

Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R

Sebastian Wallot

To cite this article: Sebastian Wallot (2017): Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R, Discourse Processes, DOI: [10.1080/0163853X.2017.1297921](https://doi.org/10.1080/0163853X.2017.1297921)

To link to this article: <http://dx.doi.org/10.1080/0163853X.2017.1297921>



Published with license by Taylor & Francis Group, LLC © 2017 Sebastian Wallot



Published online: 17 Apr 2017.



Submit your article to this journal [↗](#)



Article views: 549



View related articles [↗](#)



View Crossmark data [↗](#)

Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R

Sebastian Wallot 

Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

ABSTRACT

Processes of naturalistic reading and writing are based on complex linguistic input, stretch-out over time, and rely on an integrated performance of multiple perceptual, cognitive, and motor processes. Hence, naturalistic reading and writing performance is nonstationary and exhibits fluctuations and transitions. However, instead of being just complications for the analysis of such data, they are also informative about cognitive change, fluency, and reading or writing skill. To use and quantify such dynamics, one needs appropriate statistics that capture these aspects. In this article I introduce Recurrence Quantification Analysis (RQA) as a tool to capture such dynamic structure. After a conceptual introduction of the analysis, I present a step-by-step tutorial on how to run RQA using R. Guidance is given with regard to common issues and best practices using this time-series analysis technique. Finally, I review previous results from studies applying RQA to reading and writing and summarize current hypotheses and interpretations of RQA measures in the context of reading and writing.

Introduction

Capturing the process of reading creates substantial challenges for investigators interested in more naturalistic stimuli or more complex task setups that allow participants greater degrees of freedom for processing and comprehension. This is due to the dynamic nature of the reading process, which exhibits quite complex dynamics: For example, reading of connected text leads to highly nonstationary reading times (Wallot, Hollis, & van Rooij, 2013), where a single word can have a significant impact on how reading performance changes over the next several hundred words (Booth, Brown, Eason, Wallot, & Kelty-Stephen, 2016) and where qualitative changes in reading performance occur (Wallot, O'Brien, Haussmann, Kloos, & Lyby, 2014), which are potentially indicative of qualitative shifts in the mental representations as a function of comprehension (Stephen, Dixon, & Isenhower, 2009). Moreover, the problem with naturalistic texts is that they are highly structured stimuli that cannot be experimentally manipulated, as one can do with random word lists in lexical decision and naming tasks, because the different psychologically relevant text descriptors are highly correlated among each other (Graf, Nagler, & Jacobs, 2005) and with the syntactical structure of a text (e.g., Keller, Carpenter, & Just, 2001) and their effects change across the time-course of reading (McNerney, Goodwin, & Radvansky, 2011; Teng, Wallot, & Kelty-Stephen, 2016; Wallot et al., 2013). On the one hand, many studies have been conducted that circumvent these problems by focusing on outcomes of connected text reading only, but in doing so, leaving the process of reading in the dark. On the other hand, it is clear that many of the experimental word or sentence reading tasks that are the workhorses of reading process research cannot incorporate naturalistic text materials.

CONTACT Sebastian Wallot  sebastian.wallot@aesthetics.mpg.de  Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany.

Published with license by Taylor & Francis Group, LLC © 2017 Sebastian Wallot

This is an Open Access article. Non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly attributed, cited, and is not altered, transformed, or built upon in any way, is permitted. The moral rights of the named author(s) have been asserted.

In short, reading of naturalistic texts is a very integrated performance of coordinated cognitive processes, based on complex stimulus material. The question is how can one go about analyzing such an integrated activity, exhibiting nonstationary data and qualitative shifts in performance? One way to begin is to use analysis techniques that have been designed to investigate systems that exhibit behaviors as complex as the process of naturalistic text reading and that are tailored to the problems that such data-series pose. Natural candidates are all analysis techniques from the field of nonlinear dynamic systems, such a fractal (e.g., Peng, Havlin, Stanley, & Goldberger, 1995) or entropy analysis (e.g., Bandt & Pompe, 2002). In the present article I introduce one of these analysis techniques, Recurrence Quantification Analysis (RQA; Webber & Zbilut, 1994), that is particularly versatile and has already been applied to the analysis of reading (O'Brien & Wallot, 2016; O'Brien, Wallot, Haussmann, & Kloos, 2014; Wallot et al., 2013, 2014; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012), writing (Wallot & Grabowski, 2013, 2017) and the analysis of texts and transcripts of conversations (Angus, Watson, Smith, Gallois, & Wiles, 2012; Fusaroli & Tylén, 2016; Orsucci, Giuliani, Webber, Zbilut, Fonagy, & Mazza, 2006). RQA captures coordination properties of higher-dimensional systems (such as the integrated performance of multiple perceptual and cognitive processes during reading) and quantifies stability, flexibility, and complexity of their dynamics. Global patterns, such as start-up transients or switching between different modes of reading behavior, are some of the phenomena that can only be observed across multiple observations of the reading process, and RQA is effective at capturing these behaviors, which are potentially informative about the perceptual-motor-cognitive coordination that drives reading. Particularly, RQA-based methods have been very useful in advancing some of the issues surrounding naturalistic reading research, such as relating reading processes to reading comprehension (O'Brien & Wallot, 2016; Wallot et al., 2014). I summarize this research in more detail later on.

With regard to the next sections, the article is structured as follows. First, I introduce the concept of recurrence and the recurrence plot (RP), which is the basis for RQA. Then, I introduce RQA conceptually and discuss with regard to applications in psychological research. This is followed by a practical step-by-step tutorial on how to apply the analysis technique in R. For the examples, simplistic and artificial data are used to better communicate what the analysis does. Finally, some points regarding the application to real behavioral data and the analysis of multiple data sets are discussed.

Concept of recurrence and the RP

"How do you write a story that is 800 words long when you have only 50 different words? Well, you just use the same words again and again."

—Charles Webber

The above quote summarizes most of what you need to know to understand the basics of RQA. Additionally, it leads to a nice and straightforward example of what RQA captures. When I first met Chuck Webber and was listening to one of his lectures on applying RQA, he used the example of the rhyming children's book "Green Eggs and Ham" by Dr. Seuss to illustrate what recurrence is. Let us consider the text snippet "Do you like green eggs and ham? I do not like them, Sam-I-am. I do not like green eggs and ham" from this book. The book is written as a poem, featuring rhyme and meter, which are to a certain extent reflected in the written language: Obvious repetitions that make up the rhythm of the text are direct repetitions of words (such as "green"). Others are reflected in similar syllables, which are similarly written (such as "Sam" and "ham," whose phonological similarity is expressed in the written language by them sharing the letters "am").

We can visualize—and quantify—the rhythmic structure of the text by converting it into an RP. For illustration purposes, we first simplify the line of text by removing all punctuation and spaces and turning all letters into uppercase, leaving us with the following string: "DOYOU LIKE GREEN EGGS AND HAM I DO NOT LIKE THEM SAM I AM I DO NOT LIKE GREEN EGGS AND HAM." This string is now 69 characters long, so it has 69 different positions, which

are filled by 16 different characters. To assess their repetitiveness, we can now simply go through the string from beginning to end and examine which character occurs at what position. For example, the character “D” occurs five times, on positions 1, 21, 26, 46, and 66. At this point, the RP comes in handy: An RP is depicting instances of repetitions in a sequence, and the RP shown in Figure 1 is a representation of this string: “DOYOU LIKE GREEN EGGS AND HAM I DO NOT LIKE THEM SAM I AM I DO NOT LIKE GREEN EGGS AND HAM.” As can be seen, an RP shows recurrences of individual characters in the series as individual black dots (recurrence points) and recurrences of larger sequences of characters as diagonal lines of adjacent recurrence points.

The RP does not represent specific values on specific positions (such as the value “D” on position 1 in the sequence), but rather shows when a specific value repeats itself recurs. Repetitions are marked in black, also called recurrent points. The absence of recurrence in a sequence is marked by a white space. For example, when we look at the first column (i.e., the left-most column) of the plot, we see recurrence points at the positions 1, 21, 26, 46, and 66—everywhere, where we observe a “D” in the sequence. If the character “D” would only occur once in this letter string, it would not generate a recurrence point, and we would see white spaces instead. Note that the first position is the very lower-left in this plot. Similarly, if we count the number of recurrent points in the second column, we see where—and how often—the letter “O” recurs in the sequence.

Note two things at this point. First, you could also count the number of recurrences in the first row (i.e., the lower-most row) of the plot, and you would obtain the same result of recurrences of “D.” This is because the RP is symmetrical about its main diagonal. Second, note the continuous line of recurrence along the main diagonal, from the lower-left to the upper-right. This diagonal is always present in RPs and simply means that the sequence is always the same with itself (i.e., if we look at the “D”s on the vertical axis and the “D”s on the horizontal axis, we see that they occur at the very same positions on both axis).

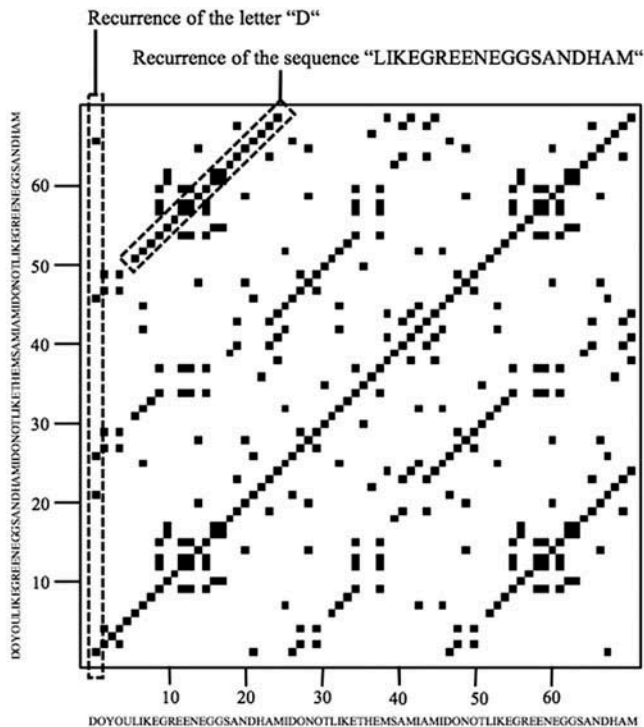


Figure 1. RP for the letter string.

Recurrence points would also give us a first indication of the repetitiveness of the sequence: Counting the number of recurrence points (without the main diagonal, which is always present) in the plot will tell us about the number of instances of individual letters recurring. However, it becomes immediately clear that the plot shows further properties: Diagonal lines on the plot marks repetitions of whole sequences, such as recurrences of the one-word substring “GREEN” or the phrase-string “GREENEGGSANDHAM.” Recurrent patterns can be quantified in several ways. Table 1 gives definitions for four RQA variables that quantify the recurrence of individual elements and larger patterns of connected recurrences.

Such larger patterns are important, because they provide measures for different ways in which a sequence can recur and are informative when comparing different sequences. For example, consider Figure 2. Figure 2a displays the original RP of the “green eggs and ham...” sequence, and Figure 2b displays a shuffled version of the same sequence, where the positions of all letters have been randomly assigned. Although the overall number of recurrences is still the same (i.e., proportion of each letter in the sequence has not changed), the diagonal patterns, which are indicative of larger connected sequences, are gone.

Table 1. Names and definitions of four RQA measures of temporal structure.

Variable Name	Definition	Quantifies
Percent recurrence (%REC)	Sum of recurrent points in RP /size of RP	Repetitiveness of the elements in the sequence
Percent determinism (%DET)	Sum of diagonally adjacent recurrent points /sum of recurrent points in RP	How many of the individual repetitions occur in connected trajectories
Average diagonal line length (ADL)	Average diagonal lines in RP	How long the average repeating trajectory is
Maximum diagonal line length (MDL)	Length of longest diagonal line in RP (excluding the main diagonal)	How long the longest repeating trajectory is

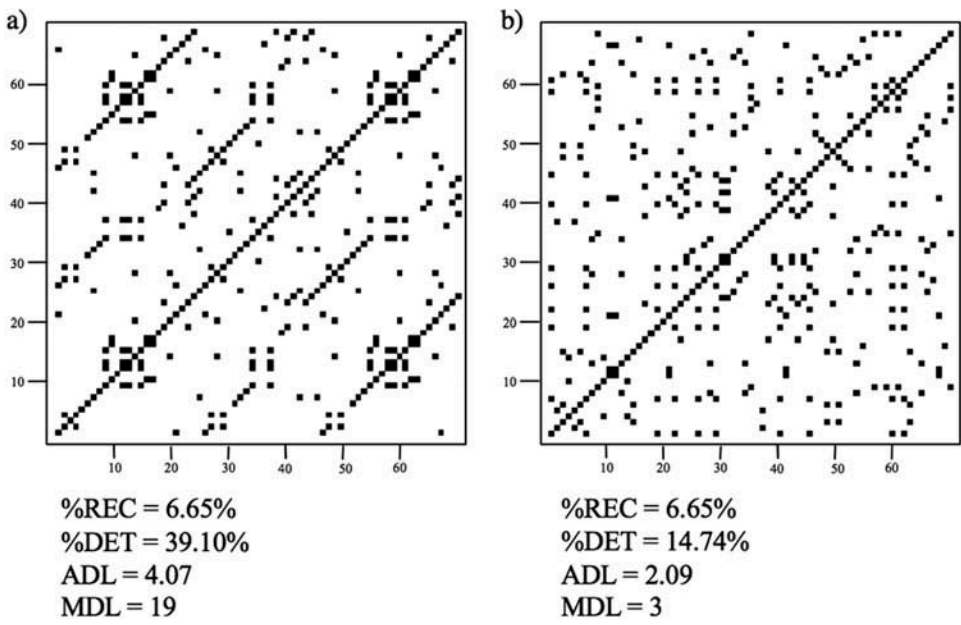


Figure 2. Comparison of RPs of (a) the original “green eggs and ham...” sequence with (b) the sequences when characters are shuffled. Both plots show the same number of individual recurrences (%REC). However, recurrences are much more isolated and evenly distributed in the shuffled string b compared with the original sequence a. This difference can be seen in the three measures that capture the “connectedness” of recurrence points (i.e., %DET, ADL, MDL), which are higher for the original sequence (a) compared with the shuffled sequence (b).

Further down, we will turn to the analysis of interval-scale data, but this example already illustrates the flexibility of RQA, which can be used on interval and nominal variables. To perform such a text analysis, one must, however, replace each letter (and symbol) by a unique number, converting a sequence of characters into a sequence of numbers that uniquely identifies each element. This way, RQA can be used for the analysis of linguistic rhythm, for example, of poems (Orsucci et al., 2006) but also other text types, and assess repetitions, breaks in text structure, or introduction of new language. For example, Lyby (2015; cited in Wallot, Lyby, & van Rooij, 2015) investigated whether recurrence structure of essays during an expressive writing intervention would predict short- and long-term change of stress and depressive symptoms in patients with breast cancer. Increases in %DET of the text structure predicted alleviation of short- and long-term psychological distress and depression. This was interpreted as an increase in the coherence of wording and grammatical structure, which in turn might indicate an increase in the coherence of thinking about one's own situation, instead of a disjoint, incoherent thinking indicating a passive way of coping on the side of patients.

To sum up, an RP is simply a means of displaying repeating values in a sequence. However, the way sequences exhibit repetitions on an RP can be informative. Moreover, RPs allow for a quantification of patterns of repetition, which makes a statistical analysis of such pattern possible. For interval data, such as continuously sampled signals or series of response times, further steps are necessary to obtain a proper RP. However, before walking through the different steps of how to concretely perform such an analysis in R, I summarize some of the research that has been done using RQA, particularly in the area of reading, to illustrate how RQA can be used and how RQA measures can be interpreted.

Applications of RQA in reading and writing research

There is an increasing body of research in psychology that used RQA or its variants to investigate such different topics as motor control and coordination (Abney, Warlaumont, Haussman, Ross, & Wallot, 2014; Farina, Fattorini, Felici, & Filligoi, 2002; Kuznetsov & Riley, 2010; Rhea & Kiefer, 2013; Riley & Turvey, 2002; Riley, Balasubramaniam, & Turvey, 1999), joint action (Mønster, Håkansson, Eskildsen, & Wallot, 2016; Ramenzoni, Davis, Riley, Shockley, & Baker, 2011; Shockley, Santana, & Fowler, 2003), and conversation (Dale & Spivey, 2006; Fusaroli & Tylén, 2016; Louwerse, Dale Bard, & Jeuniaux, 2012; Richardson, Dale, & Kirkham, 2007). However, applications to more classical cognitive tasks and research questions are less developed. A few studies have used RQA in the context of reading, noticeable text reading: O'Brien et al. (2014) had second, fourth, and sixth graders, as well as adults, read a children's story in a self-paced manner, word by word. Figure 3 presents two example data sets from this self-paced reading study. Figure 3a shows data from a second grader and Figure 3b from an adult, both reading silently in a self-paced manner. As can be seen, the time-series of word reading times show a higher degree of intermittent fluctuations and periods of comparatively long reading times in the second grader, whereas reading times of the adult are much more homogenous. Moreover, as the RP suggests, adult reading times are more structured, as if adults read the text in "bursts" of connected stretches of words, whereas second graders read the text more in a word-by-word manner.

Word reading times were analyzed using RQA, and the results revealed that reading time-dynamics became more structured with age: The older—and putatively the more competent—the reader, the higher the %REC, %DET, and ADL values of reading times, suggesting that more competent readers seem to read texts more smoothly, with less intermittent periods composed of long and short reading times, and in strokes of multiple words in quick succession.

Wallot et al. (2014) presented a reanalysis of the data and showed that %DET of reading times was predictive of text comprehension (each child and adult answered comprehension questions about the text right after reading) and reading speed and was a significantly better predictor for comprehension than reading speed. Additionally, O'Brien and Wallot (2016) showed that the relation between %

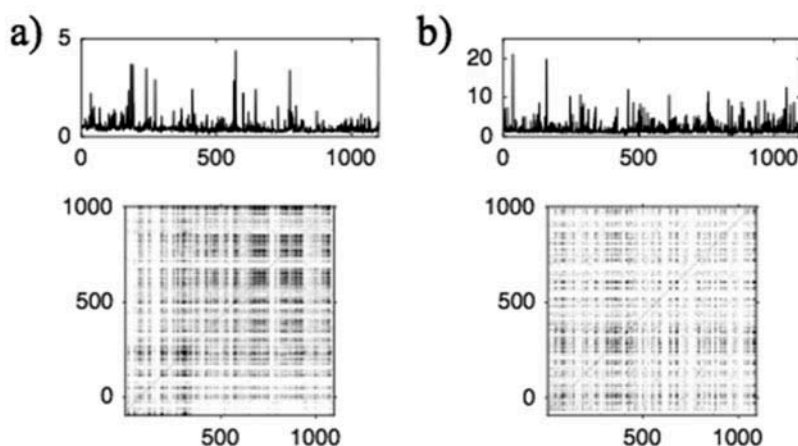


Figure 3. Example time-series and RPs of word-reading times for (a) a second grader and (b) an adult, reading the same text of at about 1100 words of length.

DET in reading times and comprehension also held up for bilinguals. Note that only %*DET* was used in these studies, because the three measures based on diagonal lines in the RP (i.e., %*DET*, *ADL*, *MDL*) showed very high interrelations and %*DET* of reading times provided the overall best single predictors for text comprehension. In interval-type data, such as word reading times, the different RQA measures can often not be dissociated.

Finally, Wijnants et al (2012) showed that dyslexic readers—compared with an unimpaired control group—exhibited significantly lower %*DET* in their reading times during a word naming task, indicating a perturbation of reading performance on the side of the dyslexics. In general, %*DET* of reading times seem to be an indicator of reading process fluency, correlating positively with reading speed and comprehension and negatively with the presence of reading problems. The results are surprising in that earlier studies that have tried to relate process and outcome measures of reading (i.e., comprehension) have often found null results, that is, no relation between measures of the reading process and subsequent comprehension tests (LeVasseur, Macaruso, & Shankweiler, 2008; LeVasseur, Macaruso, Palumbo, & Shankweiler, 2006; but see Schroeder, 2011), whereas RQA measures seem to be much more sensitive and reliable quantifiers of this part of the reading process.

However, somewhat contradictory results come from a study of repeated text reading by Wallot et al. (2013), where adult readers read a long story of 12,000 word repeatedly in a self-paced manner: Here, %*DET* in reading times dropped off with the second reading, despite the fact that rereading of text is generally supposed to increase situational reading fluency (Samuels, 1979). However, Wallot et al. (2013) also noted that participants would greatly speed up with the second reading, often pressing the response key to reveal new words at the limit of the speed of repeated button presses. This measurement-based limitation, and a potential effect of repeated exposure related to boredom or fatigue on the side of participants, might have actually lead to a situation where participants just “glanced” over the text, and response times were not reflective of individual word reading times anymore, leading to the increase of error variance in the reading time series and hence lower %*DET*.

Even fewer studies have been conducted on writing using RQA: Wallot and Grabowski (2013) investigated how different complexity measures, among them RQA, characterized qualitatively different writing tasks, namely repeated typing of the same sentence, copy-typing of text, and open writing. These tasks have shown qualitatively different cognitive demands (Grabowski, 2008), but applications of complexity measures showed that these qualitatively different writing tasks can actually be localized on a continuum of cognitive complexity, perhaps increasing the greater demand for the integration of multiple cognitive functions from the simplest tasks (repeated typing) to the most complex one (open writing). Furthermore, in a different study, Wallot and Grabowski (2017)

investigated copy-typing performance: Participants either copy-typed a text in their native language or a text in a language they did not comprehend, while their key-strokes were being measured.

Figure 4 shows data of key-stroke times of participants copy-typing text in their native language (Fig. 4a) or text in a language they could not comprehend (Fig. 4b). Similar to the results in reading, copy-typing a text one can understand proceeds in bursts of key-strokes (like the reading times of a competent adult reader), whereas copy-typing of a text that is incomprehensible proceeds similarly more in a single-key-stroke manner (like the reading times of a beginning reader in second grade). Accordingly, %*DET* of typing performance was higher for copy typing of the comprehensible text, suggesting that typing of comprehensible texts proceeds in bursts of typing activity, whereas copy-typing of incomprehensible text is more akin to individual button pressing. In line with the results on reading reviewed above, higher values of RQA %*DET* also seem to be indicative of fluency of the writing process.

Hence, RQA measures that capture aspects of the sequential structure of the reading process can well be interpreted as indicators of reading (and writing) fluency. This psychological notion of flow also nicely coincides with a statistical notion of flow, where high values of %*REC*, %*DET*, *ADL*, and *MDL* are also indicative of smooth, laminar flow in physical systems. Laminar flow in reading process measures can be interpreted as an indicator of reading fluency. Hence, measures of temporal structure (i.e., %*REC*, %*DET*, *ADL*, *MDL*) in time-series of reading performance might be an indicator of reading fluency, particularly effortlessness of the reading process and ease-of-comprehension during reading (Wallot, 2015, 2016).

This interpretation also fits with studies that have used a windowed RQA approach to investigate qualitative changes in representations following spontaneous mathematical discoveries in participants. For example, Stephen, Dixon, and Isenhower (2009) had participants solve a series of gear tasks, where participants were presented with a set of interlinked gears on each trial. Their task was to predict in which direction the last gear of a particular series of gears would turn, given that the first gear turned in one direction, left or right. During each trial, participants' finger movements were recorded. Initially, participants traced the gears with their finger to walk themselves through each series of gears and find out in which direction the last gear would turn. During the course of several such gear setups, participants spontaneously discovered that all they needed to do is to count the number of gears to find out in which direction the last gear would turn: If the number of gears was even, the last gear would turn counter-clockwise to the first one. If the number of gears was odd, the last gear would turn into the same direction as the first one. Interestingly, recurrence entropy (i.e., entropy of the diagonal line distribution in the RP) of the finger movements predicted when a

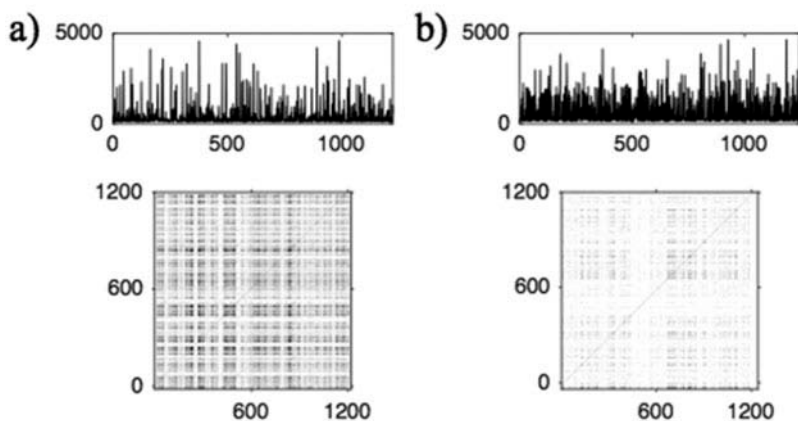


Figure 4. Example time-series and RPs of word-reading times for key-strokes during copy-typing of a (a) comprehensible and (b) incomprehensible text.

participant would understand this mathematical problem and change his or her representation of the task. This was indicated by a peak in recurrence entropy (i.e., a drop in the temporal structure of the finger-movement time-series), but remarkably this drop was observed on the trial before participants would start to use the new counting strategy.

Interestingly, a similar effect is observed in the word reading times of participants transitioning from one reading task to another. Similar to Stephen et al.'s (2009) interpretation, this transition effect is supposed to reflect a reorganization of the representational structure, which in the case of reading might be believed to be a strong need for updating one's situation model (Zwaan, Magliano, & Graesser, 1995) or the falling apart thereof (as it seems difficult to sustain a coherent situation model for random text).

Figure 5 shows unpublished data illustrating a transition during reading: Figure 5a shows an RP of a reader reading a connected text in a self-paced manner and Figure 5b a reader reading the same text, but with the order of words in the second half of the text being randomized (Wallot, & Kelty-Stephen, 2016). Although the RP of connected text reading exhibits relatively constant reading time dynamics with a long-term transient, indicating a relatively monotonic increase of recurrences over the course of reading, the RP of a reader transitioning from structured text to random word reading shows that each of these two are differently structured kinds of reading activities, indicated by the recurrences that mark the first and second half. However, adapting to the new task takes time and perturbs reading performance, as evident in the white band showing absence of recurrences in reading times connecting these two tasks.

Stephen et al.'s (2009) interpretation of changes in RQA reflecting changes in cognitive-representational structure fits the reading fluency interpretation of RQA measures of reading, where the absence of such changes (i.e., high %DET) is assumed to be an index of smooth and stable cognitive performance during reading. In addition, phase-transitions in reading (quantifiable by RQA) might indicate qualitative changes in the situation model, where existing information is not incrementally updated based on already existing dimensions in this model but perhaps rather indicating a more radical change in the architecture of the situation model. An example of this is related to the emergence of new dimensions in the model or drop-out of previously present ones.

Hence, RQA measures of temporal structure might serve as a good indicator of processing or production (and perhaps conversation) fluency in time-series and might be related to ease-of-comprehension and language skill. Additionally, RQA is suited to detect qualitative changes in process data, for example during reading, which might reflect emergence of or change in representation and comprehension.

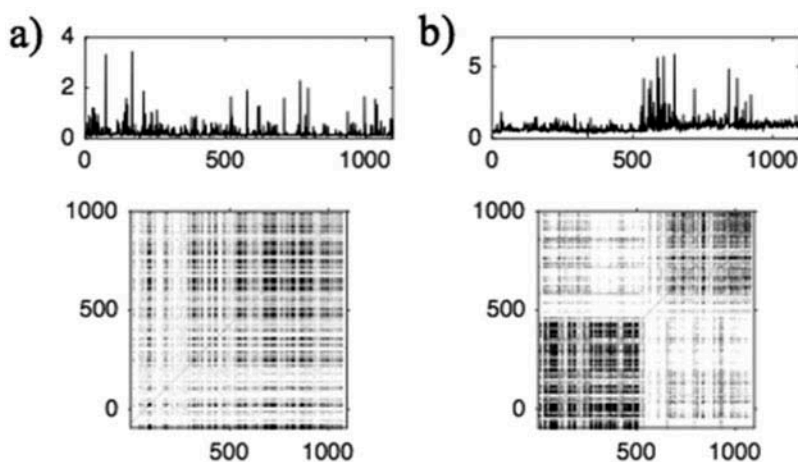


Figure 5. Example time-series and RPs of word-reading times for (a) reading of a connected text and (b) reading of the same text, with the word order in the second half of the text being randomly shuffled.

Recurrence quantification analysis

Originally, RQA was developed to characterize the behavior of time-series with nonlinear temporal properties that are the result of multiple interdependent variables (Takens, 1981). The foundations of the technique go back to problems in physics and physiology (Eckmann, Kamphorst, & Ruelle, 1987; Webber & Zbilut, 1994), but the underlying considerations readily apply to psychological data: Usually, when a phenomenon is empirically investigated, only one or few observables are available. For example, in the case of reading, we might have a one-dimensional time-series of fixation durations or sentence reading times available. However, this time-series is a result of multiple underlying variables that constitute it, for example, aspects of visual perception, lexical memory, and higher-order semantic processing. If these subcomponents are independent of each other, variability in this time-series is fully decomposable, as far as it reflects these components. If the subcomponents are *interdependent*, however, one can gain additional information about the reading process by describing the dynamics of integrated functioning of the component processes.

This can be done by representing the subcomponents in a phase-space, which is a multidimensional space where each component contributes one dimension to that space. However, because we have only a single one-dimensional time-series available, we need to infer the other dimensions of the phase-space from the one-dimensional measurement. This can be done by the method of phase-space reconstruction by time-delayed embedding (Takens, 1981). In a nutshell, the method implies that one can retrieve the multidimensional dynamics from a one-dimensional time-series by plotting that time-series against itself at a certain time delay τ . Note, however, that in the case of a phase-space reconstructed from psychological data, the dimensions are not necessarily interpretable in the sense of specific cognitive functions, as our example using perceptual, memory, and reasoning processes suggests. They might be a more abstract reflection of the statistical structure of the data—we come back to this issue when discussing the interpretation of RQA below.

In our text example, we only considered direct recurrences of the actual values of one sequence/time-series. We circumvented the step where a time-series is first converted into a phase-space profile, from which in turn an RP is computed. For many nominal variables, such as characters, the direct construction of RPs is sensible, but the basis of the RQA for interval-scale data is phase-space reconstruction through time-delayed embedding. A phase-space is a space in which all possible states of a system under study can be charted. The starting point for phase-space reconstruction is a time-series (or sequence), which is a one-dimensional observable of measured values x :

$$\mathbf{x} = (x_1, x_2, x_3, \dots, x_n) \quad (1)$$

where \mathbf{x} is a vector with values x_1 to x_n representing the time-series or sequence. To reconstruct the phase-space profile of \mathbf{x} , we need to estimate the appropriate dimensionality D of the dynamical system from which we have sampled \mathbf{x} , as well as the appropriate time-delay τ at which we have to shift values of \mathbf{x} for each reconstructed dimension. Basically, all that we are doing is to take D -times values from \mathbf{x} , each time shifted by a delay of τ to construct a coordinate in the D -dimensional phase-space. So, given some value for D and τ , the first coordinate of our reconstructed phase space, \mathbf{V}_1 , is

$$\mathbf{V}_1 = (x_1, x_{1+\tau}, x_{1+2\tau}, \dots, x_{1+(D-1)\tau}) \quad (2)$$

Note that the elements of \mathbf{V}_1 are all elements from \mathbf{x} , starting with x_1 and then using values of \mathbf{x} at later intervals of multiples of τ , such as $x_{1+\tau}$. Now we can construct the second coordinate of our phase space, \mathbf{V}_2 by starting with x_2 in the same manner. Because we lose a number of data points from \mathbf{x} equaling the number of additional embedding dimensions times the value of the delay, we can maximally construct $n - (D - 1)\tau$ such coordinates, which are then arranged in a matrix:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_{n-(D-1)\tau} \end{pmatrix} = \begin{pmatrix} x_1 & x_{1+\tau} & \dots & x_{1+(D-1)\tau} \\ x_2 & x_{2+\tau} & \dots & x_{2+(D-1)\tau} \\ \vdots & \vdots & \dots & \vdots \\ x_{n-(D-1)\tau} & x_{n-(D-2)\tau} & \dots & x_n \end{pmatrix} \quad (3)$$

Note that the rows are the D -dimensional phase-space coordinates that we set out to construct, whereas the columns are new D -dimensional data points of our new phase-space series taken from values of \mathbf{x} . Hence, the row index is a measure of time and each column index corresponds to a dimension in phase-space. Thus, the row vectors \mathbf{V}_i constitute points in the phase-space portrait of the multidimensional dynamics of the system from which the observable x was taken.

We are now interested in repeating patterns in our reconstructed phase-space. However, repetitions are never exact in empirical data, either due to intrinsic variability of the organism or due to measurement noise. Hence, we need a way to define which coordinates count as repetitions in our phase-space and which ones do not. To do so, we use a threshold parameter T , which is effectively a tolerance threshold within which we count coordinates in phase space as recurrences. An RP is effectively a thresholded representation of the phase-space profile:

$$RP_{ij} = \Theta(T - \|V_i(x) - V_j(x)\|) \quad (4)$$

Here, Θ is the Heaviside step function, effectively converting pairs of coordinates whose distance is greater than T into 0 (nonrecurrent values) and pairs of coordinates whose distance is less than T into 1 (recurrent values).

To illustrate this more concretely, let us use the examples of a sine-wave and a sine-wave with added noise. Our goal is to describe the dynamics of these two signals in terms of their phases and then in terms of their RPs that are derived from the phase-spaces. Figure 6 illustrates this: Figure 6a

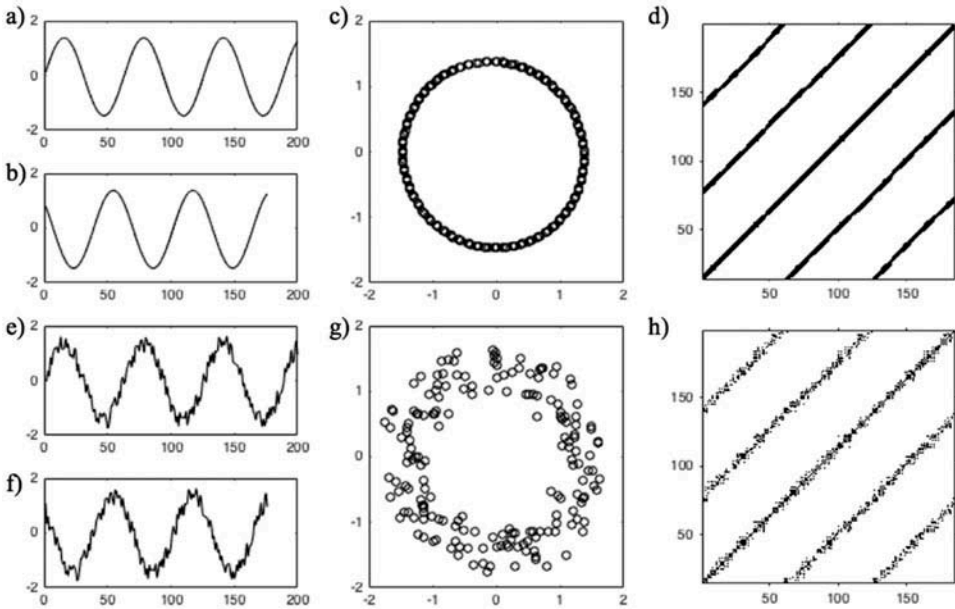


Figure 6. Illustration of phase-space reconstruction and resulting RPs of a sine-wave and a noisy sine-wave. (a) Sine-wave time-series and (b) surrogate series of the original. (c) Resulting two-dimensional phase-space obtained by plotting the original series against its surrogate and (d) RP of the sine-wave. (e) Noisy sine-wave time-series and (f) surrogate series of the noisy original. (g) Resulting two-dimensional phase-space obtained by plotting the noisy series against its surrogate and (h) RP of the noisy sine-wave. In the RPs, black dots indicate instances of recurrence, whereas white spaces indicate absence of recurrence.

shows the original sine-wave and Figure 6e the same sine-wave but with some random noise added to it. In this case, we know that the dimensionality Dof both the sine-wave and the noisy sine wave is 2, and the delay τ is a quarter of its phase, in our case equals 16 data points.

Following the steps of phase-space reconstruction outlined above, we first create a surrogate series of our original sine wave (Fig. 6b) and noisy sine wave (Fig. 6f), which is a copy of the original shifted by 16 data points. Now we plot the values for the original and the surrogate series against each other to obtain their reconstructed phase-space profiles (Fig. 6, c and g, respectively). First, note that our phase-spaces have 16 coordinates less than the original data series have data points, because after applying the delay parameter τ to obtain the surrogate embedding dimension, we have no data left for the last 16 data points of the original that we could pair with the surrogate. Second, note the similarities and differences between the phase-spaces: Although the coordinates of both phase-spaces form a circular shape on a plane, the coordinates of the original perfect sine wave sit tightly on a strict circle, but the coordinates of the noisy sine wave are much more spread out.

After applying some threshold parameter T to the distances between coordinates in the phase-space and converting them into an RP, we see that both the plot for the perfect sine-wave (Fig. 6d) and for the noisy sine-wave (Fig. 6h) show a striped diagonal line pattern. However, although the lines of the perfect sine-wave RP are straight and saturated, the lines of the noisy sine-wave RP are speckled.

How do we read such an RP? The dark dots on the plot are instances of recurrence (i.e., where we observe repetitions in the time-series) and the white spaces mark absences of recurrence. The RP basically shows correlations (recurrences) over lags. Time in the RPs runs along the central diagonal, going from the lower-left to the upper right. That means there is always a diagonal line in the RP, because a time-series is always the same with itself at lag0, and hence a phase-space coordinate is always the same with itself at lag0. As we move away from the central diagonal toward the lower-right or upper-left, we observe lagged recurrences (or absence thereof). That is, if we look at the RP in Figure 6d at 50 at the x -axis and 75 at the y -axis, we see whether and how structure in the time-series at data point 50 repeats itself at data point 75. In our example, there is a white space, so no recurrent structure. If we look, however, at 50 at the x -axis and 120 at the y -axis, we see black dots (i.e., recurrent behavior). Note furthermore that the plot is symmetrical about the central diagonal, that is, we look at 50 at the y -axis and 120 at the x -axis we see exactly the same picture of recurrences (or lack thereof). In the case of the sine-wave, we see that the lag0 data repeat themselves deterministically (in the case of the perfect sine-wave) or stochastically (in the case of the noisy sine-wave) at about 65 data points, which corresponds roughly to the phases of the sine-waves.

How much recurrences we see on an RP depends on the value of the threshold parameter T . The larger T (i.e., the greater the range within which we count coordinates in the phase-space as recurrent), the more instances of recurrence we observe. Figure 7 illustrates this: panels a through d show the phase-spaces of the perfect sine-wave and the noisy sine-wave with a small and a large threshold, indicated by the gray circle, whereas panels e through g show the corresponding RPs. As can be seen, recurrences increase with increase of the threshold value. However, at the same time the plots retain their qualitative, striped pattern—as long as the threshold is not set in a way that it is too low (i.e., yielding no recurrences) or too high (i.e., counting every coordinate in phase-space as recurrent).

Now we have seen how we can construct a RP from a time-series and how to interpret it. However, we are ultimately interested in using an RP as a statistical tool that will help us to quantify interesting aspects of our data. In particular, we use four measures that can be derived from the plot, which are percent recurrence (%REC), percent determinism (%DET), average diagonal line length (ADL), and maximum diagonal line length (MDL). Figure 8 illustrates how these measures are computed for the perfect (Fig. 8a) and the noisy sine wave (Fig. 8b), using a subsection of their respective RPs.

When we use the complete plots, we can see that the perfect sine-wave exhibits much more structure on all of the four measures compared to the noisy sine-wave: Compared with the noisy

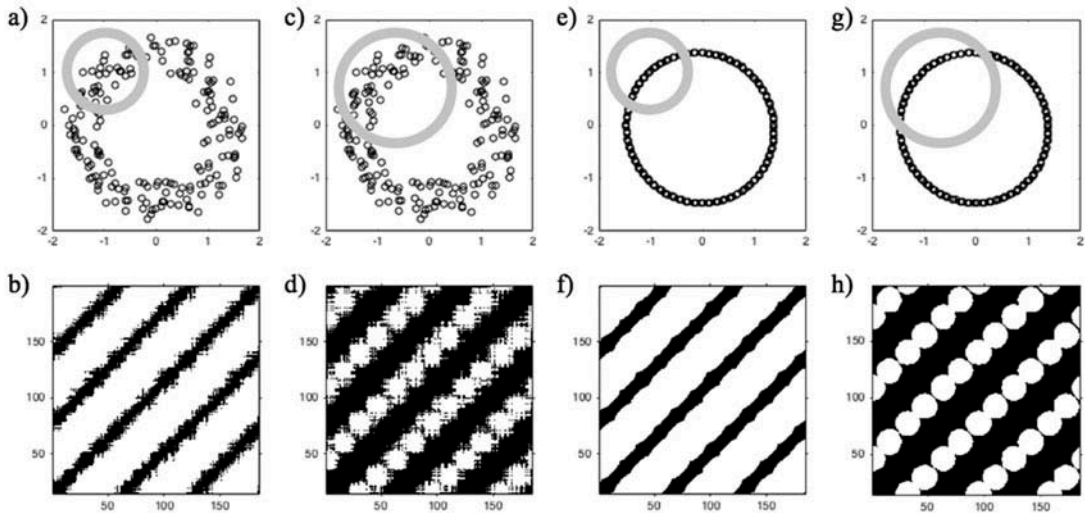


Figure 7. Illustration of the effect of the threshold parameter on the percentage of recurrence points in an RP for the perfect sine-wave and the noisy sine-wave. (a) phase-space portrait of the perfect sine-wave with a low and (c) a high threshold value with their associated RPs (b and d). (e) Phase-space portrait of the perfect sine-wave with a low and (g) a high threshold value with their associated RPs (f and h).

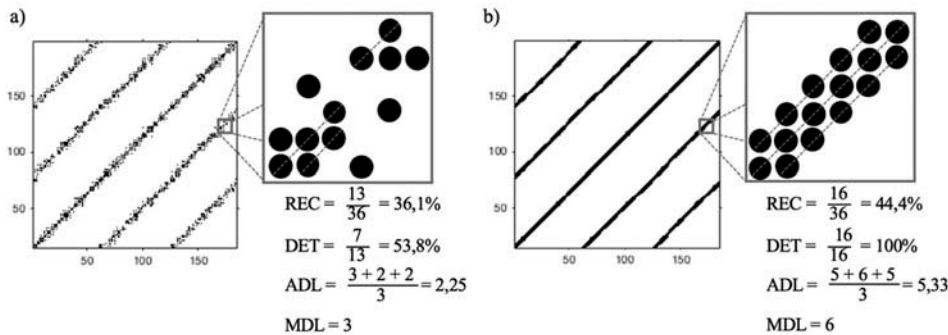


Figure 8. RPs of (a) the perfect and (b) the noisy sine-wave, each with a magnification of a sub-plot of the RPs (6 × 6 = 36 data points), illustrating the calculation of four RQA measures. Again, black dots are recurrent points, whereas white spaces mark the absence of recurrence. In addition to the overall percentage of recurrent points (%REC), diagonally adjacent points (marked by gray dotted lines) provide further information about the temporal structure in a time-series and are the basis for calculating the RQA measures of %DET, ADL, and MDL.

sine-wave, the perfect sine-wave exhibits higher %REC (7.2% vs. 4.6%), higher %DET (99.5% vs. 54.6%), higher ADL (42.3 vs. 2.7), and higher MDL (184 vs. 10). So, due to its temporal arrangement, the perfect sine-wave exhibits not only more individual recurring data points (%REC), but almost all data points are aligned in larger patterns (%DET), with the longest repeating trajectories being almost as long as the whole time-series (MDL).

A note on the R toolboxes needed

For this tutorial we use the following R-packages: ‘tseriesChaos’ and ‘nonlinearTseries’, which provide us with the functions for parameter estimation and running RQA. Moreover, we need the package ‘sparseM’ for plotting. These three packages, as well as their dependencies, need to be installed to follow the next sections. Moreover, there is an R-package for Cross-recurrence

quantification analysis called `,crqa`. We will not use the `,crqa` package because we will not delve into recurrence analysis of more than one time-series. However, the `,crqa` package is computationally very efficient, and it can also be used to reproduce RQA of a single variable (Coco & Dale, 2014). This just as a side note for those who need to analyze long time-series of several 1000 data points or more.

Example creation

We run the remainder of this tutorial on two variables, a sine-wave and a sine-wave with some of its data points shuffled in the middle. Those two signals can be created using the following R commands. To create the sine-wave, type `sine=(sin(seq(0.1,20,0.1)))` # create sine wave. To create the sine-wave with shuffled data points in the middle, type `shuffled_sine=c(sine[1:80],sample(sine[81:120]),sine[121:200])` # create sine wave with its middle 40 data points shuffled. Figure 9 displays the two time-series.

Delay parameter τ

The delay parameter τ can be estimated using the mutual average information function (Thiel, Romano, Read, & Kurths, 2004). It is the basis for unfolding the one-dimensional time-series into a multidimensional phase-space. This is done to ensure that reconstructed dimensions of the phase-space are relatively orthogonal and deliver non-confounded information about the temporal structure in a time-series. To that end, we search for the number of data points at which the time-series is uncorrelated—or minimally correlated—with itself. The mutual average information for the time-series is calculated over a number of lags, every time shifting the time-series by one data point. In practice, we are not searching for the global minimum in mutual average information but rather for a local minimum at which the mutual average information function has dropped considerably and stays relatively constant thereafter. Here, we use the average mutual information function for the package ‘tseriesChaos’ in R, which can be called via `mutual(sine,lag.max = 50)` # run average mutual information.

When estimating τ , we do not usually know how many lags to investigate in the beginning. Hence, we start with an arbitrary number of 50 lags and then see whether the mutual average information function reveals a local minimum. The resulting graph is displayed in Figure 10. As can be seen, mutual average information drops off, reaching a local minimum at lags 13 to 20, and rises again thereafter. Hence, $\tau = 16$ seems to be an appropriate value for the delay parameter (but any value in the range of 13–20 would be equally suggestive at this point). If a local minimum is not observed, one would simply start again, using longer lags, until a minimum appears in the function.

When working with interevent data, such as key-presses collecting response times or series of fixations and saccades, τ is often 1 or close to 1. This is because the interevent times already collapse the dynamics that connect adjacent keypresses or the continuous eye-movement record into one

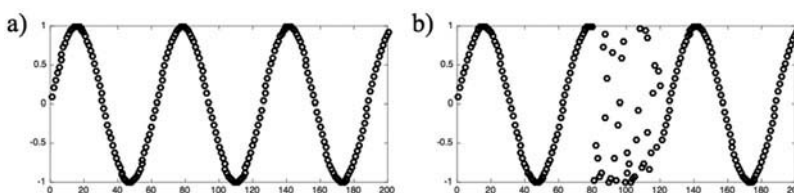


Figure 9. Illustration of the two synthetic signals: (a) a sine-wave and (b) a sine-wave with the data points in the middle randomly shuffled.

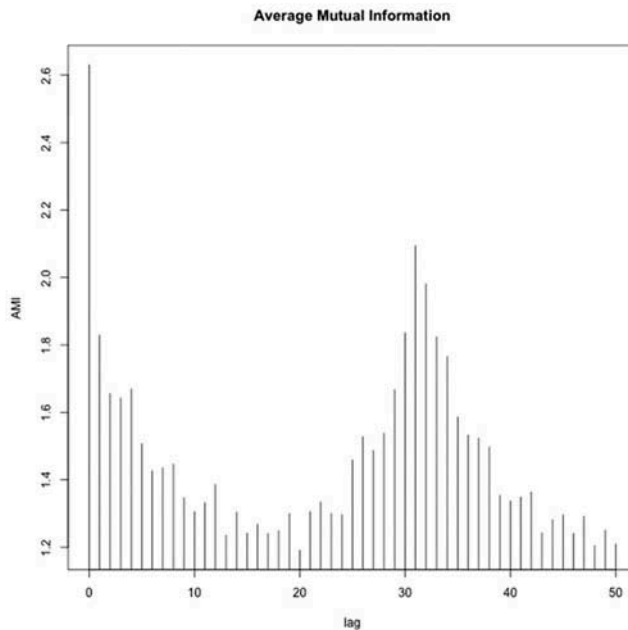


Figure 10. Average mutual information function plot, displaying average mutual information as a function of the first 50 time-lags of the sine-wave series.

single data point, removing much of the redundant information that might be inherent in equally sampled signals (such as in our sine-wave data).

Embedding parameter D

The embedding parameter D can be estimated using the false-nearest neighbor function (Thiel et al., 2004). The embedding parameter D provides the number of embedding dimensions of the phase-space (i.e., how often the time-series has to be embedded by creating surrogate-series from the original time-series), each of them shifted by an additional number of data points equal to the delay parameter τ . If $D = 1$, then we do not need to embed. The phase-space is one-dimensional, and with our time-series data, we already have a one-dimensional signal at hand. If $D > 1$, we want to unfold the one-dimensional time-series into as many dimensions so that we exhaust all information about higher-dimensional dynamics contained in the time-series but not more. To that end, we plot the number of false neighbors in phase-space as a function of the number of embedding dimensions. Again, we are looking for the first local minimum of that function or a point at which the function has exhibited a substantial decrease and remains stable thereafter. False neighbors, here, are coordinates that are close to each other in phase-space, but where their proximity is an artifact of an improper choice of embedding dimensions. The false-nearest neighbor function from the R package ‘tseriesChaos’ can be called via `fnn=false.nearest(sine,m=5,d=15,t=0)` # run false-nearest-neighbor analysis.

Here, m is the embedding dimension D and d is the delay τ . t is an additional parameter, the Theiler-window, that can help to increase the reliability of the false-nearest neighbor analysis and recurrence statistics by removing unwanted short-term temporal correlations (Marwan, 2011). However, we use this parameter here, not least because the kinds of data we are dealing with are usually very stochastic and the application of the Theiler-window often does not result in changes in the analysis outcome. To plot the results of the false-nearest neighbor function, type `plot(fnn)` # plot results of false-nearest neighbor analysis.

Starting with an arbitrary number of five embedding dimensions, we inspect the resulting false-nearest neighbor plot. The number of false-nearest neighbors drops immediately off at a value of 2. Hence, $D = 2$ for our sine-wave data. For empirical data, there are sometimes more values for D that look like reasonable candidates—just as in the plot of the mutual average information function above. In such a case, a somewhat higher embedding dimension might be a better choice and yields more consistent results. However, we do not want to pick unnecessarily high dimensionality, because this shortens the number of data points available for the final analysis by $(D - 1)\tau$ data points.

Norm parameter

As a next step, a norm is usually chosen by which the phase-space is normalized. Normalization of data series is helpful, because by using RQA we are primarily interested in the sequential properties of the time-series and not their absolute values or absolute magnitude of variation, which can be captured by measures of central tendency and dispersion. Figure 11 illustrates the issue. When one creates a copy of the original sine-wave and multiplies its amplitude by 2, the two sine-waves would still exhibit the same dynamics, but the one with the smaller amplitude would show a higher number of recurrences. In fact, this would be just the same as applying twice the magnitude of the threshold parameter to the sine-wave with the smaller amplitude (see Fig. 7). Normalization bypasses this problem by rescaling the phase-space and thus making the phase-space portraits of different time-series comparable.

Several norms can be applied to rescale the phase-space: the minimum norm rescales the phase-space according to the minimal distance between two coordinates in that space, the maximum norm rescales according to the maximum distance between two coordinates, and the Euclidean norm rescales according to the average distance between all coordinates in that space. The main point about the norm parameter is to keep it constant across all data sets (i.e., pick the same parameter to execute RQA for all data sets within a sample or between different samples you want to compare). Unfortunately, the *rqa*-function in the package ‘nonlinearTseries’ does not provide a normalization option. Alternatively, one can normalize the data by z-scoring them before subjecting them to the analysis, which will lead to a similar outcome as phase-space normalization. However, the ‘crqa-package’ by Coco and Dale (2014) provides an option to perform normalization after phase-space

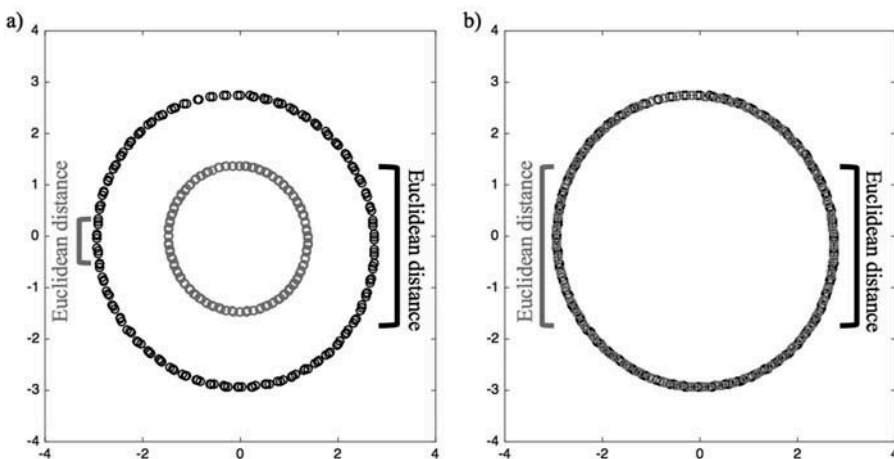


Figure 11. Example of renormalization of the phase-space. (a) Phase-space with two sine-wave portraits that possess the same sequential dynamics, where the black dotted circle comes from a sine-wave with twice the amplitude of the one with gray dotted circles. (b) Phase-space of the same sine-waves after normalization by Euclidean norm (and then rescaling to the distance of the sine-wave with the larger amplitude). Now, the time-series occupy the same neighborhood in phase-space, reflecting their similar.

reconstruction; as mentioned, even though it was developed to analyze cross-recurrences of two variables, one can also use it to reproduce standard RPs.

Selecting the threshold parameter T and running RQA

The value of the threshold parameter T provides the size of the neighborhood in phase-space, within which we count two coordinates as either recurrent (when they lie within the range of the threshold) or not (when they lie farther apart than the range of the threshold). The size of the threshold parameter is directly related to the percentage of recurrent points ($\%REC$), because the higher the threshold, the more coordinates will be counted as neighbors. The aim is to find a threshold value that does not yield too many, but also not too few recurrences, as we would lose information about the dynamics of the time-series in both cases.

There are several suggestions on how to pick the threshold parameter. Webber and Zbilut (2005) recommend a threshold that yields a percentage of recurrence points between 1% and 5%. This is a good suggestion for many physiological time-series. However, $\%REC$ can be much smaller (i.e., $< 1\%$) for signals that are not noisy and have well-defined dynamics (e.g., our sine-wave). On the other hand, $\%REC$ can be a bit higher (i.e., between 5% and 20% or more; see Wallot, O'Brien, & Van Orden, 2012) for noisy interevent-type data, for example, reading times or fixations. Fortunately, RQA results will be relatively stable over a broad range of values of T .

To find out how much percent recurrence a particular threshold T will yield, we need to actually run an RQA function that calculates $\%REC$. Hence, we need to use our parameters of $\tau = 16$, $D = 2$, and begin with an arbitrary $T = 0.5$ by calling the function *rqa* from the R package 'nonlinearTseries' to perform RQA: *rqa(time.series=sine, embedding.dim = 2, time.lag = 16, radius=.5) # run RQA*. Note here that *embedding.dim* is the embedding dimension D , *time.lag* is the delay τ , and *radius* is the threshold parameter T . After running the function with those parameters, it provides us with the following results: $REC = 0.183$, $DET = 0.995$, $Lmean = 39.86$, $Lmax = 183$. Before we look at these results, particularly regarding the amount of recurrences in relation to the selection of the threshold parameter value of $T = 0.5$, we need discuss a few discrepancies in notation and presentation of the results by the *rqa*-function in regards to our notation: REC (recurrence rate) equals our $\%REC$ (percent recurrence) and DET (determinism rate) equals our $\%DET$ (percent determinism). However, although we have so far been talking about percentages, the *rqa*-function outputs rates, so the rate values need to be multiplied by 100 to become percentages. So, $REC \times 100$ yields $\%REC = 22.1\%$, and $DET \times 100$ yields $\%DET = 99.7\%$. The other two variables are numerically equivalent to ours, where $Lmean$ (mean line) equals our ADL (average diagonal line length) and $Lmax$ (max line) equals our MDL (maximum diagonal line length).

Now back to the evaluation of our choice of $T = 0.5$. Because we are dealing with a smooth and continuously sampled time-series (i.e., a sine-wave), $\%REC$ of 22.1% seems too high. A much lower threshold value would already give stable estimates of our recurrence measures. Hence, we can now try to reduce the threshold value to $T = 0.1$ and run the *rqa*-function again: *rqa(time.series=sine, embedding.dim = 2, time.lag = 16, radius=.1) # run RQA*.

Now we obtain $REC = 0.046$, that is, $\%REC = 4.6\%$. This is much more adequate, but in the case of the sine-wave we could probably even lower our threshold toward 1% to 2% recurrence. Note that some other recurrence variables remained numerically very similar: $DET = 0.996$, $Lmax = 183$. Only $Lmean$ increased to 74.47, and this is because the high threshold parameter value yielded unnecessarily many recurrence points and part of the spurious, less coherent recurrence structure that decreased $Lmean$. This can be seen when comparing the RPs for our two choices of the threshold value, $T = 0.5$ versus $T = 0.1$, which we can investigate by calling the *recurrencePlot* function of the 'nonlinearTseries' package: *recurrencePlot(time.series=sine, embedding.dim=2, time.lag=16, radius=.5) # generating RP with $T = 0.5$ and recurrencePlot(time.series=sine, embedding.dim=2, time.lag=16, radius=.1) # generating RP with $T = 0.1$.*

Figure 12 shows both RPs. First, notice another difference to the presentation of our RPs so far compared with the output of the *recurrencePlot*-function: Although in our RPs time at lag0 was running along the main diagonal from the lower-left to the upper-right, in the RPs here, time at lag0 runs from the upper-left to the lower-right. Just like with the different notation for the RQA measures, one needs to be aware that these two kinds of notations as well as these two ways of presenting RPs have both been firmly established, unfortunately. Regarding the plots, in the left plot (Fig. 12a) one can see thicker lines for recurrences but also that these lines start to exhibit some wave-like curvature on their edges, which are absent in the plot on the right panel that has fewer recurrences but also much straighter lines (Fig. 12b). It is these wave-like recurrences that lead to the lower-value of *L*mean when choosing a comparatively large threshold value of 0.5. Overall, however, RQA measures are relatively stable across a large range of parameter selections, and what is even more important for comparison of different data sets, or samples of data sets, is that their *relative* stability is even higher. We discuss usage of RQA in the comparison of multiple data sets or samples in more detail further below.

Windowed RQA

RQA can not only be used to quantify the dynamics of a whole time-series, but it can also be used to investigate changes of the dynamics within the series. Using RQA this way is potentially very interesting for investigators of natural reading and writing tasks, as these tasks are stretched out in time, featuring a longer stimulus set and observation durations without interventions during those periods, within which cognitive changes are to be expected, and windowed RQA is potentially a very powerful method to detect such changes.

Running windowed RQA is very similar to running RQA. The difference is simply that the RQA procedure is performed on a subset of the whole time-series. For example, using the sine-wave data with a part of it shuffled in the middle, we can now investigate how the temporal structure of that series changes by chopping it into five adjacent, non-overlapping windows of 40 data points and calculate RQA on each of them. Calculating RQA for the first window, we use the RQA-function as follows: `rqa(time.series=shuffled_sine[1:40], embedding.dim = 2, time.lag = 16, radius=.1)` # run RQA for the first 40 data points of the shuffled_sine variable.

To compute RQA for the following windows, the index of the *shuffled_sine* variable simply needs to be moved to the respective windows, i.e., [41:80], [81:120], [121:160], and [161:200]. The RP of the *shuffle_sine* variable is displayed in Figure 13a, with the gray boxes indicating the subwindows.

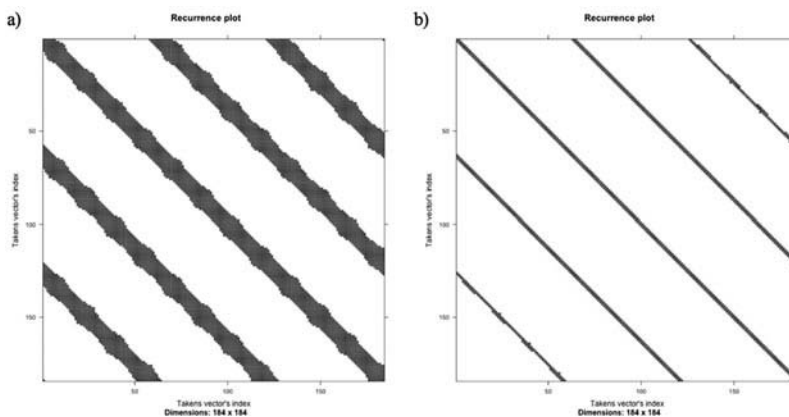


Figure 12. Comparing RPs with different threshold settings. (a) RP of the sine-wave with a comparatively large threshold value of $T = 0.5$. (b) RP of the same sine-wave with a more adequate threshold value of $T = 0.1$.

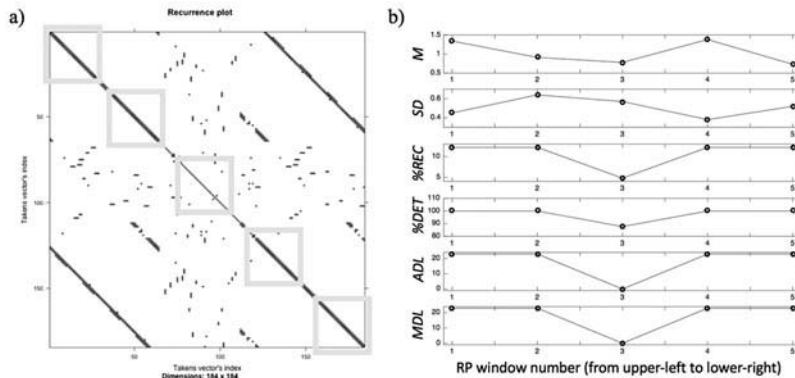


Figure 13. Illustration of windowed RQA. (a) RP of the *shuffled_sine* variable, with gray squares indicating the actual sub-windows from which the statistics are computed. (b) Values for the mean, standard deviation, %REC, %DET, ADL, and MDL for each of the five sub-indows of *shuffled_sine*.

(Note: The windows are not covering the whole original plot around the diagonal, because we are losing 16 of the 40 data points for each subseries due to embedding with a time-delay of 16.) The resulting RQA values of %REC, %DET, ADL and MDL for each window are shown in Figure 13b, together with the mean and the standard deviation of each window of the original time-series.

As can be seen in Figure 13, although the mean and standard deviation are insensitive to the change of dynamics in our time-series, all recurrence measures clearly indicate a dip in recurrence structure where the sine-wave data was shuffled in the middle. Note that %DET is still very high—this is because the *rqa*-function in the package ‘nonlinearTseries’ includes the central diagonal into the computation of %DET (and ADL/Lmean), thus overestimating %DET in the time-series. Hence, using the output *\$LmeanWithoutMain* instead of *Lmean* gives a better picture of the real average diagonal line length. Also, if we are interested in the relative differences between different data sets, the computation of %DET including the main diagonal will not bias these results, as long as the time-series are of approximately the same size.

Using RQA to compare different samples of data sets

So far we have seen how RQA functions for the analysis of a single time-series. However, we are usually not interested in the investigation of a single time-series from a single participant but in comparing samples of data. In that case, treatment of the RQA measures (%REC, %DET, ADL, MDL) is no different from treating other statistical indices, such as the averages of word reading times. That is, RQA measures are obtained for each participant, condition, or trial and are subsequently put into statistical models that are used to analyze other statistics of reading process measures, such as ANOVA, regression, or linear mixed models (e.g., Richter, 2006).

RQA is a time-series (or sequence) analysis technique, so one needs to have multiple data points for each participant to apply the analysis. The more data points, the more reliable RQA measures will be; however, this depends also on the dynamics in the time-series. For word reading times, for instance, several hundred data points need to be collected. With some kinds of data, however, as few as 40 to 50 data points may be sufficient (e.g., RQA applied to the analysis of DNA sequences; e.g., Cao, Tung, Gao, & Qi, 2005). Again, the most important part is to capture the phenomenon one wants to investigate sufficiently and do so at an (average) sampling rate that sufficiently covers the changes in the observable over time.

Comparing sample data sets is usually done by first selecting a single set of parameters for all of them and then to run an RQA on each and obtain RQA measures for each data set. Hence, in our example we would actually proceed through each data set and estimate and note down values for τ

and D . After estimating these parameters for all data sets, we can now pick a single value for τ and D that gives a good characterization of the whole sample, even though the values might differ among some of the data sets. That is, if we want to compare reading times of two groups of readers, we estimate the parameters as described above for each reader individually. Then, from this distribution of embedding parameters, we pick a single value that seems to be good characterization across all participations and conditions and use one value for each, τ and D , across all participants when running the analysis. A concrete example in the next section illustrates the process. The average (rounded to the next integer) is a good value to start with. Selecting the same parameters for all data sets ensures a common standard across the data sets against which the RQA outputs can be evaluated (Wallot et al., 2012). If the sample is relatively diverse, one should select parameter values for τ and D that are somewhat above the average, as RQA is robust against moderate degrees of over-embedding (Webber & Zbilut, 2005).

Similarly, a single threshold parameter τ should be picked for the whole sample: Just as in our example above, one would start again with an arbitrary value for r , run all data sets with this threshold, and then inspect the distribution of percent recurrence $\%REC$ in the sample. As mentioned earlier, $\%REC$ should be low, but not too low, to obtain meaningful RQA results. For interevent-times such as reading times, fixations, or key-strokes features in our examples above, the lowest data set in the sample should have not much less than 1% of recurrence (i.e., $\%REC \approx 1\%$), with most data sets between $\%REC$ 5% and 10%. Hence, the threshold parameter might have to be adjusted a couple of times before a satisfying solution is found. Again, if one uses nominal data, the threshold should be set to 0 or to a tiny value, so that recurrences in all data sets are only based on the repetition of identical instances.

If the data sets cannot be reasonably fitted with a single radius parameter—that is, for some value of τ , some of the data sets are at $\%REC = 100\%$, and at the same time some are at $\%REC = 0\%$ —then instead of keeping the threshold τ constant across all data sets, one can keep the percent recurrence constant across all data sets (e.g., $\%REC = 5\%$ for each data set). Of course, in this case $\%REC$ needs to be omitted from the inferential statistical analysis, as it should be very similar across all data sets in all samples, but the other RQA measures, $\%DET$, ADL , and MLD , can be still analyzed. Also, as a stand-in for $\%REC$, one can instead include analysis of the parameter T that is now different for each data set (Wallot et al., 2012).

If one sees a great dispersion of one of the parameter values or $\%REC$ across the sample, one can also explore the parameter space. This means that the parameters τ , D , and T are systematically varied to check whether the resulting solutions are stable across these variations. A common first exploration of the parameter space would be to run RQA on the lowest, highest, and average values of each τ and D observed across the individual data sets, and pick 3 different values of T that would yield a low ($\%REC \approx 1\%$ to 3%), moderate ($\%REC \approx 5\%$ to 10%), and high ($\%REC \approx 15\%$ to 20%) average amount of $\%REC$ across the sample. If the results converge across the different parameter settings, then one can accept the occurrence of a few individual cases where $\%REC = 0\%$ or 100% .

Another issue that arises when using RQA is how to deal with the multiple outcome variables. In this article I have focused on only four variables ($\%REC$, $\%DET$, ADL , MDL), but other such variables exist or are being developed. In principle, all outcome variables can be dissociated, but although they have more specific meanings in mathematical equation systems or physical systems, this is not so much the case for psychological data and, depending on the task, even less so for interval/noise-type of data. With regard to our four variables, $\%REC$ provides information about the recurrence of individual values given a certain threshold and is hence directly related to the dispersion of values in a data set. The other three measures— $\%DET$, ADL , and MDL —are related to recurrences of sequences of values, capturing larger time-scale patterns in the data that are not necessarily related to dispersion. However, for embedded data, $\%REC$ and dispersion do not necessarily translate into each other directly.

So, although relatively high %REC tells us that many individual instances in a data set are recurrent, %DET tells us something about the probability that multiple of these individual instances are connected. If %DET is high and ADL and MDL are low, this means we have a high but rather homogenous autocorrelation in the time-series. If %DET is low and ADL and MDL are high, this means that the time-series is very heterogeneous and is probably composed of erratic epochs and epochs of high correlation in the data, which might be indicative of qualitative motor-cognitive changes during the recorded performance. If ADL is high and MDL is comparatively low, this could be indicative of bursting behavior, where the observed performance is composed of bursts of activity (such as during type-writing, where within-word key-strokes come in relatively quick succession, with longer gaps between words), but when MDL is high as well, performance might rather exhibit a longer-lasting “settling down” instead of multiple, equally distributed individual patterns of activity. One example could be the buildup of a situation model during reading, where initially incoming information leads to comparatively large changes in the (buildup) of a representation, whereas later information provides only comparatively minor updates of an otherwise established model. The data in [Figure 5a](#) might be an example of such a process, where the beginning of the time-series is marked by relatively few recurrences, but recurrences increase and become more densely connected in the course of reading.

However, the described relations between %REC, %DET, ADL, and MDL should only be seen as a rule of thumb, as an initial suggestion. An inspection of the associated RPs will usually quickly reveal which kind of pattern is actually captured between these for variables. Moreover, as discussed above, in noise-type data, such as key-presses, many of the RQA variables are often highly correlated and do not actually dissociate different dynamics. In such cases one can pick one (or a few) of the variables that seem to capture the composition of the corresponding RP best. Alternatively, the different RQA measures can also be subjected to principle component analysis, and the extracted component(s) can be treated as a dependent (or predictor) variable that captures generalized stability or autocorrelation in the data, depending on the respective factor loadings.

A final, but important question—raised by one of the reviewers of this manuscript—pertains to the application of RQA on performance in the context of linguistic material: For example, it is common practice to investigate, remove, or correct performance (such as word or sentence response times, or fixation durations) for lexical variables such as word frequency, length, or other features of the text material. Whether this should be done with the data before the application of RQA depends on the research question, however. There are some cases where one usually wants to removed unwanted correlations—for example, in the case of self-paced sentence reading times, where removal of sentence length seems warranted because it conflates interesting variables that influence cognitive processing during reading with the simple, spatial extend of the text material. Whether influences of lexical variables on eye movements, for example, are to be removed depends more on what one wants to know: If one is interested in residual variation that might indicate endogenous cognitive coordination during reading beyond the influence of certain (lexical) features, then such correlations should be removed from the data before applying RQA. However, there might be other cases where this is not a good idea. For example, I have proposed (Wallot, 2015, 2016) that recurrence properties can be used to infer how strongly cognitive processes are coupled to linguistic information and that would include correlations between reading times and word frequency, for example. In this case, removing the shared variance between word frequency (and other, similar variables) and reading times would not be sensible.

For further discussions of applications of RQA specifically to reading and writing, see articles by Wallot et al. (2012) and Wallot and Grabowski (2017), respectively. For discussions of potential pitfalls and best practices in recurrence analysis, I recommend a summary article by Marwan (2011). Furthermore, for those who want to get a more mathematical in-depth presentation of RPs and their variants, the overview article by Marwan, Romano, Thiel, and Kurths (2007) is recommended.

Example: Applying RQA reading times of children and adults

In this section we briefly go step-by-step through an RQA analysis of reading self-paced time data, based on the process of parameter selection and analysis presented above. The data are 12 sets of self-paced word reading time-series from four second graders (g2), four sixth graders (g6), and four adults (ad) (undergraduate students), taken from O'Brien et al. (2014) and Wallot et al. (2014). In this study participants were simply asked to read a text word-by-word in a self-paced manner from a computer screen and answer comprehension questions thereafter. All participants read the same, ≈ 1100 words long text. The data can be found here: <https://figshare.com/s/0ae395ae2d4dd1339658>; doi: 10.6084/m9.figshare.4644967.

To load the data into R, one needs to download the 12 data files into a folder and use the `setwd()`-command in R to set the working directory to this folder. Then, the `read.csv()`-command can be used to load each data file into R. For example, the following command would load the data file 'ad_1.txt' and store it in the variable 'data': `data=read.csv('ad_1.txt') # load data file 'ad_1.txt'`.

It is always worthwhile to plot the data first, to get an idea of what the time-series looks like. First, we start with the estimation of the delay and dimension parameters, τ and D . We do this, by applying the average mutual information procedure and the false-nearest-neighbor procedure:

- `mutual(data$ad_1,lag.max = 30) # run average mutual information for data$ad_1`
- `fnn=false.nearest(data$ad_1,m=10,d=1,t=0) # run false-nearest-neighbor analysis`
- `plot(fnn) # plot results of fnn analysis of ad_1`

We chart the resulting values for τ and D and proceed this way with each of the remaining 11 data sets. Although the average delay estimated for each participant is clearly 1, values for the embedding dimension are a little more varied, ranging from 3 to 8. The average across the 12 data sets is 4.17. So, 4 would be a good candidate, but also 5, because small to moderately larger parameters values for noisy data are unproblematic and also fit a little better to some of the higher numbers observed as well. However, the reader is encouraged to play around with different solutions and examine the effects of parameter selection on the results.

Next, we need to find an appropriate value for the threshold parameter T . To that end, we use the `rqa()`-function with our delay ($\tau = 1$) and embedding ($D = 5$) parameters and an initial arbitrary radius/threshold parameter of $T = .1$, e.g., `rqa(time.series=data$ad_1, embedding.dim = 5, time.lag = 1, radius=.1) # RQA for ad_1`

Going through our data sets this way to examine %REC, we see that $T = .1$ yields too little recurrences for some of the data sets, particularly of the younger readers (recurrence rate being lower than 0.01, equaling %REC of lower than 1%). After increasing/changing the parameter a couple of times, it turns out that $T = .4$ is a good setting, because the data set with the lowest recurrences ('g2_3.txt') now scores above a recurrence rate of 0.01, namely 0.012 (i.e., %REC = 1.2%), which is a good minimum, given that some of the other data sets now yield comparatively high percentage of recurrence already. Now, we have all our parameters together: delay $\tau = 1$, embedding dimension $D = 5$, and radius/threshold parameter $T = .4$. Hence, we can now run the `rqa()`-function on each of the data sets with those parameters and note down the resulting RQA-outcome variables for %REC, %DET, ADL, and MDL (the resulting RQA-measures are summarized in Table 2): `rqa(time.series=data$ad_1, embedding.dim = 5, time.lag = 1, radius=.4) # RQA for ad_1`.

Now that we have a table of participants and associated RQA measures for their reading performance, we can simply proceed by analyzing the data with conventional inferential statistics—for example, a one-way ANOVA with the factor Age Group (3 levels: second graders vs. sixth graders vs. undergraduate students). Of course, 12 cases for three cells are too little data points, but for illustrative purposes the results reveal significant (at $\alpha = .05$) effects of increasing RQA measures with age for %REC ($F(1, 10) = 9.55$, $p = .011$), for ADL ($F(1,10) =$

Table 2. RQA measures for each data set.

Data Set	%REC	%DET	ADL	MDL
ad_1.txt	20.73%	92.87%	4.88	37
ad_2.txt	46.57%	97.80%	8.06	57
ad_3.txt	72.42%	99.71%	17.20	192
ad_4.txt	38.34%	96.87%	7.33	74
g6_1.txt	16.39%	91.69%	4.64	37
g6_2.txt	38.06%	97.14%	6.70	70
g6_3.txt	17.28%	92.79%	4.61	33
g6_4.txt	12.50%	90.02%	4.22	29
g2_1.txt	4.63%	84.47%	3.60	19
g2_2.txt	6.08%	82.51%	3.32	15
g2_3.txt	1.24%	76.10%	2.99	10
g2_4.txt	25.03%	95.41%	5.70	43

6.51, $p = .029$), and for *MDL* ($F(1,10) = 5.25.51$, $p = .045$), as well as a marginal effect for *ADL* ($F(1,10) = 4.50$, $p = .060$). For the full results and discussion, see O’Brien et al. (2014) and Wallot et al. (2014).

Conclusion

This article introduced a time-series analysis technique, RQA, provided a tutorial in R on how to run the analysis and summarized previous research on the application of RQA to reading and writing data. RQA provides statistics for complex time-series data (i.e., time-series that exhibit nonstationary fluctuations and transitions), which might be particularly suited to the investigation of naturalistic reading and writing processes that stretch out in time, are based on complicated stimulus material, and stimulate endogenous cognitive change. RQA can be used to detect and quantify such changes, and prior research suggests that such transitions are indicative of cognitive change. Moreover, RQA seems to provide a good measure of reading fluency, relating the concept of reading fluency to flow and smoothness of the change in process measures of reading and writing over time, relating cognitive processing of language to comprehension and understanding. Still, relatively few studies have used RQA to investigate reading and writing, but first results indicate that such investigations yield promising results.

Acknowledgments

I thank Anna Wallot for comments and critical reading of earlier versions of this manuscript and John Holden and two anonymous reviewers for helpful comments and suggestions.

ORCID

Sebastian Wallot  <http://orcid.org/0000-0002-3626-3940>

References

Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. *Frontiers in Psychology*, 5, 771.

Angus, D., Watson, B., Smith, A., Gallois, C., & Wiles, J. (2012). Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PloS One*, 7, e38014.

Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88, 174102-1–174102-4.

Booth, C. R., Brown, H. L., Eason, G. E., Wallot, S., & Kelty-Stephen, D. (2016). Expectations on hierarchical scales of discourse: multifractality predicts both short- and long-range effects of violating gender expectations in text reading. *Discourse Processes*. doi:10.1080/0163853X.2016.1197811

- Cao, Y., Tung, W. W., Gao, J. B., & Qi, Y. (2005). Recurrence time statistics: Versatile tools for genomic DNA sequence analysis. *Journal of Bioinformatics and Computational Biology*, 3, 677–696.
- Coco, M. I. & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Quantitative Psychology and Measurement*, 5, 510.
- Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56, 391–430.
- 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, 1153–1161.
- Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, 5, 973–977.
- Farina, D., Fattorini, L., Felici, F., & Filligoi, G. (2002). Nonlinear surface EMG analysis to detect changes of motor unit conduction velocity and synchronization. *Journal of Applied Physiology*, 93, 1753–1763.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, 40, 145–171.
- Grabowski, J. (2008). The internal structure of university students' keyboard skills. *Journal of Writing Research*, 1, 27–52.
- Graf, R., Nagler, M., & Jacobs, A. (2005). Faktorenanalyse von 57 Variablen der visuellen Worterkennung. [Factor analysis of 57 variables of visual word recognition.] *Zeitschrift für Psychologie*, 213, 205–218.
- Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, 11, 223–237.
- Kuznetsov, N. A., & Riley, M. A. (2010). Spatial resolution of visual feedback affects variability and structure of isometric force. *Neuroscience Letters*, 470, 121–125.
- LeVasseur, V. M., Macaruso, P., & Shankweiler, D. (2008). Promotion gains in reading fluency: A comparison of three approaches. *Reading and Writing*, 21, 205–230.
- LeVasseur, V., Macaruso, P., Palumbo, L. C., & Shankweiler, D. (2006). Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics*, 27, 423–445.
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36, 1404–1426.
- Lyby, M. (2015). Is change in text structure during Expressive Writing Intervention associated with change in psychological symptoms? A re-analysis of essays from an Expressive Writing Intervention with a subsample of Danish breast cancer survivors. Unpublished Master thesis. Aarhus University, Aarhus, Denmark.
- Marwan, N. (2011). How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos*, 21, 1003–1017.
- Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438, 237–329.
- McNerney, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes*, 48, 453–474.
- Mønster, D., Håkansson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & Behavior*, 156, 24–34.
- O'Brien, B. A., & Wallot, S. (2016). Silent reading fluency and comprehension in bilingual children. *Frontiers in Psychology*, 7, 1265.
- O'Brien, B. A., Wallot, S., Haussmann, A., & Kloos, H. (2014). Using complexity metrics to assess silent reading fluency: a cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading*, 18, 235–254.
- Orsucci, F., Giuliani, A., Webber, C., Zbilut, J., Fonagy, P., & Mazza, M. (2006). Combinatorics and synchronization in natural semiotics. *Physica A*, 361, 665–676.
- Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5, 82–87.
- Ramenzoni, V. C., Davis, T. J., Riley, M. A., Shockley, K., & Baker, A. A. (2011). Joint action in a cooperative precision task: nested processes of intrapersonal and interpersonal coordination. *Experimental Brain Research*, 211, 447–457.
- Rhea, C. K., & Kiefer, A. W. (2013). Patterned variability in gait behavior: How can it be measured and what does it mean. In L. Li & M. Holmes (Eds.), *Gait biometrics: basic patterns, role of neurological disorders and effects of physical activity* (pp. 17–44). Hauppauge, NY: Nova Science Publishers.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological science*, 18, 407–413.
- Richter, T. (2006). What is wrong with ANOVA, and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, 4, 221–250.
- Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior*, 34, 99–125.
- Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, 32, 403–408.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103, 877–896.
- 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, 326–332.

- Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 1811–1822.
- Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898, 366–381.
- Teng, D.W., Wallot, S. & Kelty-Stephen, D.G (2016). Single-word recognition need not depend on single-word features: narrative coherence counteracts effects of single-word features that lexical decision emphasizes. *Journal of Psycholinguistic Research*, 45, 1451–1572.
- Thiel, M., Romano, M. C., Read, P. L., & Kurths, J. (2004). Estimation of dynamical invariants without embedding by recurrence plots. *Chaos*, 14, 234–243.
- Wallot, S. (2015). From “cracking the orthographic code” to “playing with language”: Toward a usage-based foundation of the reading process. *Frontiers in Psychology*, 5, 891.
- Wallot, S. (2016). Understanding reading as a form of language-use: A language game hypothesis. *New Ideas in Psychology*, 42, 21–28.
- Wallot, S., & Grabowski, J. (2013). Typewriting dynamics: What distinguishes simple from complex writing tasks? *Ecological Psychology*, 25, 267–280.
- Wallot, S., & Grabowski, J. (2017). A tutorial introduction to Recurrence Quantification Analysis (RQA) for keylogging data. In E. Lindgren & K. Sullivan (Eds.), *Observing writing: logging handwriting and computer keystrokes*.
- Wallot, S., Hollis, G., & van Rooij, M. (2013). Connected text reading and differences in text reading fluency in adult readers. *PloS One*, 8, e71914.
- Wallot, S. & Kelty-Stephen, D. (21 March, 2016). Temporal structure in reading times predicts switching between different reading tasks—An inquiry into the language game hypothesis (LGH) of reading. Paper presented at the Tagung Experimentell arbeitender PsychologInnenen 2016, University of Heidelberg, Heidelberg, Germany.
- Wallot, S., Lyby, M., & van Rooij, M. J. D. (2015). Recurrence Quantification Analysis as a general-purpose tool for bridging the gap between qualitative and quantitative analysis. In G. Huber (Ed.), *Qualitative psychology nexus* (Vol. 13, pp. 69–87). Tübingen: Center for Qualitative Psychology.
- Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, A., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1745–1765.
- 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, and C. Westbury (Eds.), *Methodological and analytical frontiers in lexical research* (pp. 395–430). Amsterdam: John Benjamins.
- Webber, C. L., Jr., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76, 965–973.
- Webber, C. L. Jr., & 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. 26–96). Retrieved from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>.
- Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia*, 62, 100–119.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386–397.