Potential solutions to this issue can be using either the sign of linear term or the sign of the most important term of each input.

3. Selection of Window Functions

This section aims to investigate how the selection of window function and size affects the performance of the proposed WERR method, particularly in accuracy and precision of onset detection for rapid changing systems, accuracy of tracking slow changing systems, and the success rate for detection of short-duration correlation.
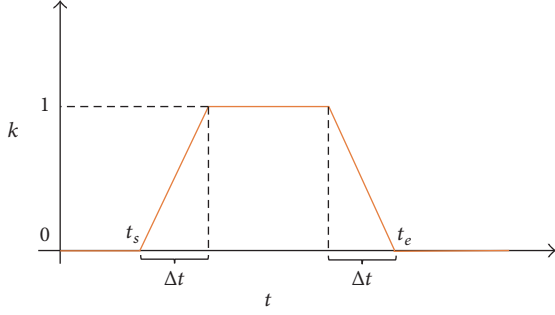
FIGURE 1: The function of k by time.

A total number of 500 data points were generated using a one input one output time-varying model, written as

$$y(t) = k(t)$$
$$\cdot [\theta_1 u(t-1) + \theta_2 u(t-2) - \theta_3 u(t-1) u(t-2)] \quad (30)$$
$$+ \varepsilon_y(t),$$

where $\varepsilon_y(t)$ is a white noise sequence with zero mean and the standard deviation of 0.1 and $u(t)$ is a random input data sequence uniformly distributed in $[-1, 1]$. The parameter $k(t)$ is the weight of correlation between u and y, which controls the change speed of correlation. The function of k can be illustrated by Figure 1, where the variables t_s and t_e are the start and end time of the correlation and Δt denotes the period of transition. The parameters were randomly selected as

$$\theta_1 = -0.07$$
$$\theta_2 = 0.32 \quad (31)$$
$$\theta_3 = -1.$$

3.1. Precision and Accuracy of Onset Detection.
Example 1 aims to assess the detection of the onset of interaction. In this example, the time when correlation starts was chosen as $t_s = 200$, the time when correlation ends was chosen as $t_e = 300$, and the period of transition Δt was chosen as 0. Five commonly used window functions, including Rectangular, Triangular, Hann, Blackman, and Welch, were tested. The mathematical representation of these windows is shown in Table 1. The NFIR model order q was chosen as 2, and the maximal time lag n was chosen as 4. The initial polynomial model therefore can be written as

$$y(t) = \theta_0(t) + \theta_1(t) u(t) + \theta_2(t) u(t-1) + \cdots$$
$$+ \theta_5(t) u(t-4) + \theta_6(t) u(t)^2$$
$$+ \theta_7(t) u(t) u(t-1) + \cdots$$
$$+ \theta_{10}(t) u(t) u(t-4) + \theta_{11}(t) u(t-1)^2$$
$$+ \theta_{12}(t) u(t-1) u(t-2) + \cdots$$

TABLE 1: Mathematical representation of tested windows functions, where N denotes the window size.

Window function	$w(n)$
Rectangular	1
Triangular	$1 - \left\| \dfrac{n - (N-1)/2}{N/2} \right\|$
Hann	$0.5 \left(1 - \cos \left(\dfrac{2\pi n}{N-1} \right) \right)$
Blackman	$0.42 - 0.5 \cos \left(\dfrac{2\pi n}{N-1} \right) + 0.08 \cos \left(\dfrac{4\pi n}{N-1} \right)$
Welch	$1 - \left(\dfrac{n - (N-1)/2}{(N-1)/2} \right)^2$

$$+ \theta_{14}(t) u(t-1) u(t-4) + \cdots$$
$$+ \theta_{20}(t) u(t-4)^2 + \varepsilon(t). \quad (32)$$

The total number of the candidate terms M is 21. The window size, H, was varied from 30 data points to 200 data points with a step of 1 point. For those times when the number of available data samples is smaller than the window size, such as the areas adjacent to the approaching the left and right boundaries, the calculation of WERR_s was neglected. The computed correlation strength between the input and the output using the proposed method is illustrated in Figure 2, as well as the ground truth. The ignored time interval (dark blue regions) increases following the increase of the windows size, which justifies the shape of reversed trapezoid for all five tested window functions. Inspection of the top right graph of Figure 2 shows that the value of the calculated correlation strength is dramatically affected by the window size when the Rectangular window is applied. A large window size leads to a lower strength than the ground truth due to the inclusion of data outside of $[t_s, t_e]$, which justifies the shape of the reversed triangular pattern. Results produced using the other four window functions present a more similar pattern to the ground truth, where the correlation strength is much less influenced by the window size.

A test to quantify the performance of each window function in onset detection of interaction has been conducted. The procedure can be summarised as follows:

(1) Produce the simulation data based on (30).

(2) For a selected window size H of a selected window function W, the proposed method was applied to calculate the value of $\mathrm{WERR}_s(y, u, t)$.

(3) Calculate the significance threshold $\tau(A, H, W, q, n)$, where $A = 1; q = 2; n = 4$ in this example.

(4) The first time when the calculated $\mathrm{WERR}_s(y, u, t)$ is above the threshold $\tau(A, H, W, q, n)$ was recorded and denoted by $\hat{t}_s$.

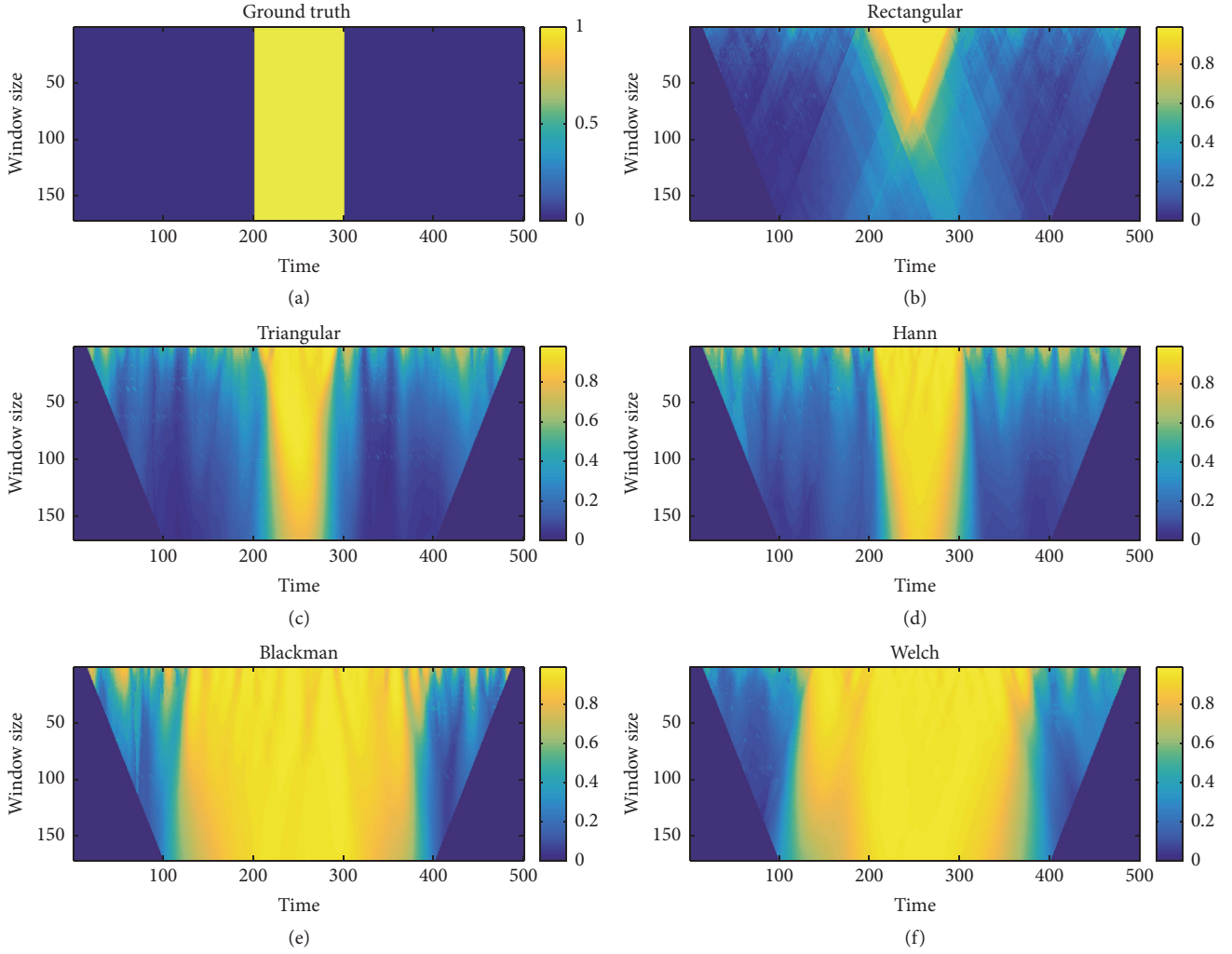(5) Repeat steps (1) to (4) for 100 times to achieve statistical significance of testing.

FIGURE 2: Visualisation of computed WERR$_s$ against the window size using different window functions for Example 1. The x-axis denotes the time (data point), the y-axis denotes window size, and the color intensity denotes the value of WERR$_s$. (a) shows the ground truth.

(6) Repeat step (1) to step (5) by varying the window size from 30 points to 200 points with the step of 1 point.

(7) Compute the mean, standard variation, and histogram of $\widehat{t}_s$.

(8) Repeat step (1) to step (7) by changing the window function.

Figure 3 illustrates the computed τ for five tested windows, where it can be observed that the significance threshold exponentially decreases following the increment of window size. It can also be observed that, for the same window size, the Rectangular window consistently has the lowest significance threshold and the Blackman window has the highest threshold. Note that the value of τ needs to be recalculated whenever any parameters are changed.

Figure 4 shows the histogram of the estimated $\widehat{t}_s$ for all tests, where the window size varies from 30 to 200 points with a step of 1 point, and each window size was implemented for 100 times. The ground truth of $\widehat{t}_s$ is 200. It can be clearly
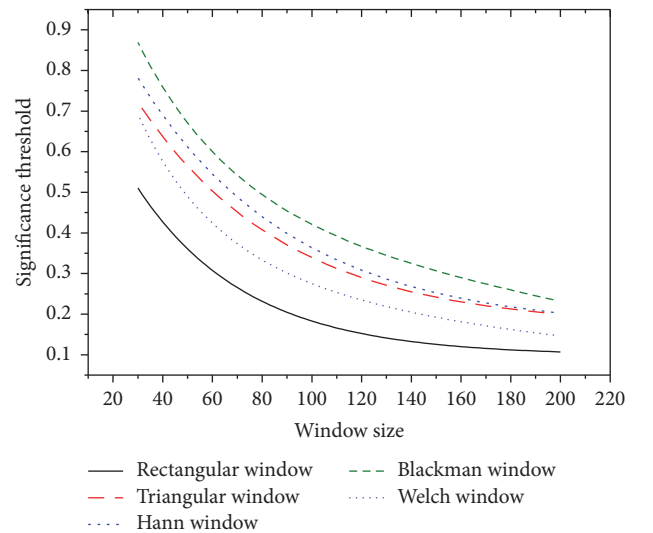
FIGURE 3: Significance thresholds against the window size for the five window functions, where the parameters are $A = 1; q = 2; n = 4$.
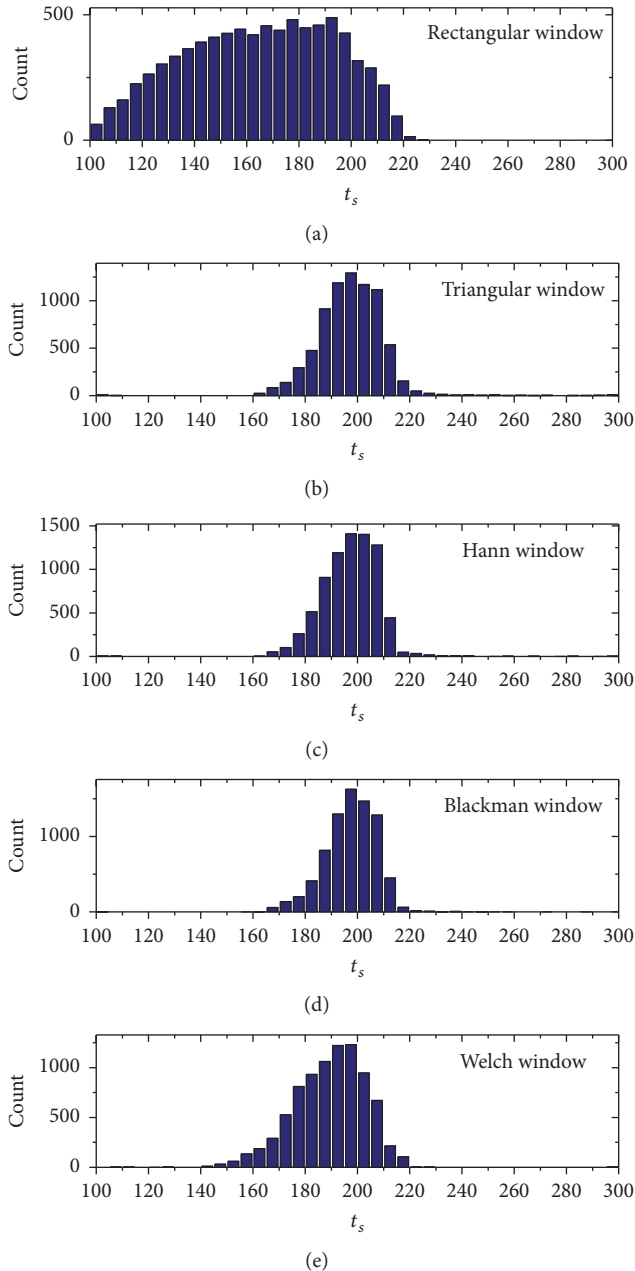
(a)

(b)

(c)

(d)

(e)

FIGURE 4: Histograms of the point of the significant correlation start for five windows functions. The window size varies from 30 to 200 and each window size was implemented 100 times for statistical significance.

observed that the Rectangular window function produces the worst result since the size of window has most effect on the results. The other four window functions produced more accurate and precise results, with the distributions being closer to Gaussian. Table 2 shows the mean and standard deviation of estimated $\hat{t}_s$, which indicate the accuracy and precision of onset detection, respectively. As expected, the Rectangular window function has the significantly worst performance in both accuracy (159.21) and precision (34.39). The Welch and Blackman window functions produce the best two
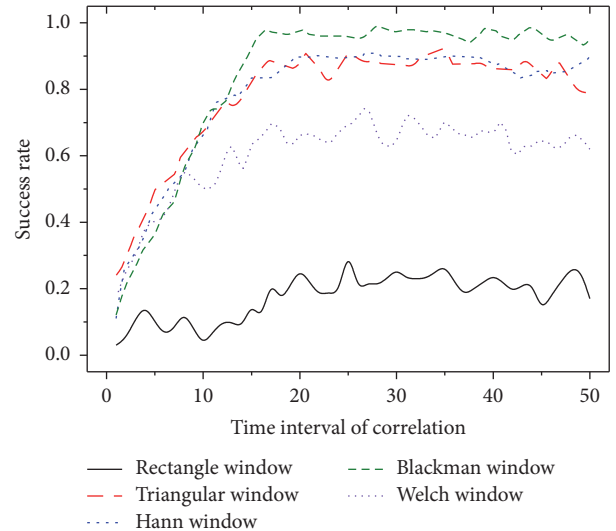
FIGURE 5: Performance to detect the short-interval correlation with different window functions, where the x-axis denotes the time interval (data points) of correlation and y-axis shows the success rate calculated based on 100 tests for each time interval. The window size was chosen as 50 data points.

TABLE 2: Precision and accuracy to detect the start of significant correction for Example 1.

Window function	Mean	Standard deviation
Rectangular	159.21	34.39
Triangular	196.81	23.92
Hann	194.92	17.94
Blackman	195.90	14.94
Welch	188.72	14.21

performances in precision, 14.21 and 14.94, respectively, while the Triangular and Blackman window functions produce the best two performances in accuracy, 196.81 and 195.90, respectively. The Blackman window function therefore has the top two performances in both accuracy and precision for onset detection.

3.2. Detection of Short-Duration Correlation. Example 2 aims to assess the performance of five window functions against short-duration interaction. Such a capability is important in some real applications, such as EEG signal processing where the correlation between channels can last for a very short time. The same model shown in (30) and (31) was used, but the time duration of correlation, $t_e - t_s$, was varied from 1 data point to 50 data points. If the detected onset time, $\hat{t}_s$, is in the range of $[t_s - \epsilon, t_e + \epsilon]$, the detection is classified as success. The symbol ϵ is a tolerance that was set as 10 data points in this example. For each selected correlation duration, a total of 100 tests were repeated and the success rate was computed. Figure 5 shows the results where the window size was chosen as 50 data points. The Rectangle window function has significantly lower performance than the other four, with the success rate being lower than 30% for
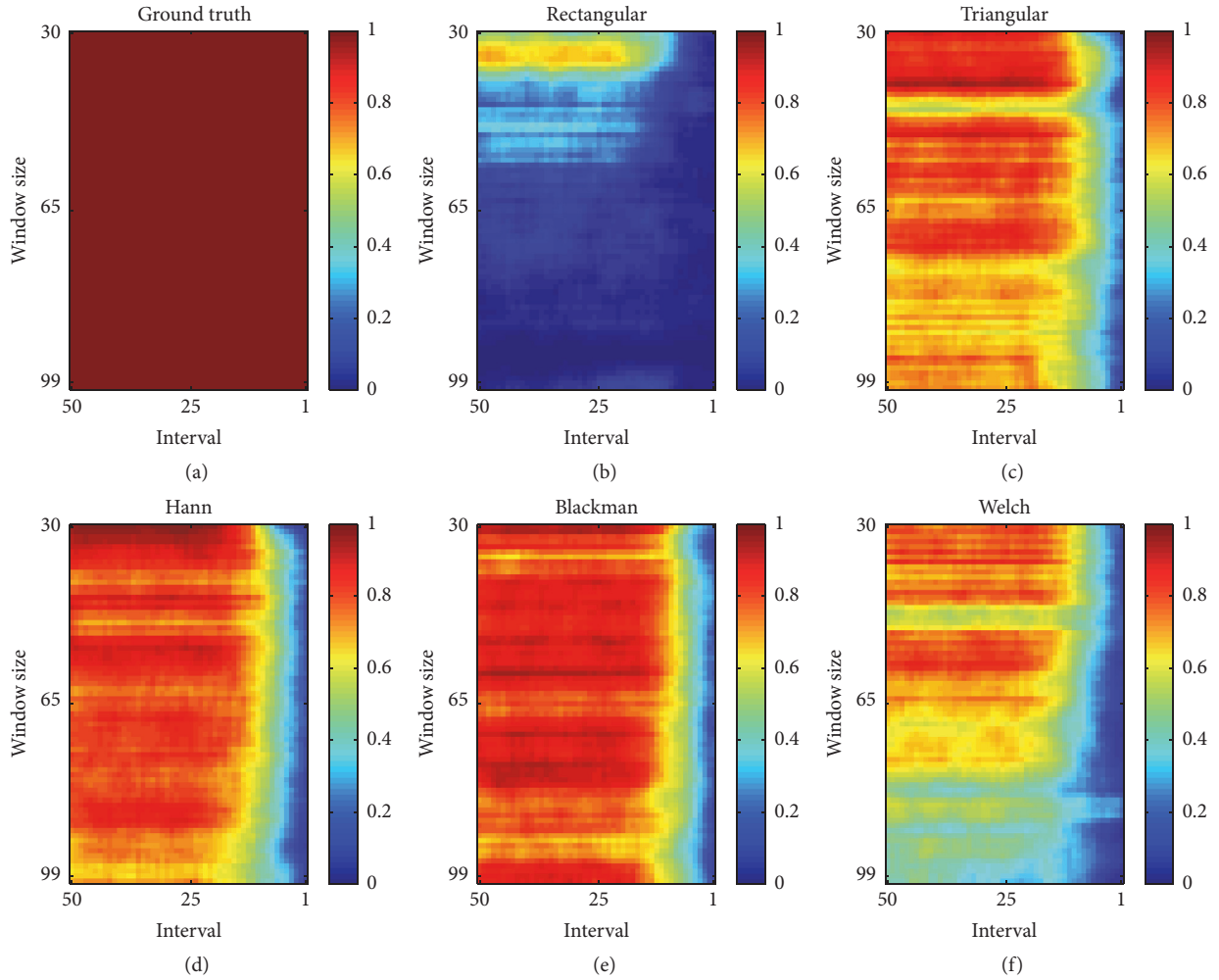
FIGURE 6: Full performance to detect the short-interval correlation with different window functions, where x-axis denotes the time interval (data points) of correlation and y-axis shows the windows size and the color intensity indicates the success rate calculated based on 100 repeated tests.

all considered durations. The Welch window has the second worst performance where the success rate is no higher than 80%. The remaining three functions present similar patterns that when the duration is more than 15 data points, the success rate is more than 80% and consistent, and when the duration is smaller than 15 data points, the success rate starts to drop dramatically. Furthermore, the Blackman window function, again, produces the best performance, with the success rate being almost 100% when the interval of correlation is larger than 15 data points.

The above test only considers one selection of the window size. To fully evaluate the performance under different window sizes, Figure 6 shows the success rate maps for five window functions, where the x-axis denotes the time duration (data points) of interaction, and the y-axis denotes the window size, varying from 30 to 100 data points. The color intensity indicates the success rate calculated based on 100 repeated tests. In terms of the success rate, the Blackman and Hann window functions show the best two performances, evident by the color patterns. Furthermore,

the Hann window function shows a sharper reduction of the success rate than the Blackman window function when the window size is more than 80 data points, which indicates that the latter window is less sensitive to the window size. Therefore, the Blackman window shows the best performance in both success rate and precision to detect short-duration correlation.

3.3. Tracking Slow Changing System. Example 3 aims to assess the performance of tracking a slow time-varying system. In this example, the time when correlation starts was chosen as $t_s = 100$, the time when correlation ends was chosen as $t_e = 400$, and the period of transition Δt was chosen as 100. The parameter setting of the proposed method was the same as that of Example 1. The results are shown in Figure 7, where the contour of the estimated WERR_s map was illustrated for better comparison of gradient. It can be inferred that the Rectangle window function has the worst performance in terms of the sensitivity to window size based on the observation of the shape of trapezoid of the
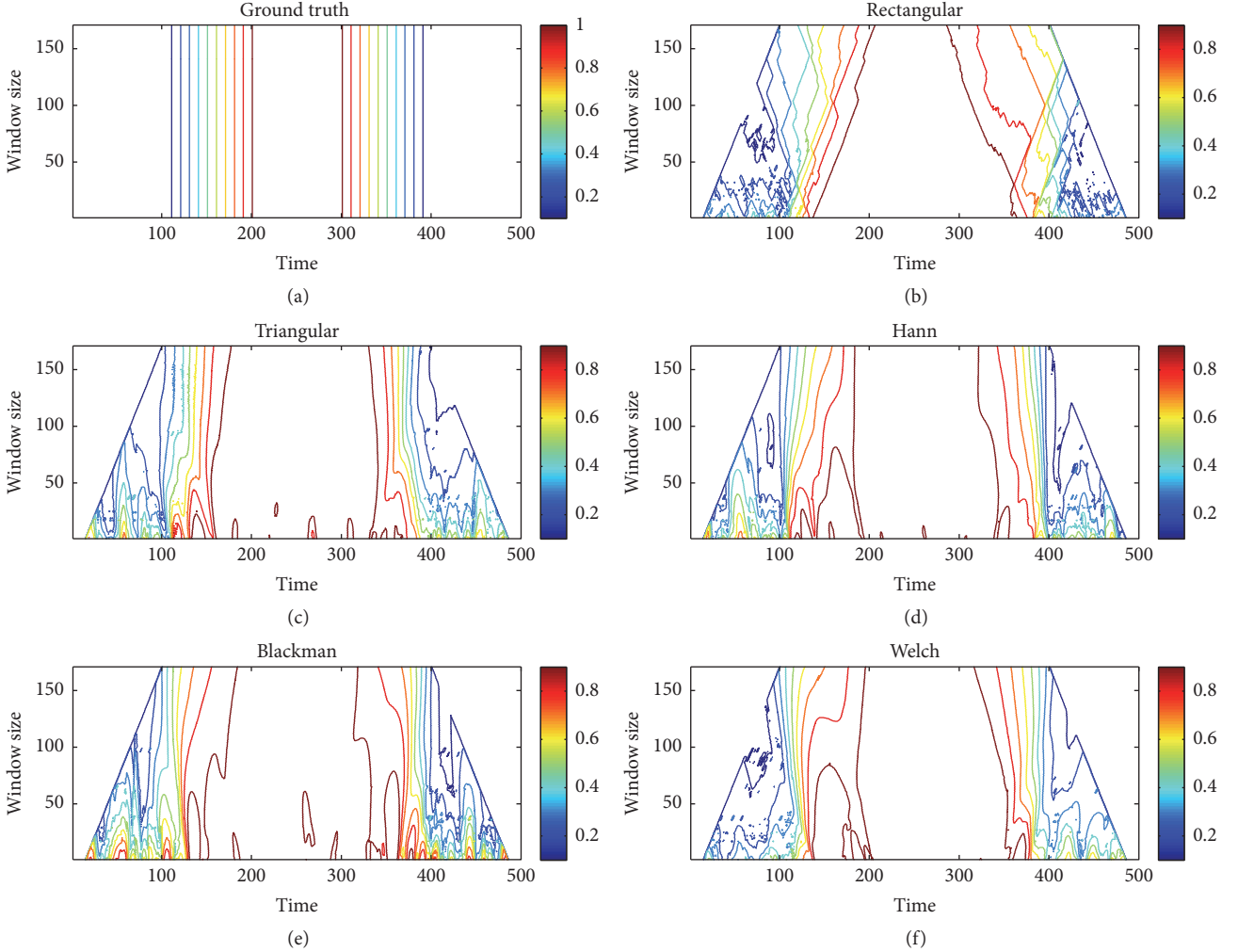
FIGURE 7: Contour visualisation of computed WERR$_s$ against the window size using different window functions for Example 3. The x-axis denotes the time (data point), the y-axis denotes window size, and the color intensity denotes the value of WERR$_s$. (a) shows the ground truth.

inner contour. The Triangular, Hann, and Blackman windows slightly compromise the performance in precision for small window size based on the observation of several islands. The Welch window appears to have the best performance in terms of both sensitivity to window size and the number of sampled data. These observations indicate that the Welch window has a slightly better performance to track a slow time-varying system.

3.4. Sensitivity Analysis of λ. This section aims to investigate how the selection of λ in (23) affects the model structure determination and strength estimation. In this example, a total number of 500 data points were generated using the model (30) and corresponding parameter setting (31) where $k(t)$ was set as 1 and the Blackman window was chosen. The setting of other parameters is $q = 2, n = 4$, and $H = 100$. Table 3 shows the produced PESR values against different values of λ, where the first valley is highlighted by bold font. It has been observed that when λ is chosen from 3 to 7, the selected number of terms is 3 which is correct based on the

model (30). When λ is larger than 7, less terms are selected, which introduces a false negative indication of correlation. When λ is smaller than 3, more terms are selected, which introduces a false positive indication of correlation. Further inspection of Table 3 shows that although the selection of λ can affect the selection of model structure, the effect on the estimated strength of correlation is not significant.

4. Application to Environmental Change Data

The proposed approach can be applied in a complex system where variables are measurable in spatial and temporal domains, and the correlation between them is of interest. This section introduces a real application of the WERR method to understand the variation of climatic and glaciological contributions to West Greenland iceberg discharge in the twentieth century.

Iceberg discharge is a major component of the mass balance of the Greenland Ice Sheet (GrIS) [21, 22]. While bulk estimates of discharge variation over time exist, inferred

TABLE 3: PESR values for different numbers of selected model terms against different values of λ. The final number of model term N is chosen when the PESR reaches the first valley (highlighted by bold font).

Number of terms	λ						
	2	3	4	5	6	7	8
1	0.08319	0.08670	0.09042	0.09440	0.09864	0.10318	0.10803
2	0.02305	0.02515	0.02754	0.03029	0.03348	0.03719	**0.04157**
3	0.01603	**0.01840**	**0.02134**	**0.02505**	**0.02981**	**0.03607**	0.04453
4	**0.01533**	0.01864	0.02314	0.02950	0.03889	0.05360	0.07854
5	0.01634	0.02120	0.02859	0.04064	0.06230	0.10731	0.22704
6	0.01762	0.02461	0.03676	0.06077	0.11911	0.33086	2.97778
WERR	0.99340	0.99084	0.99084	0.99084	0.99084	0.99084	0.09830

remotely from measurements of grounding line ice velocities or Surface Mass Balance calculations, few detailed measurements of discharge itself from individual marine-terminating glaciers exist until recent years [23]. However, it has recently been shown, through a combination of ocean-iceberg modelling and nonlinear system identification, that the century-long record of iceberg numbers crossing $48°N$ in the West Atlantic is a good first-order proxy for discharge from at least South and West Greenland [24]. This example will measure, track, and explore the time-varying linear and nonlinear correlation of ice sheet, oceanic and climatic forcing of iceberg discharge from these areas over the twentieth century.

The input variables to be used in developing the windowed NFIR model were chosen, as in Zhao et al. [25], to be Surface Mass Balance (SMB), North Atlantic Oscillation (NAO), and the Labrador Sea Surface Temperature (LSST). The output variable to be modelled is I48N, and the monthly iceberg counts from the US Coast Guard's International Ice Patrol over 1900–2015. There were in total 1392 months of data. The parameters selected for the proposed WERR approach are as follows:

(a) The number of inputs, A in (1), was chosen as 3.

(b) The maximal time lag for each input, n_i in (2), was chosen as 48 months, which was determined based on prior knowledge and expert experience relating to Greenland calving processes and the mean speed of ocean circulation in the NW Atlantic.

(c) The model order, q in (5), was chosen as 2, which means that there are 11026 candidate terms. The third- or higher-order model is unfeasible in this case due to the computation complexity and limitation of sampling number (e.g., a third-order model has 518665 candidate terms).

(d) The Blackman window function was chosen due to its superior performance against other windows functions, as shown in Examples 1 and 2 above.

(e) The window size, H in (10), was chosen as 30 years (360 months) based on the dynamical properties of the original signals and the complexity of the chosen model structure. As discussed in the above section, the influence of window size on the correlation tracking has been significantly reduced in the proposed

method due to the employment of the Blackman window function.

(f) The sliding step of the window was chosen as 12 months, which is reasonable small comparing with the window size of 30 years.

(g) A boundary condition was considered in this example to ensure the same length of sampling data for each window. Therefore, zero values were added to the beginning and end of the input and output signals.

After applying the proposed method, the computed properties of the correlation are illustrated in Figure 8. The first graph plots the comparison between the observed annual I48N (solid) and the corresponding model output (dot). Note that although monthly data for I48N were sampled and modelled, the annual I48N is plotted in Figure 8 for easier inspection of fitting performance. The second graph plots the correlation strength, including both linear and nonlinear components, for each input. The significance threshold was also calculated and plotted. The linearity and nonlinearity of correlation are illustrated by the third graph. To understand the latency between the inputs and the output, a correlation map plotted in the time-latency domain for each input is shown in the fourth, fifth, and sixth graph, respectively.

Figure 8 shows clearly that the flux of icebergs reaching the Grand Banks (I48N) is a strongly nonlinear process, with significant temporal variation in the contribution of the main inputs to the signal. The direct atmospheric contribution to iceberg calving and motion, encapsulated within the NAO variable, has been below significant levels through much of the twentieth century; it has only been since global warming became a major environmental signal in the last 20 years [26] that this variable has been consistently significant, but with its WERR value still only being ~0.2. However, the main significant correlations leading to the dramatic variability of I48N are with SMB, a measure of the balance between melting and growth of the Greenland Ice Sheet, and LSST, a measure of the ocean temperature affecting tidewater glacier calving. These two variables have oscillated in importance, on a decadal scale, although it is noteworthy that the SMB has had the largest WERR for a large part of the total period. LSST's importance is strong particularly in the 1970s and 1980s, when the recent period of very large iceberg fluxes was
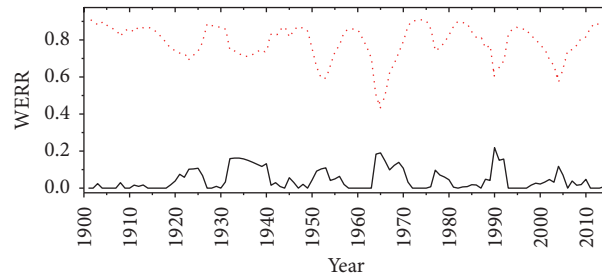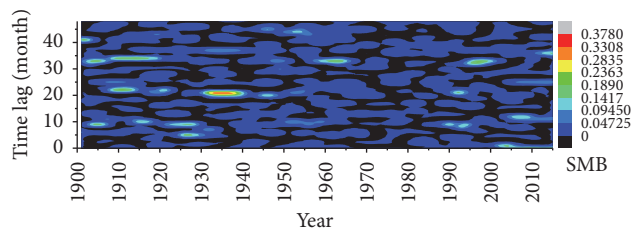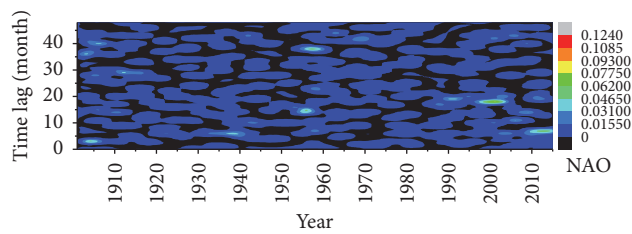
(a)

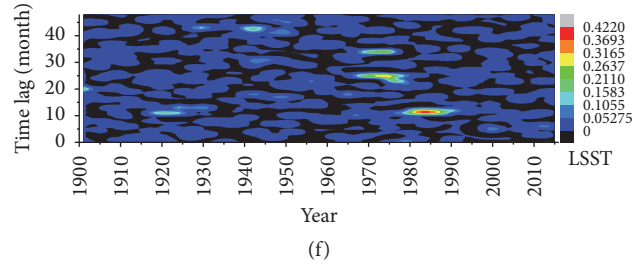(b)
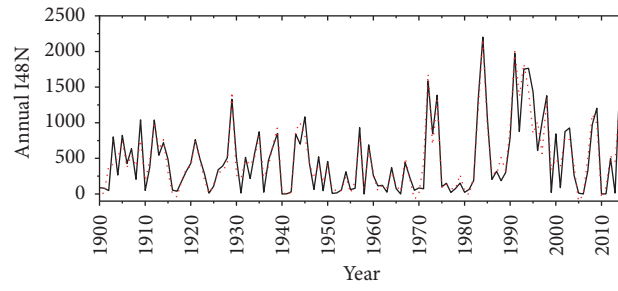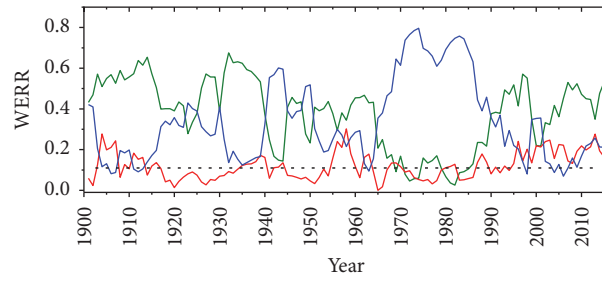
(c)

(d)

(e)

FIGURE 8: Continued.

(f)

FIGURE 8: Results produced by the proposed WERR approach to understand the variation of climatic and glaciological contributions to West Greenland iceberg discharge. (a–f) Comparison of the measured annual iceberg numbers and the model output; the WERR value for each input variable with the significance threshold; the linear and nonlinear contributions; the WERR maps against time and time lag for each input.

initiated. The time-latency correlations show a distinct peak around 10 months during this period for LSST, indicating that it was the ocean temperature of the previous summer that was driving the I48N variation. There is also a noticeable periodicity in the time-latency plots, consistent with signals from each of the past 3 years playing important roles in producing iceberg calving at different stages of the last 115 years.

5. Conclusions

This paper has introduced a Windowed Error Reduction Ratio (WERR) method to quantify the properties of correlation between measurements of complex systems, particularly for a system that has very little prior knowledge and the correlation is nonlinear, time-varying, rapid changed, or with short duration. Established upon the Nonlinear Finite Impulse Response (NFIR) model for a multiple inputs single output system, the introduced method quantifies the strength of the linear and nonlinear interdependencies between the inputs and the output in a percentage form. Other properties, such as significance level of correlation, latency between the input and the output, correlation type, and linearity and nonlinearity, are also computed to form a relatively comprehensive theory of correlation analysis. A sliding nonrectangular window function was used to track the time-varying correlation. Examples 1 and 2, employing prescribed simple functional relationships, clearly demonstrate the superior performance of the proposed method using the Blackman window function to detect the onset of interaction and short-duration interaction. A common problem in tracking technologies is the difficulty in selecting an appropriate window size that should balance the accuracy of correlation measurement and the response speed to the correlation changes. Although the window size still needs to be chosen due to the nature of the sliding window technique, it has been inferred from the simulation results that the window size in the new method has significantly less influence on correlation measurement performance than existing approaches. Such a characteristic is important because it will improve the applicability and robustness of correlation analysis for complex systems. For a slow time-varying system, it appears that the Welch window has a slightly better performance than the Blackman window.

The application of the WERR method to study the Greenland iceberg discharge problem has been presented and the results are encouraging. It has been found that the process leading to iceberg calving is distinctly nonlinear, with distinctly decadal signals in the dominant variable underlying this flux. These observations show that the proposed method has the potential to better understand complex systems by revealing and characterising some hidden information contained by measurements, which is usually "invisible" for conventional methods.

It should be noted that, similar to the cross prediction methods, the proposed method focuses on quantification of coupling strength between inputs and outputs, where the direction of causality is known a priori. Therefore, the research context is different with other studies which investigate the causality (cause-effect direction) between measures. One difference of the proposed methods with the cross prediction methods and Granger causality tests is that an unbiased full model is not required and predetermined and there is no prediction process involved. The model in the proposed method is to be built up term by term in a manner that exposes the significance of each new term that is added. This feature is especially important for a complex system where the underlying relationships are nonlinear and dynamic and the measured observations are noisy because, unless a complete and full model which accounts for any potentially nonlinear noise effects is estimated, the other methods' result could be compromised. One limitation of the proposed method is that the number of candidate terms could dramatically increase by increasing d and n (see (9)), which will lead to a high-computational cost. One potential solution is to detect the range of model order and time lag before applying this method.

Additional Points

Highlights. (i) Sensitivity of dependence estimation to the sample length is significantly reduced. (ii) Rapid and nonlinear dependence between measurements can be better revealed. (iii) Nonlinear dynamic systems with limited prior knowledge become more analyzable. (iv) An application to environmental data demonstrates the potential of the method.
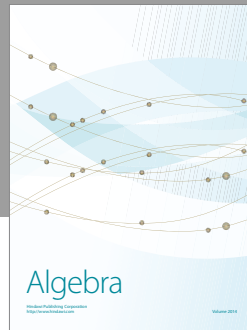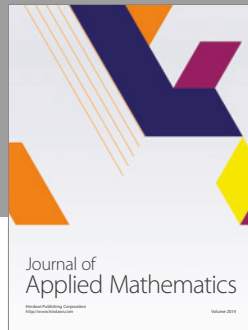
Conflicts of Interest

Acknowledgments

References

[1] J. A. S. Kelso, *Understanding Complex Systems*, Springer, Berlin, Germany, 2005.

[2] D. Aerts, B. D'Hooghe, and N. Note, "Worldviews, Science and Us, Singapore: Worl Scientific, 2005".

[3] Y. Goyal, D. Parikh, and D. Batra, "Towards transparent ai systems: interpreting visual question answering models," in *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, Las Vegas, USA, 2016.

[4] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston, "Dynamic causal modeling of evoked responses in EEG and MEG," *NeuroImage*, vol. 30, no. 4, pp. 1255–1272, 2006.

[5] S. G. Douma, X. Bombois, and P. M. Van den Hof, "Validity of the standard cross-correlation test for model structure validation," *Automatica*, vol. 44, no. 5, pp. 1285–1294, 2008.

[6] Z. Miao, M. R. James, and I. R. Petersen, "Coherent observers for linear quantum stochastic systems," *Automatica*, vol. 71, pp. 264–271, 2016.

[7] T. M. Lenton, H. Held, E. Kriegler et al., "Tipping elements in the Earth's climate system," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 105, no. 6, pp. 1786–1793, 2008.

[8] Millennium-Ecosystem-Assessment, *Ecosystems and Human Well-being*, Island Press, Washington, DC., USA, 2005.

[9] P. G. Sarrigiannis, Y. Zhao, H.-L. Wei, S. A. Billings, J. Fotheringham, and M. Hadjivassiliou, "Quantitative EEG analysis using error reduction ratio-causality test; validation on simulated and real EEG data," *Clinical Neurophysiology*, vol. 125, no. 1, pp. 32–46, 2014.

[10] H. Park, *Nonlinearity Detection for Condition Monitoring Utilizing Higher-order Spectral Analysis Diagnostics*, University of Texas, Austin, Tex, USA, 2008.

[11] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–238, 1969.

[12] L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 5, Article ID 051112, 2011.

[13] S. Chen and S. A. Billings, "Representations of nonlinear systems: the NARMAX model," *International Journal of Control*, vol. 49, no. 3, pp. 1013–1032, 1989.

[14] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*, John Wiley & Sons, London, UK, 2013.

[15] S. A. Billings and H.-L. Wei, "Sparse model identification using a forward orthogonal regression algorithm aided by mutual information," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 81, no. 5, pp. 306–310, 2006.

[16] Y. Zhao, H. L. Wei, and S. A. Billings, "A new adaptive fast cellular automaton neighborhood detection and rule identification algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 4, pp. 1283–1287, 2012.

[17] Y. Zhao, S. A. Billings, H. Wei, F. He, and P. G. Sarrigiannis, "A new NARX-based Granger linear and nonlinear casual influence detection method with applications to EEG data," *Journal of Neuroscience Methods*, vol. 212, no. 1, pp. 79–86, 2013.

[18] T. Hesterberg, "Bootstrap," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, no. 6, pp. 497–526, 2011.

[19] W. Max-Moerbeck, J. L. Richards, T. Hovatta, V. Pavlidou, T. J. Pearson, and A. C. S. Readhead, "A method for the estimation of the significance of cross-correlations in unevenly sampled red-noise time series," *Monthly Notices of the Royal Astronomical Society*, vol. 445, no. 1, pp. 437–459, 2014.

[20] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77–94, 1992.

[21] M. McMillan, A. Leeson, A. Shepherd et al., "A high-resolution record of Greenland mass balance," *Geophysical Research Letters*, vol. 43, no. 13, pp. 7002–7010, 2016.

[22] M. R. Van Den Broeke, E. M. Enderlin, I. M. Howat et al., "On the recent contribution of the Greenland ice sheet to sea level change," *The Cryosphere*, vol. 10, no. 5, pp. 1933–1946, 2016.

[23] I. M. Howat and A. Eddy, "Multi-decadal retreat of Greenland's marine-terminating glaciers," *Journal of Glaciology*, vol. 57, no. 203, pp. 389–396, 2011.

[24] G. R. Bigg, H. L. Wei, D. J. Wilton et al., "A century of variation in the dependence of Greenland iceberg calving on ice sheet surface mass balance and regional climate change," *Proceedings of the Royal Society a Mathematical, Physical and Engineering Sciences*, vol. 470, no. 2166, Article ID 20130662, 2014.

[25] Y. Zhao, G. R. Bigg, S. A. Billings et al., "Inferring the variation of climatic and glaciological contributions to West Greenland iceberg discharge in the twentieth century," *Cold Regions Science and Technology*, vol. 121, pp. 167–178, 2016.

[26] D. Hartmann, A. Klein Tank, and H. Rusticucci, "Observations: atmosphere and surface," in *Climate Change 2013: The Physical Science Basis*, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge, UK, 2013.

[27] I. Janssens and P. Huybrechts, "The treatment of meltwater retention in mass-balance parameterization of the Greenland ice sheet," *Annals of Glaciology*, vol. 31, pp. 133–140, 2000.