

Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) – A Method for Quantifying Correlation between Multivariate Time-Series

Sebastian Wallot


To cite this article: Sebastian Wallot (2018): Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) – A Method for Quantifying Correlation between Multivariate Time-Series, Multivariate Behavioral Research, DOI: [10.1080/00273171.2018.1512846](https://doi.org/10.1080/00273171.2018.1512846)

To link to this article: <https://doi.org/10.1080/00273171.2018.1512846>



© 2018 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#) 



Published online: 20 Dec 2018.



[Submit your article to this journal](#) 



Article views: 126



[View Crossmark data](#) 

Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) – A Method for Quantifying Correlation between Multivariate Time-Series

Sebastian Wallot

Department of Language and Literature, Max Planck Institute for Empirical Aesthetics

ABSTRACT

In this paper, Multidimensional Cross-Recurrence Quantification Analysis (*MdCRQA*) is introduced. It is an extension of Multidimensional Recurrence Quantification Analysis (*MdRQA*), which allows to quantify the (auto-)recurrence properties of a single multidimensional time-series. *MdCRQA* extends *MdRQA* to bi-variate cases to allow for the quantification of the co-evolution of two multidimensional time-series. Moreover, it is shown how a Diagonal Cross-Recurrence Profile (*DCRP*) can be computed from the *MdCRQA* output that allows to capture time-lagged coupling between two multidimensional time-series. The core concepts of these analyses are described, as well as practical aspects of their application. In the supplementary materials to this paper, implementations of *MdCRQA* and the *DCRP* as MatLab- and R-functions are provided.

KEYWORDS

Multidimensional
Cross-Recurrence
Quantification Analysis;
multivariate time-series;
MdCRQA; *DCRP*; R; MatLab



Introduction


Cross-Recurrence Quantification Analysis (*CRQA*) is a nonlinear correlational analysis technique that has become increasingly prominent in psychological research. What makes it attractive is, that it can be used to analyze nominal and interval-scale data (Dale, Warlaumont, & Richardson, 2011), that it makes very few assumptions (i.e., no assumptions about linearity or distribution type), and is very robust against outliers (Fusaroli, Konvalinka, & Wallot, 2014; Marwan, Romano, Thiel, & Kurths, 2007; Shockley, 2005). Additionally, *CRQA* results in multiple output variables that allow a detailed description of the co-evolution of two signals (Marwan et al., 2007; Webber & Zbilut, 2005). Accordingly, *CRQA* has become a prominent tool to assess correlation – or coupling – in research settings that investigate variables with complex dynamics.

Particularly, applications of *CRQA* have been very prominent in joint action research, investigating coupling of behavioral and physiological measures between interacting participants over longer stretches of time or in complex task setups. For example, *CRQA* has been used to investigate behavioral (Abney, Paxton, Dale, & Kello, 2015; Fusaroli & Tylén, 2016)

and physiological markers (Mønster, Håkonsson, Eskildsen, & Wallot, 2016) during group action, joint physiological arousal during fire ritual performance (Konvalinka et al., 2011), facial movements and gesture (Louwerse, Dale, Bard, & Jeuniaux, 2012), eye movements (Dale, Kirkham, & Richardson, 2011; Richardson & Dale, 2005; Warlaumont et al., 2010), and postural sway during conversation and interpersonal coordination (Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Santana, & Fowler, 2003), speech properties of infants and care givers (Abney, Warlaumont, Oller, Wallot, & Kello, 2017), interlocutors during conversation (Coco, Dale, & Keller, 2018; Fusaroli & Tylén, 2016), or joint gaze behavior of mothers and infants (Nomikou, Leonardi, Rohlfing, & Rączaszek-Leonardi, 2016).

However, some behaviors are inherently multivariate – such as position measurements of eye movements, hand and arm movements, or facial muscle movements, for example. All of these have been investigated using *CRQA*, but conventional Cross-Recurrence Quantification Analysis can only be used to compute coupling of one-dimensional time-series. Thus far, this limited the analysis of inherently multivariate time-series to the analysis of coupling of their individual components.

CONTACT Sebastian Wallot  sebastian.wallot@ae.mpg.de  Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany.

 Supplemental data for this article can be accessed on the publisher's website.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hmbr.

© 2018 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

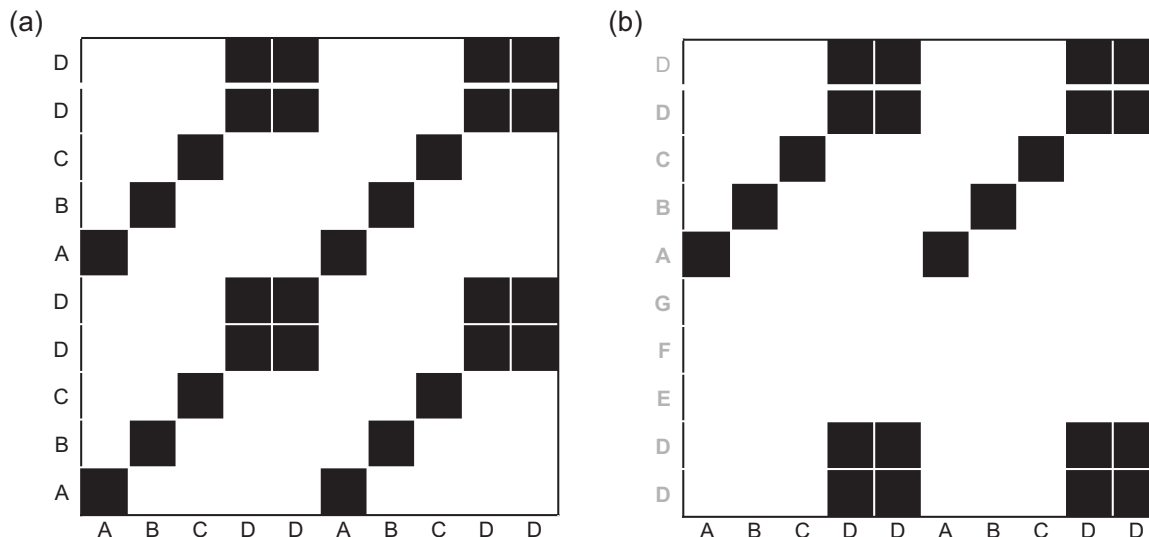


Figure 1. Illustration of recurrence of letters in the sequence “ABCDDABCDD” (a), and cross-recurrence of letters in the sequence “ABCDDABCDD” with “DDEFGABCDD” (b). The black dots in the matrices indicate the recurrence of a letter, and white spaces indicate the absence of recurrence. The distribution of recurrent points on the recurrence plot can be quantified to yield statistics of the repetitive patterns in a sequence.

For example, Louwse et al. (2012) measured 15 features of hand and face movements (such as laughing, tightening of lips, raising eye brows, etc.) to investigate coupling of facial movements during conversation, but due to the constraints of *CRQA*, analysis needed to be restricted to coupling of individual features, and computation of whole-face-movement coupling across multiple features was not possible.

In this paper, I present an extension of *CRQA* – Multidimensional Cross-Recurrence Quantification Analysis (*MdCRQA*) – that generalized *CRQA* to the application of multivariate time-series. In effect, *MdCRQA* is the cross-recurrence version of Multidimensional Recurrence Quantification Analysis (*MdRQA* – Wallot, Roepstorff, & Mønster, 2016). Particularly, the strength of the method is to provide a unified quantification for two time-series with multiple, potentially interacting dimensions. Moreover, it allows to properly assess follow-leader dynamics between two multidimensional time-series particularly in contexts where changes in one dimension of the leading time-series are responded to by changes in another dimension of the following time-series, or where changes are distributed over multiple dimensions simultaneously in both time-series.

In the following sections, the concepts of recurrence and cross-recurrence are introduced, which are at the heart of the analysis technique. Then, estimation of parameters in recurrence analyses is briefly introduced. Finally, *MdCRQA* and its application to model systems and empirical data are described. In the Appendix to this paper, Matlab and R functions implementing *MdCRQA* are provided. These

languages were chosen, because two well-developed toolboxes exist to conduct *CRQA* – the “*crqa*” package in *R* (Coco & Dale, 2014) and the “*CRP-Toolbox*” in Matlab (Marwan, 2017), and are supposedly the most widely used languages for these applications.

Recurrence and cross-recurrence

As the name implies, the core-concept of recurrence-based analyses is *recurrence* – repetition of elements or patterns in a sequence. The core tool of these analyses is the recurrence plot or recurrence matrix, which is a means of displaying and charting repetitions in a sequence. As we will see further in the following sections, the analyses are usually not performed on the original one-dimensional sequence or time-series, but on its phase-space portrait. However, how the recurrence plot (*RP*) captures recurrences in a sequence can be easily shown using a simple one-dimensional nominal sequence, “ABCDDABCDD” (Figure 1a).

Similarly, we can examine cross-recurrences between two sequences with each other, as in Figure 1b. While the *RP* in Figure 1a possesses a diagonal of recurrent points (which simply means that every element in the sequence is recurrent with itself at lag 0), the cross-recurrence plot (*CRP*) in Figure 1b does not possess such a diagonal. Moreover, while the *RP* is fully symmetrical about its main diagonal, this is not the case for the *CRP*.

The *CRP* is not just a useful tool to display the sequential properties of a series, but can be used to quantify these properties. For example, counting the recurrent points (i.e., the black squares in Figure 1b)

Table 1. The most common CRQA measures.

Variable name	Definition	Quantifies ...
Percent recurrence (<i>REC</i>)	Sum of recurrent points in <i>CRP</i> / Size of <i>CRP</i>	Repetitiveness of the elements between sequences.
Percent determinism (<i>DET</i>)	Sum of diagonally adjacent recurrent points / Sum of recurrent points in <i>CRP</i>	How many of cross-repetitions occur in connected trajectories.
Average diagonal line length (<i>ADL</i>)	Average diagonal lines in <i>CRP</i>	How long the average cross-repeating trajectory is.
Maximum diagonal line length (<i>MDL</i>)	Length of longest diagonal line in <i>CRP</i>	How long the longest cross-repeating trajectory is.

Further RQA measures exist, and others are being developed – for the description of additional measure, see for example Marwan et al. (2007).

in a *CRP* tells us something about how many individual elements between two sequences are shared, and we refer to this quantity as percent recurrence (*REC*). Counting all recurrent points that have diagonally adjacent recurrent points and dividing them by *REC* tells us something about the degree to which elements between the two sequences repeat in terms of larger, connected patterns, which is referred to as percent determinism (*DET*). Counting the average length of all diagonally adjacent lines of recurrent points tells us something about the average size of the repeated patterns (Average Diagonal Line, *ADL*).

There are many ways to quantify recurrence structure on a *CRP*, and Table 1 charts only the most common measures and their definition. Concretely, calculating cross-recurrence measures from the *CRP* in Figure 1b is done as follows: We observe 22 instances of recurrence (i.e., where a letter in the sequence on the *x*-axis is repeated in the sequence on the *y*-axis) out of 100 possible (i.e., the size of the *CRP* matrix). Hence, $REC = 22/100 = 0.22 = 22\%$, indicating that individual values cross-recur in 22%. For *DET*, we find that 14 out of our 22 recurrences form diagonally adjacent recurrence patterns (i.e., two diagonal lines of length 5, and two diagonal lines of length 2). Hence, $DET = 14/22 = 0.64 = 63.6\%$. This tells us, that a disproportional amount of the individual cross-recurrences between the two series are actually shared in terms of longer trajectories of values. Finally, *ADL* and *MDL* provide further information on the length of these trajectories, where average length is $ADL = (2 + 2 + 5 + 5)/4 = 3.5$, and maximum length is $MDL = 5$. In general, the higher the values on each of these measures, the stronger the coupling between the two time-series.

In the following section, we will apply the basic concept of recurrence and the *CRP* to multidimensional time-series, where the correlation of the common dynamics of two (or more) dimensional time-series are of interest. Before that, however, we will introduce the basic steps of parameter estimation that are necessary to conduct recurrence-based analyses, using simple CRQA as an example of how to reconstruct the multidimensional phase-space of a one-dimensional time-series, on which the *CRP* is computed.

Parameter estimation

For continuous data, particularly for time-series with strong deterministic dynamics, *CRPs* are usually not computed based on the values of the (one-dimensional) time-series *per se* (as the example in Figure 1), but on the so-called phase-space portrait of the time-series. Calculation of the phase-space portrait is performed to mitigate distortions in the recurrence patterns resulting from projection of the data from a potentially high-dimensional system onto the one-dimensional measurement – that is, the autocorrelation structure of a measured time-series can potentially hold information about other latent variables that determine the dynamics of the time-series (Takens, 1981).

One-way reconstruction of a multidimensional phase-space from a one-dimensional time-series is the method of time-delayed embedding (Packard, Crutchfield, Farmer, & Shaw, 1980). If the dynamics of latent variables that co-determine the dynamics of an observed one-dimensional time-series are coupled, then one can reconstruct the dynamics of these latent dimensions from the observed one-dimensional time-series by effectively plotting the values of that time-series (multiple times) against itself at a certain lag. The resulting coordinates in higher dimensional phase-space approximate the phase-space of the actual multidimensional system from which the original time-series was recorded.

In order to apply the method of time-delayed embedding, two parameters are needed: the delay parameter τ , which is the lag at which the time-series is plotted against itself, and the embedding dimension parameter D , where $D - 1$ is the number of times that the time-series is plotted against itself. If these two parameters are known, the original phase-space from which one-dimensional data were sampled can be approximately reconstructed. However, these parameters are usually unknown for empirical time-series data and have to be estimated.

Two standard methods to estimate these parameters in one-dimensional time-series are the computation of the Average Mutual Information (*AMI*) function and the False Nearest Neighbor (*FNN*) function, where the first local minima of those functions

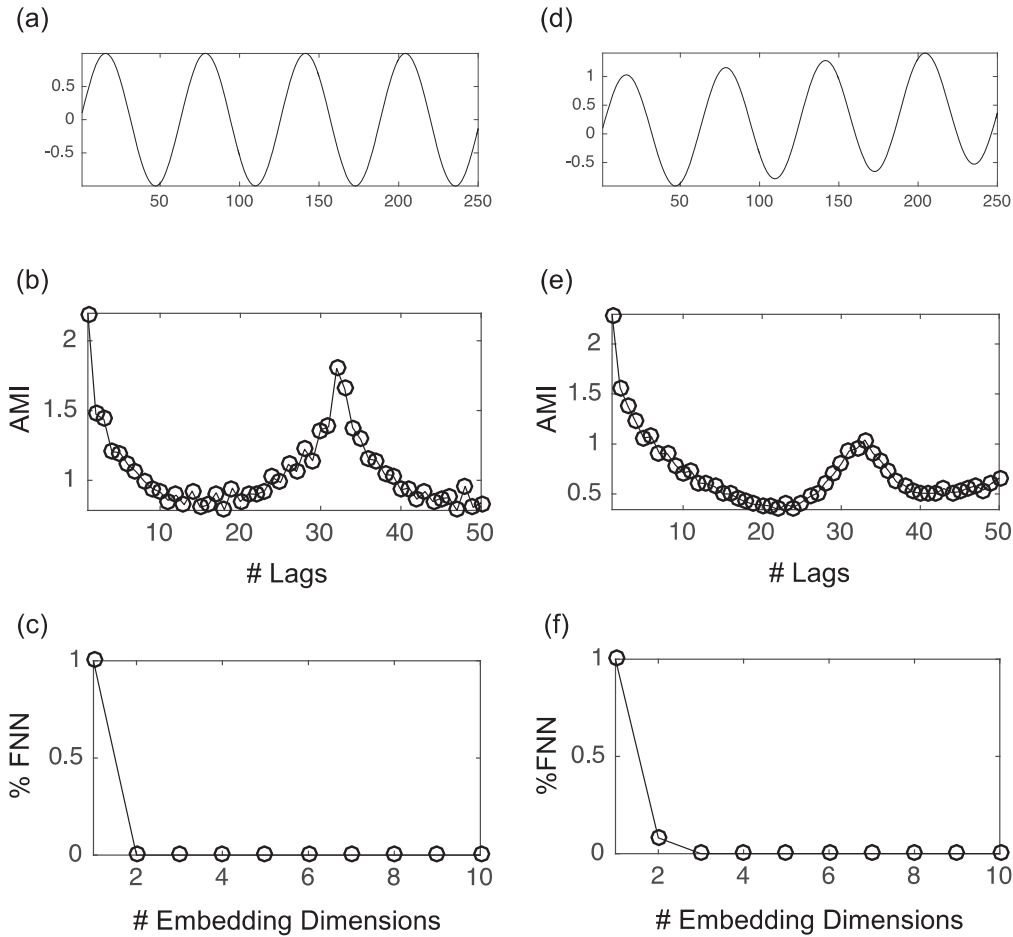


Figure 2. Illustration of the parameter estimation process with Average Mutual Information (AMI) and False Nearest-Neighbors (FNN) using a sine wave (a) and a sine wave with a linear trend (b) as an example. The data of the sine wave (a) is first subjected the AMI-function, examining the change in AMI for the first 50 lags. The first local minimum of the function is around 15, hence $\tau = 15$ would be the estimate for the delay parameter for the first time-series (a). Then, the data is subjected to the FNN-function using $\tau = 15$ and examining the first 10 embedding dimensions. Here, the first local minimum – or more precisely, the point at which the function becomes stable is clearly 2. Hence, $D = 2$ would be the estimate for the embedding dimension parameter. Thereafter, these steps are separately taken for the second time-series (d). Here, the AMI- and FNN-functions suggest parameter estimates of $\tau = 22$ and $D = 2$ or 3. Hence, to subject these time-series to CRQA, the parameter combination of $\tau = 18$ and $D = 2$ (average) or $\tau = 22$ and $D = 3$ (maximum) seem reasonable choices.

(or the point at which those functions level-off) are indicative of the delay and embedding dimension (Abarbanel, 1996; Kennel et al., 1992).

To estimate these parameters, the following procedure has been advocated (Marwan et al., 2007): first, in order to estimate the delay parameter τ , the AMI-function for each of the two time-series is computed over a range of lags. The lag of the first local minimum of that function is used as an estimate for the delay parameter τ for each time-series. Because one has to choose a single value τ for both time-series to conduct CRQA, the usual approach is to take the average lag at the first local minima of the two time-series (rounded to the next integer value) or the bigger lag of the two, because higher parameter values usually do not substantially impact the results.

Then, the embedding dimension D is estimated by subjecting the two time-series – again separately – to the FNN-function, and investigate their FNN-functions over a range of embeddings. Similarly, the estimation of the delay parameter τ , one uses the first local minimum – or the first point at which the FNN-function becomes stable – as an estimate for D . Again, because one can only set a single value for D in CRQA, one has to take the average (or higher) value of the two estimates. Figure 2 graphically illustrates the process of parameter estimation using a sine wave and a trending sine wave as examples.

For non-categorical data, one also has to estimate a threshold parameter r that is the size of the neighborhood in phase-space, which effectively provides a tolerance range within with similar, but not identical,

coordinates in phase-space are counted as recurrent. This is necessary, because empirical data – due to endogenous fluctuations or measurement error – are never perfectly repetitive. The higher the value for r , the higher the percentage of recurrence points (REC – see Table 1). Hence, r must be selected to yield lower than 100%, but more than 0% REC , because otherwise the recurrence measures cannot be sensibly computed. As a rule of thumb, Webber and Zbilut (2005) recommend an average REC of $\approx 1\%$ for well-sampled physiological data with a high signal-to-noise ratio, and $REC = 2\text{--}5\%$ for behavioral data, though certain kinds of data (i.e., inter-event-times such as time-series of reaction times) might warrant $REC \approx 20\%$ (Wallot, O'Brien, and Van Orden, 2012).

There are also algorithmic methods for choosing “optimal” parameters (Coco & Dale, 2014), but the author advises users to also inspect the AMI - and FNN -functions as briefly described above in order to check for the sensibility of such solutions.

Figure 3 provides a schematic overview over the embedding procedure and the application of the threshold parameter r to transform a phase-space into a CRP . It is beyond the scope of the current paper to discuss all potential pitfalls and special cases of the parameter estimation procedure. However, the parameter estimation procedures and best practice have been described elsewhere in detail, and readers are referred to practical introductions for existing software packages that provide estimation methods using R (Wallot, 2017; Wallot & Leonardi, 2018) or Matlab (Wallot & Grabowski, 2018).

Before we proceed to the introduction of Multidimensional Cross-Recurrence Quantification Analysis ($MdCRQA$), one last note is in order about parameter selection for time-series with multiple dimensions, as we will be dealing with using $MdCRQA$. If we are interested in analyzing the cross-recurrence properties of two time-series with more than one dimension each, such a set of measured dimensions could boast yet-higher dimensional dynamics than the two (or more) measured dimensions at hand. Then, it would still be necessary to infer the appropriate dimensionality and reconstruct the phase-space by the method of time-delayed embedding (Takens, 1981). Here, one can start by assessing the delay and embedding parameters from the individual dimensions of the time-series that are eventually fed to $MdCRQA$ using the methods of parameter estimation for one-dimensional time-series and apply them to each constituent dimension of each time-series.

That is, before running $MdCRQA$ between two time-series, one can test each time-series' individual dimension's embedding parameters. If the average estimated dimensionality D – using the FNN method – of the individual dimensions for each time-series is estimated to be equal or lower to the dimensionality d of the multidimensional time-series, then potentially no further embedding is necessary. If the average estimated dimensionality D is substantially higher than d , one could embed the multidimensional time-series before computing the $MdCRP$ (Wallot, Roepstorff, & Mønster, 2016).

For example, if one wants to cross-recur two three-dimensional time-series (i.e., $MdCRQA3$), then the dimensionality of these time-series is already $d = 3$. If the average dimensionality of the individual dimensions of each time-series is estimated to be close to $D = 3$ as well, no embedding might be necessary. If, D is estimated to be substantially bigger, for example $D = 6$, then one could embed the three-dimensional time-series once to achieve an overall phase-space dimensionality of 6 (i.e., $d * D/d = D$; here: $3 * 6/3 = 6$).

Note, however, that the estimates for τ and D when based on the average of the individual dimensions of the multidimensional time-series might not be the same as estimates of τ and D that are based on the multidimensional time-series. However, a first solution to the problem of parameter estimation based on multidimensional time-series has recently been provided by Wallot, & Mønster (2018).

Multidimensional Cross-Recurrence Quantification Analysis ($MdCRQA$)

Multidimensional Cross-Recurrence Quantification Analysis ($MdCRQA$) is an extension of $CRQA$, which can be used for one-dimensional time-series, to time-series with arbitrary dimensionality. It yields recurrence measures that capture the strength and type of coupling – or correlation – between two multidimensional time-series, and can also be used to quantify time-delayed coupling between these time-series using the Diagonal Cross-Recurrence Profile ($DCRP$). In the next sections, we will describe how a CRP is computed for the case of one-dimensional time-series, and then show how this can be extended to multidimensional time-series. Then, we will describe how the Diagonal Cross-Recurrence Profile ($DCRP$) can be computed for multidimensional time-series, as well as issues of parameter estimation for the multidimensional case.

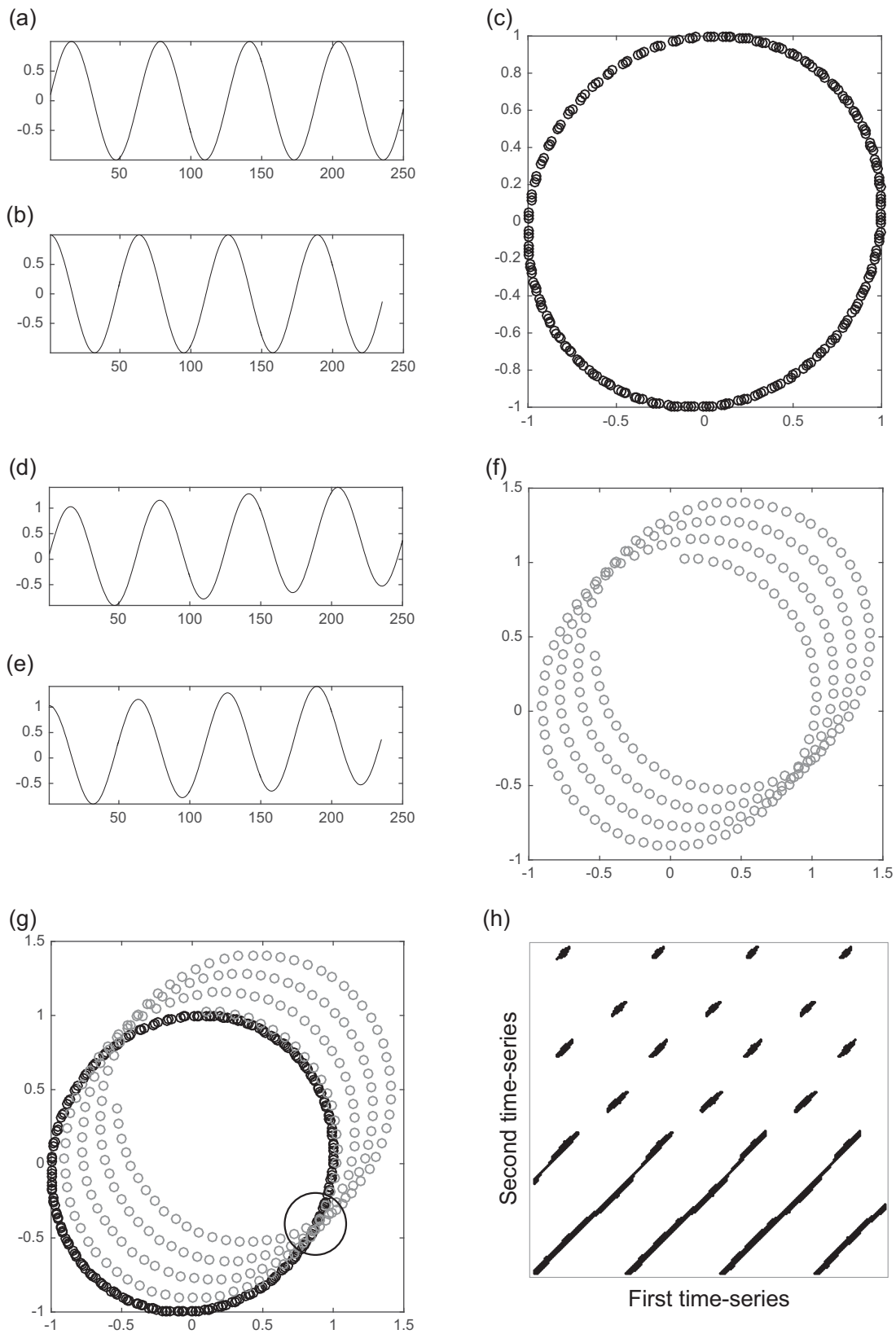


Figure 3. Schematic of the application of the embedding parameters and the extraction of a *CRP*. The delay parameter $\tau = 18$ is applied to the data of the first original time-series (a) to yield its time-delayed surrogate series (b). When the first original time-series is plotted against its surrogate, this yields the reconstructed phase-space profile (c). Similarly, the delay parameter $\tau = 18$ is applied to the data of the second original time-series (d) to yield its time-delayed surrogate series (e). Again, when the original and surrogate series are plotted against each other, this results in their phase-space portrait (f). Because we only applied the delay parameter τ once to create a single surrogate series for each original time-series, the dimensionality of the phase-space is $D = 2$

Computation of Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA)

MdCRQA is an extension of CRQA to assess coupling between multidimensional time-series. A CRP is computed by comparing all possible combinations of the values of two time-series $x = (x_1, x_2, x_3, \dots, x_m)$ and $y = (y_1, y_2, y_3, \dots, y_m)$, or the values of the coordinates of the reconstructed phase-space profiles from those two time-series V (Equation (1)) and W (Equation (2)), given some value for the delay parameter τ and the embedding dimension D , which must be the same for both time-series:

$$V = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_m \end{pmatrix} = \begin{pmatrix} x_1 & x_{1+\tau} & x_{1+2\tau} & \dots & x_{1+(D-1)\tau} \\ x_2 & x_{2+\tau} & x_{2+2\tau} & \dots & x_{2+(D-1)\tau} \\ x_3 & x_{3+\tau} & x_{3+2\tau} & \dots & x_{3+(D-1)\tau} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{m-(D-1)\tau} & x_{m-(D-2)\tau} & x_{m-(D-3)\tau} & \dots & x_m \end{pmatrix} \quad (1)$$

and

$$W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ W_m \end{pmatrix} = \begin{pmatrix} y_1 & y_{1+\tau} & y_{1+2\tau} & \dots & y_{1+(D-1)\tau} \\ y_2 & y_{2+\tau} & y_{2+2\tau} & \dots & y_{2+(D-1)\tau} \\ y_3 & y_{3+\tau} & y_{3+2\tau} & \dots & y_{3+(D-1)\tau} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{m-(D-1)\tau} & y_{m-(D-2)\tau} & y_{m-(D-3)\tau} & \dots & y_m \end{pmatrix} \quad (2)$$

Now, a cross-recurrence (CR – see Equation (3)) can be defined by applying a threshold r to the distances between all possible pairs of values for the time-series x and y , or their reconstructed phase-space profiles V and W , where distances $\leq r$ are counted as recurrent instances, and distances $> r$ are counted as non-recurrent instances:

$$CR_{ij}^{X,Y}(r) = \Theta(r - \|X_i - Y_j\|); \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (3)$$

Here, Θ is the Heaviside step function and r is some arbitrary threshold distance. Multidimensional cross-recurrence plots extend this by starting with two multidimensional time-series V (Equation (4)) and W (Equation (5)), which have a dimensionality d of $d > 1$:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_m \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,d} \\ x_{3,1} & x_{3,2} & x_{3,3} & \dots & x_{3,d} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{m,1} & x_{m,2} & x_{m,3} & \dots & x_{m,d} \end{pmatrix} \quad (4)$$

and

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \dots & y_{1,d} \\ y_{2,1} & y_{2,2} & y_{2,3} & \dots & y_{2,d} \\ y_{3,1} & y_{3,2} & y_{3,3} & \dots & y_{3,d} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{m,1} & y_{m,2} & y_{m,3} & \dots & y_{m,d} \end{pmatrix} \quad (5)$$

The analysis can either be performed on the unembedded time-series (Equations (4) and (5)), or they can be further embedded by the method of time-delayed embedding as described above – again, using a delay parameter τ and an embedding dimension D to re-construct phase-space portraits of coordinates of the multi-dimensional time-series X and Y (analogously to Equations (1) and (2)):

as a result of embedding the original time-series $D-1$ times. Then, the coordinates of the two phase-space profiles are plotted in the same phase-space (g) and a threshold parameter of $r=0.25$ (of the maximum distance between coordinates in the phase-space) is applied to identify cross-recurrent coordinates between the two profiles (gray circle with black border in the lower-right). Charting cross-recurrences for all possible coordinate pairs results in the cross-recurrence plot of the two time-series (h).

$$V = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_m \end{pmatrix} = \begin{pmatrix} X_1 & X_{1+\tau} & X_{1+2\tau} & \dots & X_{1+(D-1)\tau} \\ X_2 & X_{2+\tau} & X_{2+2\tau} & \dots & X_{2+(D-1)\tau} \\ X_3 & X_{3+\tau} & X_{3+2\tau} & \dots & X_{3+(D-1)\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{m-(D-1)\tau} & X_{m-(D-2)\tau} & X_{m-(D-3)\tau} & \dots & X_m \end{pmatrix} \quad (6)$$

and

$$W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ W_m \end{pmatrix} = \begin{pmatrix} Y_1 & Y_{1+\tau} & Y_{1+2\tau} & \dots & Y_{1+(D-1)\tau} \\ Y_2 & Y_{2+\tau} & Y_{2+2\tau} & \dots & Y_{2+(D-1)\tau} \\ Y_3 & Y_{3+\tau} & Y_{3+2\tau} & \dots & Y_{3+(D-1)\tau} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{m-(D-1)\tau} & Y_{m-(D-2)\tau} & Y_{m-(D-3)\tau} & \dots & Y_m \end{pmatrix} \quad (7)$$

Now, to define multidimensional cross-recurrences *MdCRs* (Equation (8)) of a multidimensional *MdCRP*, one simply substitutes phase-space portraits of the one-dimensional time-series, V and W , in Equation (5) with the multidimensional time-series X and Y (or their embedded phase-space portraits V and W):

$$MdCR_{i,j}^{X,Y}(r) = \Theta(r - \|X_i - Y_j\|); \quad i = 1, \dots, n; \\ j = 1, \dots, m \quad (8)$$

The resulting *MdCRP* can now be quantified the same way as any other recurrence or cross-recurrence plot, for example, using the measures presented in Table 1 [for additional measures that can be computed from (cross-)recurrence plots, see for example, Marwan et al. (2007)]. These measures reflect the overall coupling dynamics of the underlying (multivariate) time-series (Shockley, Butwill, Zbilut, & Webber, 2002). However, *CRPs* can also be used to quantify the direction of coupling, such as leader-follower relationships, by examining the diagonal cross-recurrence features (Dale et al., 2011). This is of course also possible for *MdCRPs*, as will be shown in the next section.

Computation of the Diagonal Cross-Recurrence Profile (DCRP)

Cross-recurrence plots cannot only be used to assess overall coupling between two time-series, but they can also be used to quantify leader-follower relationships, that is lagged cross-recurrences between two time-series. Dale and colleagues (2011) showed that *CRQA* generalizes lagged sequential analysis by quantifying the amount of cross-recurrences at and around the central diagonal of a *CRP*. For example, let us take *CRPs* of the stationary sine-wave and the sine-wave with a linear trend from Figure 3. If we sum-up all cross-recurrence points around ± 50 diagonals of

the central diagonal and divide them by the length of the diagonal, then we get a measure of REC_t for each of these 101 diagonals (i.e., from -50 to 50 , see Figure 4a). Note that 50 is an arbitrary time-window size – window size should be selected to cover the range of lags that one is interested in analyzing. For the plot from Figure 3h, this shows that most recurrences fall at and around the central diagonal, indicating that both time-series are mostly correlated at lag 0. This is called the Diagonal Cross-Recurrence Profile (*DCRP*).

If we shift the phase of the stationary time-series a little, and compute the *CRP* and REC_t for the ± 50 diagonals around the central diagonal again, then we see that the peak in REC_t now shifts towards the diagonals $j = -6$, where j is an index of the diagonals running from -50 to $+50$ in our example. This then indicates that the dynamics of the first time-series follows (i.e., recurs with) the dynamics of the second time-series with an average distance of six lags (Figure 4b). Hence, by picking the number of diagonals to compute the *DCRP*, we define a window of lags around lag 0 (i.e., the central diagonal) that we examine for potentially time-delayed cross-recurrences.

To compute the *DCRP*, one needs a *CRP*, which is a square $l \times l$ matrix, where l is the length of underlying (embedded) time-series, $l = m \cdot (D-1) \cdot \tau = n \cdot (D-1) \cdot \tau$. Then, one needs to choose window size w of lags that should be investigated. Here, w indicates the number of lags – that is, diagonals of the *CRP* around the central diagonal, for which time-lagged recurrence REC_t should be computed. The computation is the summation of all elements e for each diagonal j around the central diagonal, where $j = -w, -w+1, -w+2, \dots, w$. Then, the sum of elements for each diagonal needs to be divided by the length of the respective diagonal, which is $l - |j|$ to get the percentage of recurrences in that diagonal (Equation (7)):

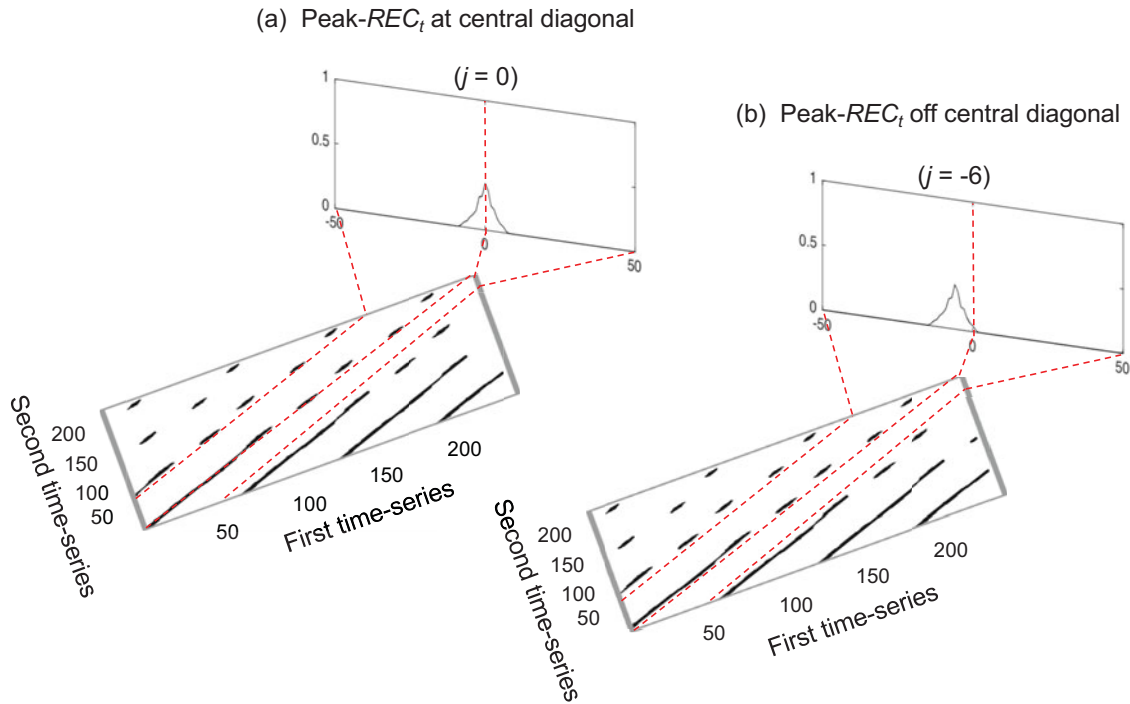


Figure 4. Illustration of the computation of the *DCRP* and leader-follower relationships. For the two sine waves with the same phase and frequency, peak- REC_t is at $j=0$, indicating synchronous dynamics at lag 0 (a). When the phase of the first time-series is shifted, peak- REC_t moves away from the central diagonal to $j=-6$, indicating that the dynamics of first time-series follow the dynamics of the second time-series with lag 6.

$$REC_t(j) = \begin{cases} \frac{\sum_{i=1}^{l-|j|} e_{(i+|j|),i}}{l-|w|} & \text{if } j \leq 0 \\ \frac{\sum_{i=1}^{l-|j|} e_{i,(i+|j|)}}{l-|w|} & \text{otherwise} \end{cases} ; j = -w, -w+1, \dots, w \quad (7)$$

While REC_t for negative values of j indicates that time-series y (or Y) follows x (or X), REC_t for positive values of j indicates that the time-series x (or X) follows y (or Y) – see Figure 4(b). High values of REC_t at $j=0$ (i.e., the central diagonal) indicate synchronous behavior of the two time-series at lag 0.

In the current example, a window size, w , of ± 50 around the central diagonal (i.e., lag 0) has been chosen. This choice is arbitrary. For empirical time-series, the choice of the w depends on the time-window of interest, within which one want to investigate leader-follower relationships. However, it is always advisable to initially investigate a larger number of w in order to get an overview over lags that might yield time-lagged recurrences, if there are no particular a-priori hypothesis which range of lags ought to be theoretically significant.

To assess the statistical significance of REC_t values (i.e., the correlation at some lag), several methods

have been proposed, such as comparing the *DCRP* of a pair of time-series to REC (i.e., the average amount of recurrences across the whole plot), to a *DCRP* of the shuffled time-series, or to false-pair surrogates (Richardson & Dale, 2005; Wallot & Leonardi, 2018).

Example 1: Using *MdCRQA* to capture common dynamics of multidimensional dynamic systems

In the following, we will apply *MdCRQA* to assess similarities in the dynamics between two multidimensional dynamic systems – the Lorenz system (Lorenz, 1963) and the Rössler system (Rössler, 1976). Both are systems of three coupled differential equations that each result in three-dimensional dynamics, which for the Lorenz-system (Equation (8)) are defined by:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= (\rho - z)y \\ \dot{z} &= xy - \beta z \end{aligned} \quad (8)$$

and for the Rössler system (Equation (9)) by:

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c) \end{aligned} \quad (9)$$

For the Lorenz system, we will use the common parameter settings $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$, and for

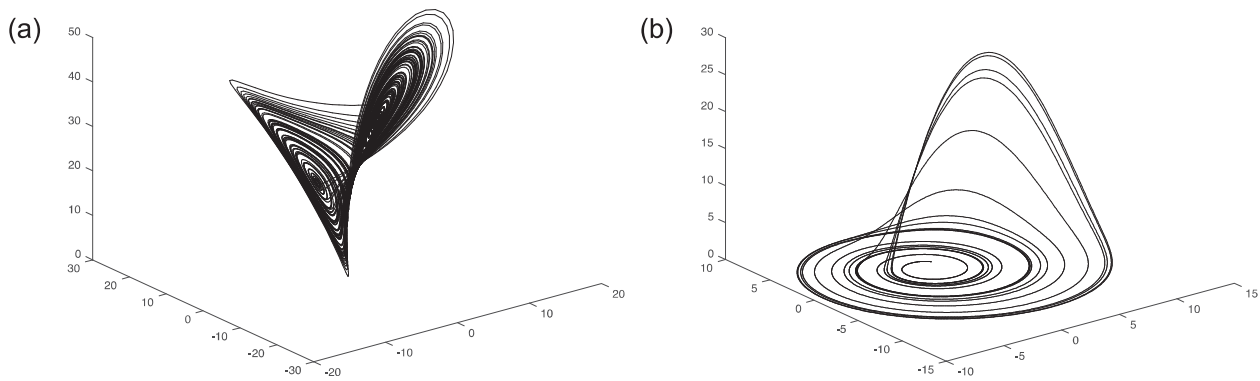


Figure 5. Three-dimensional dynamics of the Lorenz system (parameter settings: $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$) (a) and the Rössler system (parameter settings: $a = 0.2$, $b = 0.2$, and $c = 5.7$) (b).

the Rössler system, we use the common parameter settings $a = 0.2$, $b = 0.2$, and $c = 5.7$. The resulting dynamics for both systems are displayed in Figure 5. Because the dynamics are three-dimensional, examining temporal correlations between the two systems can be done using *MdCRQA3* – that is, a cross-recurrence plot between two time-series with three constituent dimensions each. Conventionally, an integer number d behind the abbreviation *MdCRQA* (i.e., *MdCRQAd*) indicates the unembedded dimensionality d of the underlying time-series (Wallot et al., 2016). Because we know that both systems are three dimensional, and because we know that we have captured each of these three dimensions properly, no embedding is necessary (i.e., $D = 1$), and thus the delay parameter is irrelevant and can be set to an arbitrary value (e.g., $\tau = 1$), as it will not be applied to reconstruct a phase-space.

Note, that when using *MdCRQA*, the order in which the dimensions of the two multidimensional time-series of interest are entered into the analysis can matter for the results. For example, if one were interested in capturing joint emotional arousal between two people, A and B, captured by multidimensional physiological measures using heart-rate, skin-conductance, and electromyography (e.g., Mønster et al., 2016), then it would be necessary to enter the dimensions of the time-series in the same order (i.e., for the first time-series: 1. heart-rate of person A, 2. skin-conductance of person A, 3. electromyography of person A; and for the second time-series 1. heart-rate of person B, 2. skin-conductance of person B, 3. electromyography of person B).

However, the relation between the dimensions of the two systems is arbitrary (i.e., the dimension x in the Lorenz system is *a-priori* not in any particular way more an equivalent to the dimension x in the Rössler system compared to the dimension y in the

Rössler system). Hence, there is no reason to match them in a particular way. Changing the order in which the dimensions of the time-series from the Lorenz and Rössler systems are entered into *MdCRQA* corresponds to a rotation of the phase-space profile in the coordinate system. Hence, by systematically changing the order in which the dimensions are entered allows us to investigate potential effects of relative rotation on the similarity of the joint dynamics of the two systems.

If you want to follow this example, you can find the respective Matlab- and R-functions in the Appendix to this manuscript, together with the commands used to generate the following results. The data for the Rössler and Lorenz system – together with .m- and .R-files are also available on GitHub: <https://github.com/Wallot/MdCRQA.git>.

Now, we can go through all possible permutations of relative pairings of the axis x , y , and z for the Lorenz and Rössler systems, and investigate their potential effects on the resulting *MdCRPs* and recurrence measures. Figure 6 presents the multidimensional recurrence plot and some of the recurrence measures (*REC*, *DET*, *ADL*, *MDL*) for the six possible combinations of relative rotation (i.e., order of entering the constituent dimensions of each time-series). As can be seen, throughout those different rotations, cross-recurrence plots look qualitatively similar, and the resulting cross-recurrence measures are of relatively similar magnitude as well.

Now we can compare the solutions of the three-dimensional *MdCRQA3* with cross-recurrence results when only individual dimensions of the two systems would be used together with phase-space reconstruction (as in conventional one-dimensional *CRQA*, equaling *MdCRQA1*). To perform phase-space reconstruction, the parameters τ and D need to be estimated. As described above, the average mutual

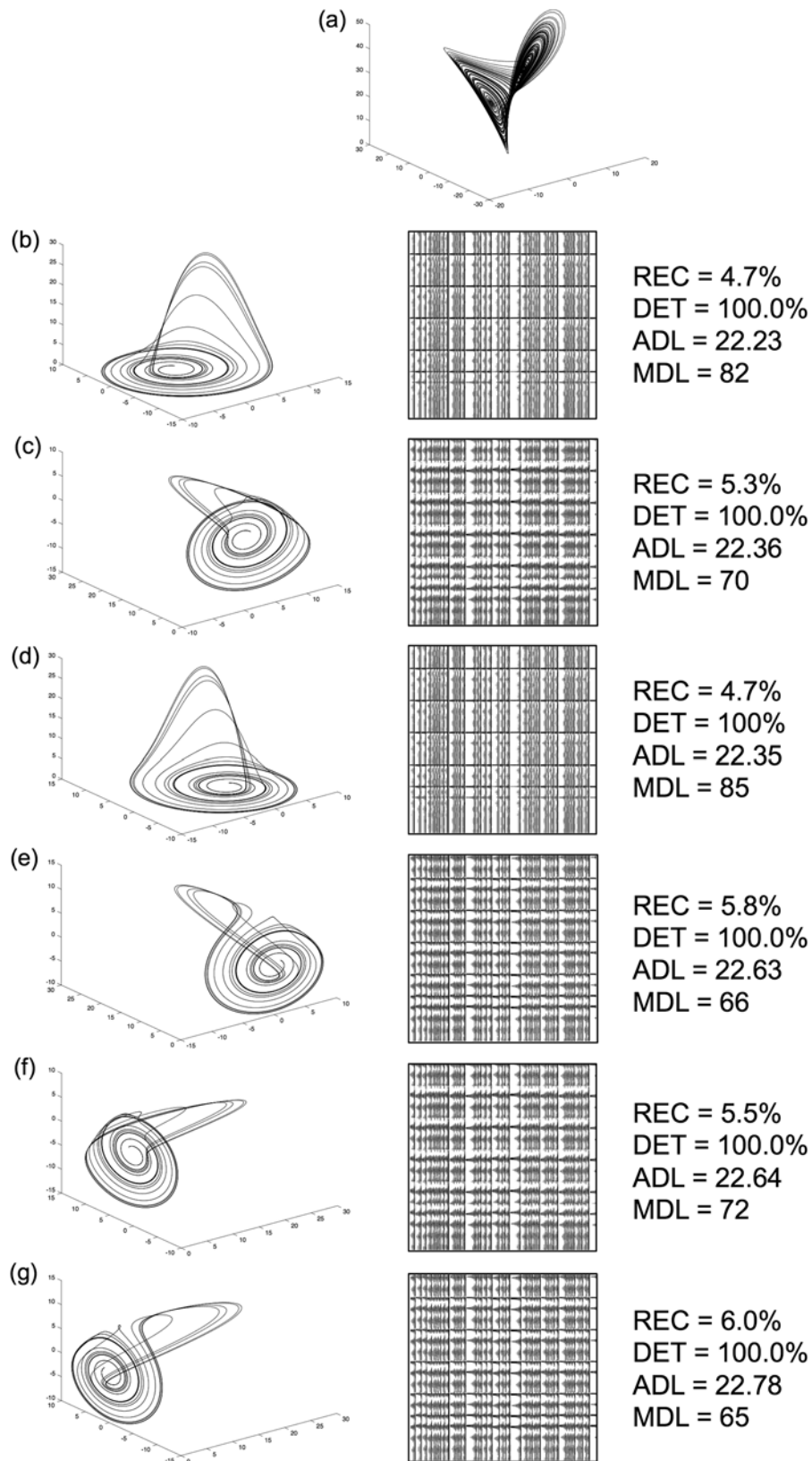


Figure 6. Multidimensional cross-recurrence between the three-dimensional Lorenz system and the three-dimensional Rössler system across all possible relative rotations (i.e., order in which the different dimensions of the time-series are entered into *MdCRQA*): For the Lorenz system, this is x, y, z (a), for the Rössler system these are x, y, z (b), x, z, y (c), y, x, z (d), y, z, x (e), z, x, y (f), z, y, x (g). Displayed are the resulting *MdCRPs* and four cross-recurrence measures (*REC*, *DET*, *ADL*, *MDL*). Parameter settings: $\tau = 1$, $D = 1$, Euclidean phase-space normalization, $r = 0.5$.

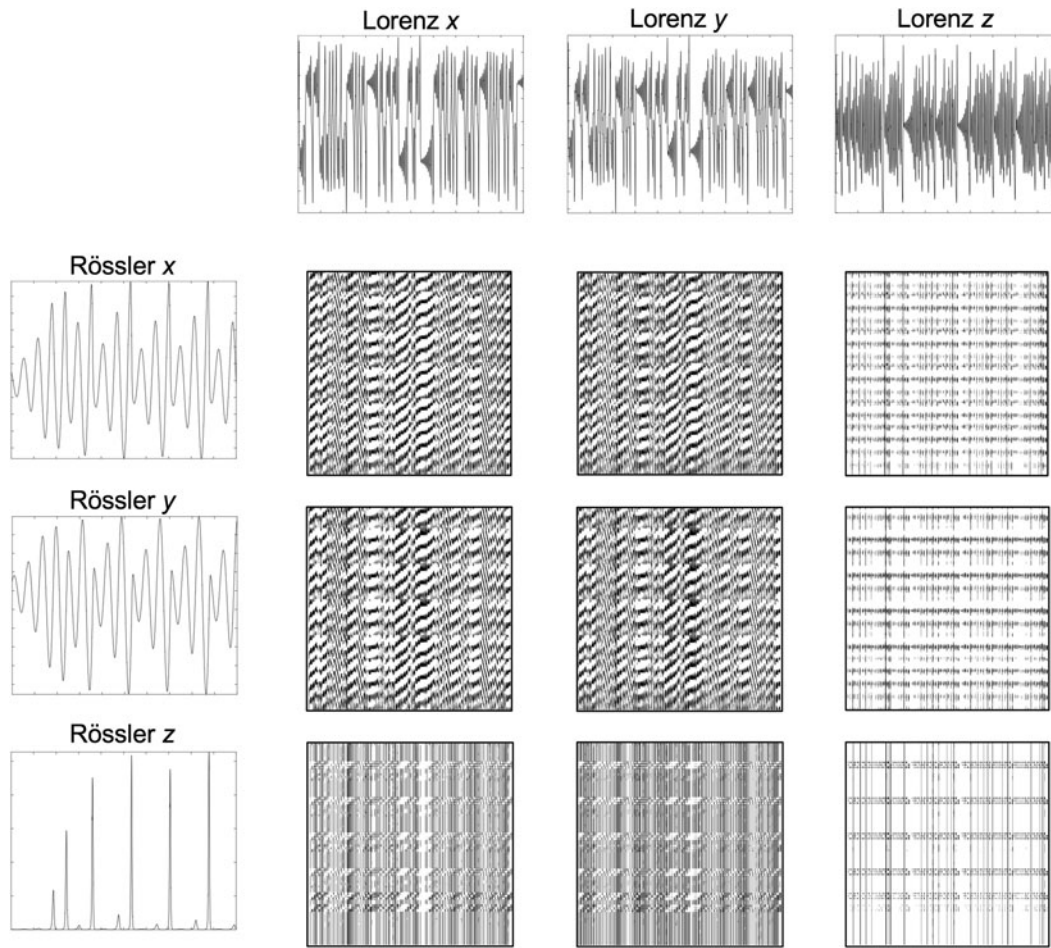


Figure 7. Individual dimensions of the Lorenz and Rössler systems, and cross-recurrence plots between the individual dimensions. Comparing the cross-recurrence plots from Figure 5 with Figure 6 shows that the estimation of cross-recurrences between the individual dimensions is more variable compared to the estimation of cross-recurrences between the complete, three-dimensional time-series in different relative rotations. Parameter settings: $\tau = 120$, $D = 3$, Euclidean phase-space normalization, $r = 0.5$. τ was chosen to be 120 after estimating an optimal delay parameter for both time-series using the mutual average information function. D was chosen to be 3 because both systems are known to be three-dimensional.

Table 2. Cross-recurrence measures for the different combinations of dimensions of the Lorenz and Rössler systems from Figure 6.

	Lorenz x	Lorenz y	Lorenz z
Rössler x	REC = 6.9%	REC = 7.31%	REC = 0.1%
	DET = 99.9%	DET = 99.8%	DET = 99.9%
	ADL = 16.18	ADL = 12.55	ADL = 13.23
	MDL = 96	MDL = 87	MDL = 30
Rössler y	REC = 6.6%	REC = 7.08	REC = 0.1%
	DET = 99.9%	DET = 99.8%	DET = 99.9%
	ADL = 16.24	ADL = 12.51	ADL = 12.07
	MDL = 98	MDL = 84	MDL = 30
Rössler z	REC = 1.73%	REC = 3.19%	REC = 0.1%
	DET = 99.9%	DET = 99.6%	DET = 99.5%
	ADL = 9.62	ADL = 11.20	ADL = 10.98
	MDL = 35	MDL = 37	MDL = 36

information function is used to estimate τ for each individual dimension and picked a value that would be a local (or absolute) minimum fitting *both*, time-series dimensions from the Lorenz and the Rössler system that are entered into the analysis, resulting in

$\tau = 120$. The embedding dimension was set to $D = 3$, because both systems are known to be three-dimensional. The phase-space normalization and threshold parameter (r) were kept the same as above.

As can be seen in Figure 7, when the joint dynamics of the two systems are assessed using pairs of the individual dimensions, CRPs appear more variable in terms of recurrence density, but also type of dynamics (while some feature a dominant strip or checkerboard pattern, others show more diagonal recurrences). Finally, examining the resulting recurrence measures for these analysis in Table 2 confirms that cross-recurrence estimates using pairs of one-dimensional time series is more variable. This suggests that *MdCRQA* indeed provides more informative estimation of recurrence properties, allowing to investigate the full, three-dimensional dynamics. It allows examine the degree of temporal correlation between two multidimensional systems, where higher RQA-measures

indicate stronger correlation for individual states (i.e., REC) or trajectories (i.e., DET, ADL, MDL). Moreover, it provides more stable estimates for cross-recurrence properties of multidimensional time-series [for example, comparing the standard deviation for the cross-recurrence measures for the univariate and multidimensional cases shows lower dispersion for in the multidimensional case for REC ($SD_{CRQA} = 3.28$ vs. $SD_{MdCRQA} = 0.54$), DET ($SD_{CRQA} = 0.13$ vs. $SD_{MdCRQA} = 0.08$), ADL ($SD_{CRQA} = 2.24$ vs. $SD_{MdCRQA} = 0.21$), and MDL ($SD_{CRQA} = 30.76$ vs. $SD_{MdCRQA} = 8.33$)].

Example II: Using DCRP of MdCRPs to capture leader-follower dynamics in multidimensional time-series

Besides calculating MdCRPs to capture overall similarity of the dynamics of two multidimensional time-series, MdCRQA also allows the calculation of leader-follower dynamics using the Diagonal Cross-Recurrence Profile (DCRP). We will exemplarily re-analyze a couple of data sets from Louwse et al. (2012), that measured multiple facial movements and types of gesture during conversation. Here, we will analyze three types of facial movements (smiling, nodding, raising the eye-brows) from two dyads over a range of conversation.

As mentioned in the introduction, in their original study Louwse et al. (2012) used one-dimensional CRQA to assess leader-follower relations among the individual features and found coupling among those during conversation (i.e., if one interlocutor nodded, this likely lead the other interlocutor to nod as well at a certain delay). In the study, both participants of a dyad were given maps, but one participant received a detailed map, while the other received a map where some of the details (the color-coding of objects) was partly removed. The participant with the fully intact map (the instruction “giver”) had to describe a path through the objects on the map to the participant with the degraded map (the instruction “follower”) to reproduce the path, and participants could talk about the information on the maps in order to successfully solve this problem.

However, because CRQA is a one-dimensional correlation technique, only individual features could be analyzed, not groups of features of whole-face (or gesture) coupling during conversation. A study by Wallot, Mitkidis, McGraw, and Roepstorff (2016), similarly collected multidimensional data, investigating hand movements of the left and right hand of participants that built model cars together in dyads. They used MdRQA to analyze those data, but because this analysis technique treads different multivariate time-

series as constituents of a *single* time-series, leader-follower relationships between measures cannot be computed with it (Wallot, Roepstorff, et al., 2016), and hence the method would have been unsuitable for the data in Louwse et al. (2012). However, MdCRQA can do so.

The example data for each dyad and session from Louwse et al. (2012) are organized similarly to the data from the Lorenz and Rössler systems presented earlier in the text. For each participant, we have a three-dimensional time-series about nodding, smiling, and movement of eye-brows, which are the three columns of each multidimensional time-series. However, the data are binary-categorical, where a “1” indicates the presence of a feature at a certain time point and “0” indicates its absence.

This simplifies the parameter selection, because the categorical data are usually not embedded, hence $\tau = 1$ and $D = 1$, even though this can be useful depending on the nature of the data, especially when more than two categories are present (cf. Orsucci et al., 2006). Moreover, the threshold parameter r needs to be set to an arbitrarily tiny value (e.g., $r = 0.00001$), because for categorical data, we only want to have exact cross-recurrences between categories.

Also, no phase-space normalization performed, because normalizing the phase-space does not make sense for categorical data as in our example. Finally, because we are not interested in the overall pattern of cross-recurrences, but in leader-follower relationships, we apply the DCRP-functions (see Appendix) to the resulting CRPs, using ± 20 lags (i.e., diagonals around the central diagonal).

The resulting DCRP based on the multidimensional CRP of the three features (nodding, raising eye-brows, and smiling) as a function of j (the lag) across eight sessions with two dyads is presented in Figure 8a. Here, it can be seen that the common smiling, nodding, and eye-brow behavior is coupled across a range of lags across the diagonal (i.e., lag 0) and that there is a peak in REC_t at lag +5, indicating that on average, the facial expression of the instruction giver preceded that of the instruction follower by 5 lags. Because the data has not been embedded, and because coding of the facial movements was done in 250 ms intervals (Louwse et al., 2012), this tells us that the facial expression of the instruction “follower” followed the facial expression of the instruction “giver” most often with a delay of 1250 ms. Also, note that the sum of REC_t across the 20 lags is higher on the side the “follower” lagging the “giver”, another indicator of asymmetric coupling.

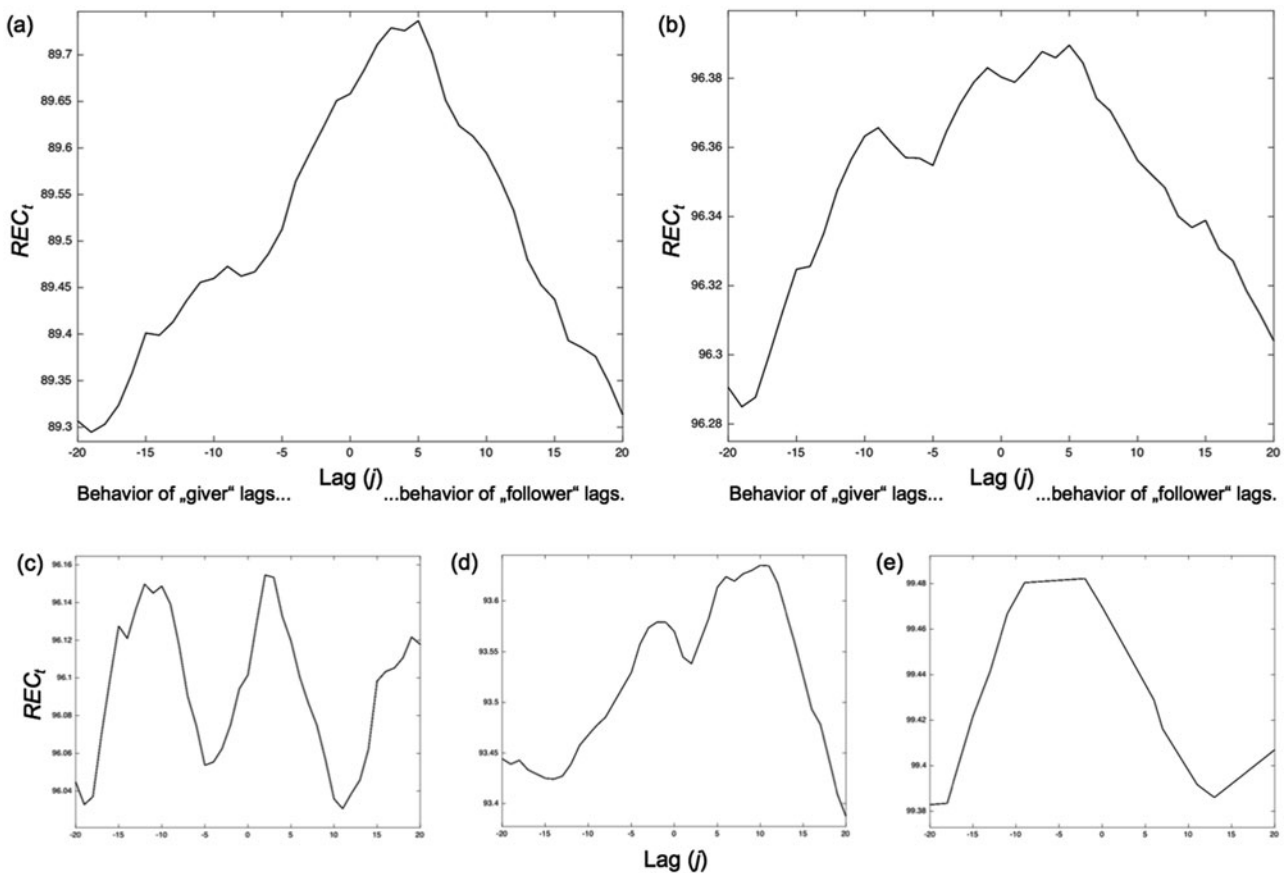


Figure 8. *DCRP* of the three-dimensional time-series with the dimensions nodding, raising eye-brows and smiling (a), average of the individual *DCRPs* of nodding, raising eye-brows and smiling, and individual *DCRP* of nodding (c), raising eye-brows (d), and smiling (e). Negative lags indicate that the behavior of the instruction “giver” followed those of the instruction “follower”, positive lags indicate that the behavior of the instruction “giver” preceded those of the instruction “follower”. As can be seen, the *DCRPs* of the individual behaviors do not show clear evidence for the instruction “follower’s” behavior being led by the behavior of the instruction “giver”. The average suggests that such behavior is happening, but leading and following behavior is relatively broadly stretched out across many lags, with a noticeable bi-modal component. In contrast, the *DCRP* based on the multidimensional analysis shows a much clearer peak at the positive lags, indicating that the behavior of the instruction “follower” is led by the behavior of the instruction “giver”.

If we compare the *DCRP* of the multidimensional analysis with the average of the *DCRPs* of the individual expressions (Figure 8b), we see that the coupling activity is much more evenly stretched out across the lags and provides less clear evidence for the instruction “follower” reacting to the instruction “giver”. This is because the “follower” reactions to the facial expression of the instruction “giver” not only by mimicking the expression, but also by reacting with other, complementary expressions (e.g., smiling in response to raising eye-brows) – an interaction that is not captured by the individual plots. Finally, as one would expect, the *DCRPs* of the individual behaviors provide the least clear picture for a leader-follower relationship, because they do not capture interactions between different facial expressions, but also because individual facial expression might not occur only a few times during a conversation.

Note again, that it is, of course, important that the order of the dimensions of each time-series are the same. If data of one participant are entered in the order “nodding, smiling, moving eye-brows” and “smiling, nodding, moving eye-brows”, then phase-space profiles of the time-series are wrongly oriented with regard to each other, and this will lead to faulty results, because in this case the dimensions have meaning relative to each other.

Further, note that we coded the presence and absence of features the same way for both participants (i.e., for both participants, a “1” indicates the presence of nodding, for example, and “0” the absence). Hence, REC_t is based on similarities of the simultaneous (or time-lagged) presence and absence of the features smiling, nodding, and moving eye-brows, which leads to very high values of REC_t for binary data of comparatively low dimensionality. If one is interested in

specifically basing cross-recurrences only of the occurrences of a complex, such as the facial expression smiling plus nodding plus moving eye-brows, one would need to code the absence of those features differently (e.g., as “0” for one participant, and as “-1” for the other participant), which eliminates cross-recurrences due to the simultaneous absence of the complex, and usually leads to very low values of *RECT* (cf. Louwerse et al., 2012).

Finally, one issue pertains to significance of a REC_t -value at a specific lag, i.e., to decide whether a certain amount of recurrences at a specific lag indicate significant (time-lagged) coupling. Several methods have been proposed to assess significance. All involve the comparison of the original *DCRPs* to *DCRPs* of those surrogate data. A liberal criterion can be defined by shuffling the original data and perform the analysis again on the shuffled data. Contrasting this surrogate with the original *DCRPs* shows significant *RECT* values compared to a baseline where all temporal correlations have been removed. A more conservative criterion can be defined by using false-pairs – i.e., pairing participant 1 of dyad 1 with participant 2 of dyad two, which performed the same tasks, but with a different partner, which leaves the overall dynamics due to task effects intact, and only shows similarities between the original time-series that are due to online interaction (e.g., Richardson & Dale, 2005).

A final note on requirements and possible applications of *MdCRQA*

The *MdCRQA* is an analysis technique for quantifying the correlation or coupling between two multidimensional time-series. However, it is important to understand that the analysis actually contains two conceptual variants that users can choose between – depending on the nature of the data and the goals of the analysis. One variant starts by assessing the need for embedding the time-series into a higher-dimensional phase-space (see the Section “Parameter estimation”). This may invite theoretical and practical problems. That is, one needs to have a certain prior understanding of the dimensions of a multidimensional time-series – for example, if one is interested in measuring the physiological dynamics of arousal, one needs to know the relevant physiological characteristics that form valid dimensions of a multidimensional arousal time-series (e.g., heart rate, breathing, body temperature...) and one needs to be able to distinguish them from unrelated variables that should not be counted toward the dimensions of a time-series.

Because the estimation of embedding parameters currently proceeds on the basis of the individual estimates for each dimension of a time-series, one needs to be aware that just adding variables to the dimensions of a time-series can only decrease the need for embedding if they really capture a valid dimension of the multidimensional system.

Alternatively, *MdCRQA* can be applied as a straight-forward correlational analysis technique for multidimensional time-series if no embedding is performed (e.g., Example II). Here, *MdCRQA* simply provides measures of the correlation strength between two multidimensional time-series. Whether one chooses to embed or not does not rule out that the results of the two types of analysis would not be similar, but it may be that they are not. Hence, users of the method need to be aware of this distinction.

Related to the issue of whether to embed or not is the question of data point requirements on a time-series. Technically, all that the analysis requires is a series of data points, i.e., two. Practically, this depends on sufficiently capturing the important dynamics of interest in the data. Hence, sufficient length of a time-series is not so much a matter of a specific number of data points, but rather of sampling fast enough as to correctly capture the fastest relevant time-scales, and measuring long enough for the important dynamics to unfold. However, embedding a time-series costs data points (see again the Section “Parameter estimation”), and if one chooses to embed, one needs to have sufficient data points to “pay” for the embedding process plus sufficient data points left for the actual analysis. Often, nominal time-series require less data points than continuously sampled ones, and applications for *CRQA* have ranged from using less than 40 data points to more than 10,000.

Finally, this brings us to a last question of how varied the composition of a time-series in *MdCRQA* can be. As has been noted above, recurrence-based analysis can be used on nominal sequences, inter-event times, or continuously sampled time-series. However, not all of these data types can be combined. Obviously, combining inter-event times with continuously sampled data mandates a continuous transformation of the inter-event times to match the sampling frequency of the continuously sampled data (such as converting RR-intervals to beats-per-minute – for specific suggestions on how to do such a conversion optimally for the purposes of recurrence analysis, see Wallot, Fusaroli, Tylén, & Jegindø, 2013). However, nominal data are not readily combinable with inter-event times or continuously sampled time-series

within the framework of *MdCRQA*. This is, because the proper definition of recurrences of nominal sequences requires exact identification of each value, where the threshold parameter needs to be set to very tiny value. The consequence of this would be a low number of recurrences for the continuous data, and even if it would yield sufficient recurrences to calculate reliable values for the recurrence measures, the result would reflect merely a weighted average of the continuous dynamics by categorical data.

Conclusions

In the present paper, I presented a new method for correlating multidimensional time-series: Multidimensional Cross-Recurrence Quantification Analysis (*MdCRQA*). Note that such a multivariate extension is also possible via multiple joint recurrence plots (Marwan et al., 2007). However, joint recurrence plots have a special limitation in this context, because they are always “dominated” by the smallest recurrence structures of the individual plots, which might not allow to capture multivariate dynamics appropriately (for a discussion of the issue, see Wallot et al., 2016). *MdCRQA* does not have such limitations. The technique can be used to assess correlation and coupling among multidimensional time-series, and it works with interval-scale as well as categorical data. Besides providing overall measures of the common dynamics of two multidimensional time-series, it also allows to capture leader-follower relationships (i.e., time-lagged coupling) between the time-series. Potential applications in psychological research are coupling between effectors during action coordination (motion capture), between multiple (neuro-)physiological measures that capture emotional or cognitive states or between multiple groups of item-response or scales that capture change of a psychological state or construct over time (such as personality characteristics and mental disorders). The analysis is very well suited to applications of multidimensional joint action analysis, such as joint gaze positions, body movements, or categorically coded data.

Article Information

Conflict of interest disclosures: The author signed a form for disclosure of potential conflicts of interest. The author did not report any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The ideas and opinions expressed herein are those of the author alone, and endorsement by the author’s institution is not intended and should not be inferred.

References

- Abarbanel, H. (1996). *Analysis of observed chaotic data*. New York: Springer. doi:10.1007/978-1-4612-0763-4.
- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing*, 16(4), 325–332. doi:10.1007/s10339-015-0648-2.
- Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2017). Multiple coordination patterns in infant and adult vocalizations. *Infancy*, 22(4), 514–539. doi:10.1111/infa.12165.
- Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time-series: An R package. *Frontiers in Psychology*, 5, 510. doi:10.3389/fpsyg.2014.00510.
- Coco, M. I., Dale, R., & Keller, F. (2018). Performance in a collaborative search task: The role of feedback and alignment. *Topics in Cognitive Science*, 10(1), 55–79. doi:10.1111/tops.12300.
- Dale, R., Kirkham, N. Z., & Richardson, D. C. (2011). The dynamics of reference and shared visual attention. *Frontiers in Psychology*, 2, 1–11. doi:10.3389/fpsyg.2011.00355.
- Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, 21(04), 1153–1161. doi:10.1142/S0218127411028970.
- Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing social interactions: The promises and challenges of using cross recurrence quantification analysis. In N. Marwan, M. Riley, A. Giuliani, & C. L. Webber Jr. (Eds.), *Translational recurrence*. Springer proceedings in

- mathematics* (pp. 137–155). London: Springer. doi:10.1007/978-3-319-09531-8_9.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, 40(1), 145–171. doi:10.1111/cogs.12251.
- Kennel, M. B., Brown, R., & Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6), 3403. doi:10.1103/PhysRevA.45.3403.
- Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E. M., Wallot, S., ... Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, 108(20), 8514–8519. doi:10.1073/pnas.1016955108.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404–1426. doi:10.1111/j.1551-6709.2012.01269.x.
- Marwan, N. (2017). Cross Recurrence Plot Toolbox 5.22 (R32.1). Retrieved November 7, 2017, from <http://tocsy.pik-potsdam.de/CRPtoolbox/>
- Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5–6), 237–329. doi:10.1016/j.physrep.2006.11.001.
- Mønster, D., Håkonsson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & Behavior*, 156, 24–34. doi:10.1016/j.physbeh.2016.01.004.
- Nomikou, I., Leonardi, G., Rohlfing, K. J., & Rączaszek-Leonardi, J. (2016). Constructing interaction: The development of gaze dynamics. *Infant and Child Development*, 25(3), 277–295. doi:10.1002/icd.1975.
- Orsucci, F., Giuliani, A., Webber, C., Zbilut, J., Fonagy, P., & Mazza, M. (2006). Combinatorics and synchronization in natural semiotics. *Physica A: Statistical Mechanics and Its Applications*, 361(2), 665–676. doi:10.1016/j.physa.2005.06.044.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical review letters*, 45, 712–716. doi: 10.1103/PhysRevLett.45.712.
- Richardson, D. C., & Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6), 1045–1060. doi:10.1207/s15516709cog0000_29.
- Rössler, O. E. (1976). An equation for continuous chaos. *Physics Letters A*, 57(5), 397–398. doi:10.1016/0375-9601(76)90101-8.
- Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 142–177). NSF. Retrieved September 22nd, 2017, from <https://www.nsf.gov/sbe/bcs/pac/nmbs/chap4.pdf>
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 201–208. doi:10.1037/0096-1523.33.1.201.
- Shockley, K., Butwill, M., Zbilut, J. P., & Webber, C. L. (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, 305(1–2), 59–69. doi:10.1016/S0375-9601(02)01411-1.
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326–332. doi:10.1037/0096-1523.29.2.326.
- Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898, 366–381. doi:10.1007/BFb0091924.
- Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes*, 54(5–6), 382–405. doi:10.1080/0163853X.2017.1297921.
- Wallot, S., Fusaroli, R., Tylén, K., & Jegindø, E.-M. (2013). Using complexity metrics with R-R intervals and BPM heart rate measures. *Frontiers in Physiology*, 4, 211. doi:10.3389/fphys.2013.00211.
- Wallot, S., & Grabowski, J. (2018). A tutorial introduction to Recurrence Quantification Analysis (RQA) for key-logging data. In E. Lindgren & K. Sullivan (Eds.), *Advances in writing research using key-logging*. Berlin: Springer.
- Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics with recurrence-based analyses: A tutorial introduction to Cross-Recurrence Quantification Analysis (CRQA), Diagonal-Cross-Recurrence Profiles (DCRP), and Multidimensional Recurrence Quantification Analysis (MdRQA) in R.
- Wallot, S., Mitkidis, P., McGraw, J. J., & Roepstorff, A. (2016). Beyond synchrony: Joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PLoS One*, 11(12), e0168306. doi: 10.1371/journal.pone.0168306.
- Wallot, S., & Mønster, D. (2018). Calculation of average mutual information (AMI) and false-nearest neighbors (FNN) for the estimation of embedding parameters of multidimensional time-series in Matlab. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2018.01679.
- Wallot, S., O'Brien, B. A., & Van Orden, G. (2012). A tutorial introduction to fractal and recurrence analyses of reading. In G. Liben, G. Jarema, & C. Westbury (Eds.), *Methodological and analytical frontiers in lexical research* (pp. 395–430). Amsterdam: John Benjamins. doi: 10.1075/bct.47.
- Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional Recurrence Quantification Analysis (MdRQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action. *Frontiers in Psychology*, 7, 1835. doi: 10.3389/fpsyg.2016.01835.
- Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. In ??, *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 121–126). Austin, TX: Cognitive Science Society.

Webber Jr, C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 142–177). NSF. Retrieved September 22nd, 2017, from <https://www.nsf.gov/sbe/bcs/pac/nmbs/chap2.pdf>

Appendix 1: Abbreviations and notation

- The dot (as in \dot{x}) denotes the time derivative.
- ADL** – Average Diagonal Line length, a recurrence measure.
- AMI** – Average mutual information.
- CR** – Cross-recurrence point.
- CRP** – Cross-Recurrence Plot (plot of cross-recurrence structure between two single one-dimensional time-series).
- CRQA** – Cross-Recurrence Quantification Analysis (analysis based on a *CRP*).
- d** – dimensionality of a multivariate time-series.
- D** – Embedding parameter for phase-space reconstruction.
- DCRP** – Diagonal Cross-Recurrence Profile to analyze follower-leader dynamics.
- DET** – Percent determinism, a recurrence measure.
- FNN** – False-nearest neighbor(s).
- MdCR** – Cross-recurrence obtained from multidimensional time-series.
- MdCRP** – Multidimensional Cross-Recurrence Plot (plot of cross-recurrence structure between two multidimensional time-series).
- MdCRQA** – Multidimensional Cross-Recurrence Quantification Analysis (analysis based on a *MdCRP*).
- MDL** – Maximum Diagonal Line length, a recurrence measure.
- MdRP** – Multidimensional Recurrence Plot (plot of auto-recurrence structure of a single multidimensional time-series).
- MdRQA** – Multidimensional Recurrence Quantification Analysis (analysis based on a *MdRP*)
- r** – Threshold parameter (also known radius parameter) that defines the neighborhood size in phase-space within which coordinates are counted to be recurrent.
- REC** – Percent recurrence, a recurrence measure.
- RP** – Recurrence Plot (plot of auto-recurrence structure of a single one-dimensional time-series).
- RQA** – Recurrence Quantification Analysis (analysis based on an *RP*).
- τ – Delay parameter (tau) for phase-space reconstruction.
- $\Theta(x)$ – Heaviside step function of x .

Appendix 2: Rössler and Lorenz example

The following text provides a brief walk-through on how to compute the results in Example I for Matlab and R. The .m- and .R-files – as well as the data – can be found as online supplements to this paper at the webpage of Multivariate Behavioral Research. If you are fluent in R as well as Matlab, it is recommended to use Matlab for running MdCRQA for longer time-series. While R and Matlab perform similarly for time-series with up to 5000 data

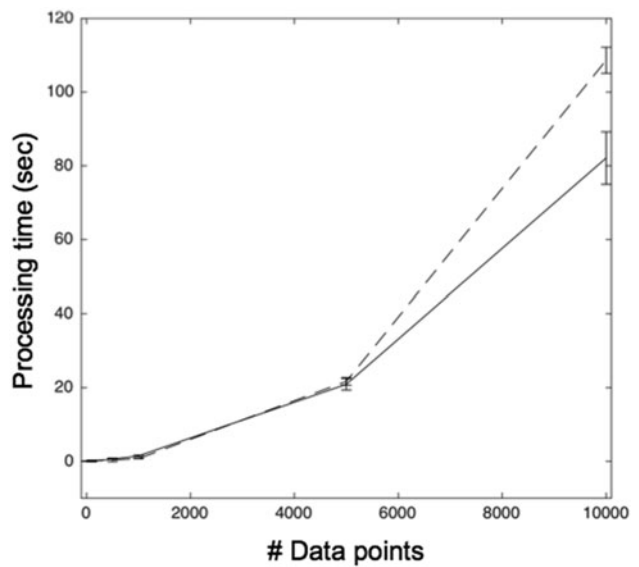


Figure 9. Average processing time for the R- and Matlab-functions with different time-series length. The solid line shows processing time for Matlab, the dotted line for R. The error bars show the standard deviation. The functions were run for sample sizes of 50, 100, 500, 1000, 5000, and 10000, each 10 times. The computer was a MacBook Pro (2.8 GHz Intel Core i7, 16GB ram, OS C 10.11.6). The R version was 3.4.0, the Matlab version was 2015b.

points (the average difference in processing time is at about 1 s), Matlab outperforms R with larger time-series of more than 5000 data points (i.e., the average difference in processing time is at about 26 s) – see Figure 9.

Assuming that we have stored the data of the Lorenz and Rössler systems in two matrices or data frames, “lorenzData” and “rosslerData”, and with three columns representing the three dimensions (x , y , z) of the of these systems, and rows equal to the number of data points, we can now compute the first MdCRQA using the Matlab and R functions presented in this paper as follows:

```
[CRP, RESULTS, PARAMETERS] = MdCRQA
(lorenzData, roesslerData, 1, 1, 'euc', 0.5)
or
results <- mdcraq(lorenzData, roesslerData, 1, 1,
'euc', 0.5)
```

Note, that if you have saved your data in individual variables, you will need to convert them to a matrix using, for example, “as.matrix(bind_cols(lorenzDataX, lorenzDataY, lorenzDataZ))” or similar expressions to enter the multivariate time-series as matrices in R, e.g.:

```
results <- mdcraq(as.matrix(bind_cols(lorenzDataX, lorenz
DataY, lorenzDataZ)), as.matrix(bind_cols(rosslerDataX,
rosslerDataY, rosslerDataZ)), 1, 1, 'euc', 0.5, 2, 2, 0,
metric = 'euclidean')
```

For in both languages, the first argument of the mdcraq-function is the variable that contains the data of time-series 1 (here: the Lorenz data), the second argument is the variable that contains the data of time-series 2 (here: the Rössler data), the third argument is the value for the embedding dimension D (here: 1), the fourth argument is the value for the delay parameter τ (here: 1), the fifth argument is the

value for the phase-space normalization procedure (here: Euclidean), the sixth argument is the value for the threshold or radius parameter r (here: 0.5), the seventh argument is the value for the minimum of the diagonal lines (here: 2), the eighth argument is the value for the minimum length of the vertical lines (here: 2), and the ninth argument is the value for z -scoring the data (here: 0, i.e., no z -scoring). For the R-function, there is also a tenth argument for setting the distance metric for the distance matrix, because this is done using the R-function `cdist` (here: `metric = 'euclidean'`). Because the data for the two systems are already correctly scaled with regard to each other, we do not need to z -score them or apply another standardization technique on the individual dimensions of each time-series, which might often be recommendable for empirical data where differences in magnitude between participants or different measures would otherwise mask differences or similarities in the sequential properties of the data that we are interested in (Shockley, 2005).

To view the resulting *MdCRPs* in Matlab and R, type:

```
im2bw((CRP-1)*-1)
or
image(results$CRP)
```

To view the resulting *MdCRQA* measures in Matlab and R, type:

```
RESULTS
or
results$RESULTS
```

To perform the analysis for the x -dimension of the Lorenz system and the x -dimension of the Rössler system (i.e., upper-left *CRP* in [Figure 7](#)), type the following in Matlab or R:

```
[CRP, RESULTS, PARAMETERS] = MdCRQA(lorenz
Data(:,1),roesslerData(:,1), 120, 3 , 'euc' , 0.5)
and
results <- mdcrqa(lorenzData[,1], roesslerData[,1], 120,
3, 'euc', 0.5)
```