On the pros and cons of using temporal derivatives to assess brain functional connectivity

Jeremi K. Ochab a,⁎, Wojciech Tarnowski a, Maciej A. Nowaka a, Dante R. Chialvo b

a Marian Smoluchowski Institute of Physics and Mark Kaz Complex Systems Research Center, Jagiellonian University, ul. Lojasiewicza 11, PL30-348, Krakow, Poland
b Center for Complex Systems & Brain Sciences (CEMSC3), Escuela de Ciencia y Tecnología, Universidad Nacional de San Martín, Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), 25 de Mayo 1169, San Martín, (1650), Buenos Aires, Argentina

A R T I C L E  I N F O

Keywords:

fmri
Resting state networks
Functional connectivity

A B S T R A C T

The study of correlations between brain regions is an important chapter of the analysis of large-scale brain spatiotemporal dynamics. In particular, novel methods suited to extract dynamic changes in mutual correlations are needed. Here we scrutinize a recently reported metric dubbed "Multiplication of Temporal Derivatives" (MTD) which is based on the temporal derivative of each time series. The formal comparison of the MTD formula with the Pearson correlation of the derivatives reveals only minor differences, which we find negligible in practice. A comparison with the sliding window Pearson correlation of the raw time series in several stationary and non-stationary set-ups, including a realistic stationary network detection, reveals lower sensitivity of derivatives to low frequency drifts and to autocorrelations but also lower signal-to-noise ratio. It does not indicate any evident mathematical advantages of the proposed metric over commonly used correlation methods.

1. Introduction

An important challenge and a very active area of research in the neuroimaging community is the study of correlations between brain regions, as a central point in the analysis of large-scale brain spatiotemporal dynamics. Usually the correlation measures are computed between several thousand time series—the BOLD (“blood-oxygen-level dependent”) signals—covering the entire brain, and averaged over a period of tens of minutes. However such evaluations fall short of characterizing the highly dynamic changes occurring in the brain (Chang and Glover, 2010; Preti et al., 2017; Yu et al., 2015; Gonzalez-Castillo et al., 2014; Thompson and Fransson, 2016) and, consequently, the current emphasis shifted to the study of time-varying correlations.

In that line, a recent report (Shine et al., 2015) introduces a metric dedicated to investigation of dynamic functional connectivity (DFC) in fMRI time series data—dubbed Multiplication of Temporal Derivatives (MTD)—based on the temporal derivative of each BOLD time series. The authors used this measure on a three-part experiment claiming that it demonstrates “the ability of this novel metric to calculate dynamic and stationary functional connectivity structure in both real and simulated data” (Shine et al., 2015). Naturally, there are numerous other DFC methods available, based on sliding-window Pearson correlation (e.g. (Chang and Glover, 2010; Hutchison et al., 2012), clustering (Liu and Duyn, 2013) or temporal ICA (Smith et al., 2012) to name just a few. Many of them can be formulated in a general conceptual framework described by Thompson and Fransson (2018)—most notably MTD method and weighted Pearson correlation to which it bears resemblance. It is of particular importance to explore and determine in which tasks these methods perform well and what properties of BOLD time series they rely on.

These notes are dedicated to carefully inspect the mathematical grounding of the MTD measure and revisit some of the scenarios for which the metric is intended. We will first discuss the case of stationary processes, and then inspect the non-stationary cases. The paper is organized as follows: the next section contains the mathematical definition of the MTD measure side by side with the mathematical expression for the Pearson correlation coefficient. To develop some intuition, Section 3 provides extremely simple examples allowing comparison of the expected results for Pearson correlation in the raw time series and its time derivative. Section 4 presents a formal framework to predict the expected behavior of the correlations for the case in which a raw time series is compared with its temporal derivative. In Section 4.1 we inspect a simple non-stationary example which lends support to the analytical expectations for the two time series (raw and derivatives). The analytical...
framework is further contrasted, in Section 5, with the results of analyzing resting state BOLD time series. In Section 6 we use surrogate data from BOLD time series to determine the behavior of the derivatives to a sudden change in covariance. Finally, in Section 7 the performance of the MTD approach in inferring the underlying network connectivity is examined. The paper closes with conclusions and a brief summary of the results. Further details are condensed in the Appendix section with the appropriate derivations.

2. Correlations of first-order derivatives

Let us start by recalling how the authors in (Shine et al., 2015) define Multiplication of Temporal Derivatives:

\[ ds_i = s_{i+1} - s_i \]

\[ MTD_{ij} = \frac{ds_i ds_j}{\sigma_i \sigma_j} \]

\[ SMA_{ij} = \frac{1}{2w+1} \sum_{t=-w}^{t+w} MTD_{ij} = \frac{1}{2w+1} \sum_{t=-w}^{t+w} \frac{ds_i}{\sigma_i} \frac{ds_j}{\sigma_j} \]

where \( 2w + 1 \) equals to the number of samples considered in a temporal window \( [t-w, t+w] \), \( s_i \) is an \( i \)-th time series, and \( \sigma_i \) is the standard deviation.

Fig. 1. Simple examples of raw time series \( s_t \) (black) and its series of derivatives \( ds_t \) (red) for a Gaussian (top), a sinusoid (middle), and a typical brain BOLD (bottom) time series. The diagrams on the right plot the time series against each other.

Fig. 2. Analytical and numerical results for the correlations of a AR(1) process of two nodes with a range of coefficients (the auto-coeff. \( a_1 \), and the cross-coeff. \( a_2 \)). In all graphs \( r_{ij} \) corresponds to the Pearson correlation for the raw (green circles) and derivative (blue squares) time series. Dashed lines represent the asymptotic analytical expectation for \( r_{ij} \), and symbols and error bars correspond to the mean and standard deviations of the numerical results. Upper panels show correlation of entire series, lower panels show sliding window correlation (i.e., SWPC of raw series and MTD). Time series length \( T = 1000 \); \( N = 1000 \) realizations; window length \( 2w + 1 = 21 \).
deviation of the entire $d_s t$ series. We call $s_t$ the raw time series as opposed to the series of derivatives $d_s t$. Usually, the series $d_s t$ is called inter-changeably increments, finite differences, temporal derivatives or differentiated time series. While the first two are terminologically more adequate, for consistency with the established name of MTD we continue to use temporal derivatives throughout. Note that (1) is a forward difference with a unit step time, but other choices are also possible.

Now, let us reflect on how these definitions relate to correlation coefficients. For any two time series $s_t$ and $s'_t$, where $i, j \in \{1, 2, ..., N\}$ are indices numbering the series and $t \in \{t - w, t - w + 1, \ldots, t + w - 1, t + w\}$ is a time index in a window around time $t$, the sample estimator of Pearson correlation coefficient for the given time window is defined as

$$r_{w} = \frac{\sum_{i=t-w}^{t+w} (ds_i - \bar{d}_s) (ds'_i - \bar{d}'_s)}{\sqrt{\sum_{i=t-w}^{t+w} (ds_i - \bar{d}_s)^2 \sum_{i=t-w}^{t+w} (ds'_i - \bar{d}'_s)^2}}$$

$$= \frac{1}{2w+1} \sum_{i=t-w}^{t+w} \frac{ds_i - \bar{d}_s}{\sigma_i} \frac{ds'_i - \bar{d}'_s}{\sigma'_i},$$

(4)

where $\bar{d}_s = \frac{1}{2w+1} \sum_{i=t-w}^{t+w} ds_i$ is the sample mean and $\sigma_i$ denote standard deviations of the time series $d_s t$ in the given time window $[t - w, t + w]$. Centering around the mean and dividing by standard deviations is necessary for the coefficient to be invariant under linear transformations ($a + b d_s t$, with constants $a$ and $b$).

The form of equations (3) and (4) are visibly similar. The difference is that the derivatives in SMA are not centered and that the standard deviations $\sigma_i$ in (3) are computed over the whole series and not just over the time window as $\sigma_i$ in (4). As regards centering, the mean $\bar{d}_s = \frac{1}{2w+1} \sum_{i=t-w}^{t+w} ds_i - \frac{\bar{d}}{2} \frac{\bar{d}}{2} \to 0$ for large window sizes; the standard deviation of the (window) sample also converges to the standard deviation of the continuous, differentiable signal.

In other words, asymptotically for large window sizes the SMA of MTD (Shine et al., 2015), i.e., the moving average of multiplication of temporal derivatives equates with the sliding-window Pearson correlation (SWPC) of temporal derivatives (not to be confused with the SWPC of temporal derivatives equates with the sliding-window Pearson correlation). In the general notation used by Thompson and Fransson (2018), $Y = R(U(X); W)$—where $X$ are raw data series, $W$ are weight vectors, $U$ is a transform of the data, $R$ is a relation function, and $Y$ is the resulting estimate of DFC—SMA of MTD can be expressed with the temporal derivative (1) standing for $U$ and Pearson correlation standing for $R$. While we believe, the “SWPC of derivatives” is a more informative term than MTD, hereafter we do not use it in order to avoid terminological confusion.

As also mentioned in (Shine et al., 2015), differencing has the high-pass filtering effect preferable in some situations. Indeed, differencing is a well-known method for reducing some time series to stationary ones (Box et al., 2015, Ch. 6.2), which one could expect to show no or, in real data, at least less dynamics. Therefore, a valid question is how is this method different from the sliding window Pearson’s correlation coefficients referred to therein or, from another angle, what information do derivatives contain that the raw series do not? To unravel this issue, in the next section we will first scrutinize the simplest examples of time series, allowing a comparison with the Pearson correlation expected for a raw time series and its temporal derivatives. Again we remark that the stationary cases will be treated first to later analyze the relevant case of non-stationarity.

3. For the sake of simplicity, let us consider how the two correlation measures behave in simple cases. First, in the case of signals which are Gaussian white noise (see Fig. 1) let us assume that $s_t = \xi_t$ where $\xi_t$ are independent and identically distributed random variables, and the $s_t$ are the outcomes of these variables (random variates). Let us remind that independence of random variables $X, Y$ is equivalent to the statement that the expectation value factorizes $E[f(X)|g(Y)] = E[f(X)]E[g(Y)]$ for any functions $f, g$. In the case of Pearson correlation, the centered time series $E[s_t - \bar{s}] = 0$ and thus, thanks to the independence between $s_t$ and $s'_t$, the correlation coefficient (4) is zero for $i \neq j$. It follows that since $d_s t$ is a sum of two random variables, its variance is twice the variance of $s_t$. Thus, in all noisy time series we can expect the signal-to-noise ratio of the derivatives to be lower than that of the raw series.

Now let us assume that instead of a discrete time series $s_t$, we have a continuous, differentiable signal $s(t) = \sin(t + \phi_1)$, which has zero mean and $\sigma^2 = 1/2$. For signals $s(t)$, $s'(t)$ differing only by their phases $\phi_1, \phi_2$, the correlation coefficient equals $r_{w} = \cos(\phi_1 - \phi_2)$. The differentiated series $d_s(t) = \cos(t + \phi_1)$ have the same mean and variance and correlation coefficients. Introducing arbitrary frequencies have no effect as well. Thus, we expect similar results for the both Pearson correlation measures for signals that are sinusoidal. For a discrete signal, the amplitude of derivatives decreases linearly with increasing sampling rate, see Fig. 1 (middle panel). Computationally, differences between the two correlation measures will only appear from an artifact when the length of the sliding windows is comparable to the period of the sinusoidal oscillation. In such case, there is an asymmetry, as the sine waves will be cut at different positions depending on their phase and frequency. For the sake of intuition, the bottom panel of Fig. 1 depicts a BOLD time series and its derivative, which illustrate the phase shift mentioned above for the case of sinusoidal signals. Note that the derivatives produce series one data point shorter than the raw time series, as visible in the figure.

Before proceeding further let us recall that the main claim of (Shine et al., 2015) was that the cross-correlation of the derivative time series is more informative of brief correlation changes.

4. Autoregressive models

To revisit and understand the results under scrutiny, in this section we study the case of two AR(1) processes for which we can manipulate their interactions and compute the expected correlations measures for both the raw and derivative time series. Despite the enormous differences between the properties of the BOLD signal and a simple AR(1) process, its analysis can bring about some general understanding on what to expect from the proposed correlation metric of derivatives. The two time series are simulated as follows:

$$x_{1,t} = a_1 x_{1,t-1} + a_2 x_{2,t-1} + \xi_{1t}, x_{2,t} = a_1 x_{2,t-1} + a_2 x_{1,t-1} + \xi_{2t},$$

(5)

where $\xi_{1t}$ are uncorrelated. The range of parameters is limited by $|a_1| + |a_2| < 1$ to ensure weak stationarity. The model (5) is a special case of VAR(q) processes—in the Appendices Appendix A, Appendix B, Appendix C we show how one can derive analytically and compute Pearson correlation (and correlation of the derivative) of such a process knowing its parameters. Thanks to that, we can predict the average behavior of SWPC and MTD for a range of parameters within that model, as well as we can reverse-engineer the real data, designing a model that exhibits specific Pearson correlations.

Notice that $a_1$ and $a_2$ above represent the auto-interaction and the cross-interaction coefficients, respectively, which can be estimated in this linear case by the correlation coefficients. Thus, in the jargon of (Shine et al., 2015) $a_1$ and $a_2$ are the values representing the “ground truth” which the numerical methods of functional connectivity shall predict.

Fig. 2 (upper panels) shows the numerical results (symbols and error bars) for the asymptotic correlations of the process of two nodes in (5) as well as the analytical results (dashed lines) provided in Appendices.
We consider the same scheme of (5), with realistic scenario, allowing us to tune the autocorrelations of the signal. The cross-interaction coefficient from null to positive Pearson correlation (from 0.1 up to 1) is essentially Gaussian signals undergoing an instantaneous change: 

The analytical result for the full parameter range is shown in Fig. 3.

Appendix A, Appendix B, Appendix C. The results show that in the case of fluctuations modelled by an AR(1) process the Pearson correlation of its derivative exhibits a negative sign. Moreover, depending on \( a_1 \) and \( a_2 \) it can be weaker or stronger than the correlation of its raw time series. A realistic, with high values of autocorrelation, is further shown in Sections 5-6. The analytical result for the full parameter range is shown in Fig. 3.

Similar conclusions can be drawn from computing sliding windows of the Pearson correlation of both the raw time series and its derivative as shown in lower panels of Fig. 2. The only difference with the results in the preceding section, and specifically from Figs. 2–3, one can see that the size of the change in correlations \( r_{ij} \) heavily depends on the parameters \( a_1, a_2 \). Also the relative size of \( r_{ij} \) between raw series and derivatives depends on them. Additionally, mark the standard deviations in Fig. 4:

Then we compute correlation measures over the raw and derivative time series and attempt to predict at which time step the change occurs. Fig. 4 shows an example of the typical results obtained. Mark that for the derivatives, we compare MTD (3) and SWPC (4), showing that despite the difference in centering and standardization they behave almost the same even for a small window size.

It is visible that correlations computed over the raw or the derivatives accurately detect the change in the coefficient from \( a_2 = 0 \) to \( a_2 = 0.5 \) as soon as the sliding window reaches the transition point. In fact, the correlation values after the transition can be read already from Fig. 3 by looking at the corresponding parameter coordinates. Moreover, from the preceding section, and specifically from Figs. 2–3, one can see that the size of the change in correlations \( r_{ij} \) heavily depends on the parameters \( a_1, a_2 \). Also the relative size of \( r_{ij} \) between raw series and derivatives depends on them. Additionally, mark the standard deviations in Fig. 4: for \( a_2 = 0 \) they are comparable between the methods; for \( a_2 = 0.5 \), MTD has the largest spread, followed by SWPC of raw series and SWPC of derivatives having the best precision. Lastly, while the raw series does not allow significant detection of change in cross-coefficient for small auto-coefficient, for high \( a_1 \) the detection is possible even sooner than for derivatives, owing to the steeper early slope during the transition.

This simple example demonstrates how strongly correlation metric of both raw series and derivatives depends on the model of the signal, which raises doubts about any general advantages of one method over the other to detect such non-stationary effects. In some scenarios,
The dots in Fig. 5 represent the cross-correlation of consecutive pairs of time series with neighboring autocorrelation values. 998 data sets. After that the cross-correlations were computed between the autocorrelation of each time series and used to sort the order to make a comparison of auto and cross-correlation properties, the entire human cortex in progressive moving average (ARMA) models (Brockwell and Davis, 2016), Furthermore, we extend the analysis of each time series to its autoregressive process with two polynomials, one for the auto-regression and the second for the moving average.

The properties discussed so far for the raw and derivative time series can be further investigated in the setting of simulated sets that contain the exact same correlation properties as the BOLD data. For that purpose we simulate time series which mimic closely the covariance and spectral properties of the BOLD data. This approach was used recently by Lauermann et al. (2016) to study the contribution of various sources of non-stationarity. The starting point is the BOLD data set from an individual subject. After applying the bandpass filters usual in most fMRI studies, two estimators are computed: the average power spectrum and the covariance matrix. After that we create a time series of random Gaussian data of size equal to the real BOLD data set. Subsequently, the spectral content is matched by multiplying these random time series by the two-sided average power spectrum obtained from the real data. Finally, the time series are projected onto the eigenvectors of the covariance matrix calculated from the real data. In this way multivariate data sets—of arbitrary length—are generated, which are stochastic realizations of the chosen BOLD data sets, with identical covariance structure and mean spectral content.

Using this simulation method, similarly as in Sec. 4.1, we test the scenario of a sudden change in the covariance, considered in the work being discussed (Shine et al., 2015). Fig. 6 illustrates an example using the same BOLD data used in Fig. 5 consisting of 998 time series. Here are plotted the covariances of a pair of time series, both for the raw and the differentiated ones. Up to the step indicated by the dashed line, the series corresponds to a realization of the selected original data (same spectra however, one or the other method might have an edge—like MTD in the particular model of head motion originally reported in (Shine et al., 2015). The following sections, and in particular Fig. 6, test these observations in a much more realistic, data-driven fashion.

5. Correlation properties of BOLD signals and its fitted ARMA model

Now we turn to study the correlation properties of real BOLD brain signals estimated for both the raw time series and its derivative. Furthermore, we extend the analysis of each time series to its autoregressive moving average (ARMA) models (Brockwell and Davis, 2016), an approach which is often used to explore the statistical properties of time series. This approach allows for a description in terms of a stochastic process with two polynomials, one for the auto-regression and the second for the moving average.

In Fig. 5 we show the results of such an approach informed by brain resting state BOLD data. These time series correspond to closed-eyes resting state fMRI, 240 vol with 2.5 s TR amounting to 10-min resting state BOLD signals estimated for both the raw BOLD and open squares for derivatives; red big circles denote binned averages for raw and squares for derivatives, with horizontally equidistant binning). Right Panel: Dashed lines are analytical expectations of cross-correlations for the fitted model given its autocorrelation. The symbols with errors bars correspond to mean cross-correlation values and standard deviations, respectively, from simulation of the fitted ARMA model, where circles denote results for the raw series and squares those from differentiated series. The model consists of two coupled ARMA(1,1) time series with three $2 \times 2$ coefficient matrices in total ($A$ for AR, $B$ for MA and $\Sigma$ for noise). The parameters $A_{11} = A_{22} = a_1$ were fixed by the autocorrelations of the BOLD pairs (green circles data points); the other parameters were fitted by least square differences between the analytical lines and the data points (both raw and differentiated).

The reason for scrutinizing the dependence of cross-correlation on autocorrelation is the following: in general, one would like to know, whether (linear) Pearson correlations of raw or derivative time series are informative of interactions between brain regions that produced the series. In Sec. 4 we have shown that it is not only the interactions (modelled by cross-coefficient $a_2$) but also the auto-coefficient $a_1$ responsible for autocorrelations that might strongly influence what is measured by the cross-correlations. Knowing that, it is worth checking how different methods measuring cross-correlations depend on autocorrelations in real data.

A straightforward visual inspection of Fig. 5 immediately reveals that, on average, most pairs of derivative time series (square symbols) exhibit a weak and constant value of the correlation (as already suggested by the results in Figs. 2 and 3), while the raw data (circle symbols) shows an increasing trend of cross-correlations for increasing autocorrelations of a given pair. This tendency can to an extent be reproduced with a VARMA(1,1) model, both simulated and analytical, as presented on the right panel of Fig. 5. The model is defined by two coupled ARMA(1,1) time series with three $2 \times 2$ coefficient matrices in total ($A$ for AR—see also Appendix A, Appendix B for MA and $\Sigma$ for noise).

These observation can be interpreted in several ways. First, correlations can be misleading in the case of raw series, because they might only be a result of autocorrelation. On the other hand, the best fit ARMA(1,1) model, although not fully adequate, cannot explain the whole range of $r_g$ by manipulating $a_1$ when the interactions (i.e., $A_{12}$ and $A_{21}$ matrix elements) are kept constant. It suggests that $r_g$ of raw series comes at least partly from the functional interactions. Secondly, the $r_g$ of derivatives can be explained more easily by the simple model, which might mean they are less informative. How generic this behavior is, however, remains to be studied analytically. Finally, the $r_g$ of raw BOLD series and of derivatives are uncorrelated, which means that SWPC and MTD methods can produce either complementary or contradictory results that are hard to interpret at the current level of understanding. More subtle temporal dynamics can be unraveled using the spectral analysis of cross-correlation matrices (Burda et al., 2010; Livan and Rebecchi, 2012; Nowak and Tarnowski, 2017) and the approach discussed in the next section.

6. Further testing with surrogate data sets with identical covariance and spectral properties as BOLD
and covariance) having high cross-covariance, approx. 0.8, and further on to a sudden change of non-diagonal entries of the covariance matrix to 0.2 we introduced in the simulation. In a similar manner, we also introduced simultaneous change in both cross- (again to 0.2) and auto-covariance (to 0.4). It can be seen that estimators of both raw series and derivatives track the change. Note the relatively smaller amplitude of covariance of derivatives, something expected by definition, as follows from discussion in Sec. 3. Next, the spread of MTD metric is much greater than that of SWPC, as illustrated by the shaded strips of standard deviations. At the same time, the intuitions gained from autoregressive processes, as demonstrated in Fig. 4, appear to hold. The autocovariance does affect the two metrics differently: MTD remains relatively insensitive to it, while it does make the step in SWPC smaller. It should be emphasized that such simultaneous amplitude modulation does not necessarily arise from any communication of the two ROIs. Since neither method is able to tell the difference, any preference in such a case is debatable.

7. Detection of ground truth network from realistic BOLD simulations

An important claim in (Shine et al., 2015) is the apparent benefit of using the MTD approach to estimate the stationary connectivity structure of a functional network. They compared the performance of the correlation of derivatives against existing methods studying a well characterized fMRI dataset as well as a previously published gold-standard simulated data set (Smith et al., 2011) obtained from FMRIB (http://www.fmrib.ox.ac.uk/analysis/netsim). For the sake of comparison with the original result, we use the exact same setting, irrespective of its suitability to test for a time-varying connectivity. The data set is simulated BOLD signals with known network structure, thus enabling evaluation of a range of connectivity models and comparing the results with the “ground truth” network structure. Briefly, this dataset consists of 28 simulations of BOLD data in 50 realizations (TR = 1.5–3 s; 200–1000 individual time points; 5–50 nodes; and differing levels of noise and hemodynamic response function variability). Each simulated dataset was created using an fMRI forward model based on dynamic causal modeling (DCM; Friston et al., 2003), combined with a non-linear balloon model (Buxton et al., 1998) to simulate vascular dynamics (see Smith et al., 2011, for details of the simulation).

Here we restrict our attention to the network labeled “sim4”, which simulates 50 BOLD signals. The simulated network comprises 10 regular modules interconnected by a few links, forming a typical small-word graph. There is a total of 50 nodes interconnected by 60 positive off-diagonal interactions, 40 of them corresponding to nearest neighbors (link weights are 0.4 ± 0.03; mean ± s.d.). There are also 50 negative (diagonal) self-interactions which determine the characteristic time scale for the dynamics. The data set contains 50 stochastic realizations which meant to represent fMRI records of different human individuals.

We proceed to use the same dataset and check how well the methods perform in predicting from the time series the underlying graph. Specifically, we check how well the correlation matrices obtained from both the raw BOLD dataset and its derivative describe the ground-truth network. We used partial correlations (instead of Pearson’s) since that was the optimal method reported in the original Smith et al. (2011) report.

To gauge each method we used the receiver operating characteristic curve (ROC; Fawcett, 2006), which benchmark specificity and sensitivity as a function of a given parameter. To determine whether a connection between two nodes is predicted or not we choose a decision threshold of...
1.75 σ at the entries \((i,j)\) of the partial correlation matrices, interpreting any value larger than that as a connection between such nodes. The presence of each link predicted in this way is compared with the respective entry in the adjacency matrix of the sim4 network, resulting in a false positive or a true positive event. The same procedure is applied for a range of time series lengths and repeated for the raw time series (green points) and the derivatives (blue points).

Fig. 7 illustrates the results: there is a family of curves corresponding to various lengths of the considered time series. The shortest data \((T = 200\) samples) give the less confident results and the longest \((T = 1000\) samples) a very good estimation of the network connections. The area under the curve (AUC), plotted in the right panel of Fig. 7, is a good estimate of the performance of the method, where a value of 1 corresponds to a perfect prediction and a value of 0.5 is equivalent to chance. The results in this figure clearly show that the estimates based on the derivatives perform worse than the ones using the raw BOLD time series, suggesting that MTD offers no advantage over standard methods.

8. Conclusions

In this paper we have analyzed the basis of a metric dubbed Multiplicative Temporal Derivatives recently proposed as a novel way to determine changes in functional correlations between regions of interest in the brain. Even though we focused on properties of only two DFC methods, MTD together with sliding-window Pearson correlation, many others are available and it is noteworthy that there are efforts (Thompson et al., 2017) to systematically assess their performance.

A formal comparison of the MTD formula with the Pearson correlation of a time derivative reveals two differences, namely that derivatives are not centered in MTD and that their variance is computed over a range of time steps. The matrices \(A^{(k)}\), which are of size \(N \times N\) (in general non-symmetric) with elements \(A_{ij}^{(k)}\) and \(\xi_{it}\) are Gaussian random variables which are uncorrelated in time, i.e., \(\xi_{it}\) and \(\xi_{jt}\) are independent if \(t \neq t'\). General VAR models allow for correlated noise with true covariances \(E[\xi_t^T \xi_{t+k}] = \Sigma_k\), where \(\Sigma\) is symmetric positive definite. By \(\langle \cdot \rangle\) we denote the sample average of the time series. For instance, in this case the estimator of the true covariance matrix reads \(\langle \xi_t^T \xi_{t+k} \rangle = \frac{1}{T-k-1} \Sigma_k\), where \(T\) is the length of the series. For infinite time series length the true covariance matrix and its estimator are identical.

Acknowledgements

Work conducted under the auspice of the Jagiellonian University-UNSAM Cooperation Agreement. JKO thanks Valeria Pattacini from the Office of International Relations (ORI) of Universidad de San Martin, UNSAM, (Argentina) for facilitating his visit and the UNSAM’s hospitality. JKO was supported by the National Science Centre (Poland) grant DEC-2015/17/D/ST2/03492. WT appreciates the financial support from the Polish Ministry of Science and Higher Education “Diamond Grant” 0225/DIA/2015/44. Authors thank Steve M. Smith and Mark Woolrich (Imperial College, London) for sharing information on the Netsim package. DRC was supported in part by CONICET (Argentina) and Escuela de Ciencia y Tecnología, UNSAM. The authors are grateful to Ignacio Cifre for valuable discussions.
From this point on our aim is to calculate the Pearson correlation coefficients for VAR(1) process and its temporal derivative. We define the time-lagged cross-covariance matrix as

\[ C(\tau)_{ij} := \langle x_i x_{j,t+\tau} \rangle = C(-\tau)_{ij}. \]  

For brevity we denote \( C \equiv C(0) \) and \( A \equiv A^{(1)} \). For simplicity of calculation let us start by deriving \( C(1) \).

\[
C(1)_{ij} = \langle x_i x_{j,t+1} \rangle = \left\{ x_i \sum_{k=1}^{N} A_jk x_k + \xi_{j,t+1} \right\} = \sum_{k=1}^{N} A_jk \left\{ x_i x_k \right\} + \langle x_i \xi_{j,t+1} \rangle \tag{A.3}
\]

The first equality is the definition, in the second one we substitute (A.1) for \( x_{j,t+1} \), in the third we expand the multiplication. The last term vanishes because the only case in which it could give a contribution is when \( x_i \) contained \( \xi_{j,t+1} \)—but according to the definition (A.1), it does not. Moreover, we recognize the first term as the equal-time covariance matrix. Therefore

\[
C(1)_{ij} = \sum_{k=1}^{N} A_jk A_{ik} \tag{A.4}
\]

The results can be presented in a clearer fashion with the use of matrix notation

\[ C(1) = CA^T, \tag{A.5} \]

where the sum over matrix elements has been replaced with matrix multiplication, and the reversed order of indices of matrix elements \( A_{jk} \) has been replaced with the transpose of the matrix \( A \) denoted by the superscript T. Such notation provides a considerably more manageable equations.

It is worth noting that the assumption \( \langle x_i \xi_{j,t+1} \rangle = 0 \) is true only for long time series \( T \to \infty \). The empirical sample average might yield a non-zero result. This might be partly responsible for the discrepancy between analytical values and simulations observed in Fig. 2.

The equal-time covariance matrix can be obtained by repeating the above steps:

\[
C_{ij} = \langle x_i x_j \rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ik} A_jk \langle x_{i-1} x_{j-1} \rangle + \langle \xi_{i} \xi_{j} \rangle = \sum_{i=1}^{N} A_{ik} A_jk + \Sigma_{ij},
\]

where we already omitted the substitution of (A.1), and by virtue of stationarity of the time series \( \langle x_i x_j \rangle = \langle x_{i-1} x_{j-1} \rangle \). In matrix notation

\[ C = ACA^T + \Sigma \tag{A.6} \]

We solve this equation in App. Appendix C.

**Appendix B. Time derivative of VAR(1)**

Let us rewrite (A.1) for \( q = 1 \) in matrix notation

\[ x_t = Ax_{t-1} + \xi_t, \tag{B.1} \]

where \( x_t = [x_t, \ldots, x_N] \) and \( \xi_t = [\xi_t, \ldots, \xi_N] \) are \( N \)-element vectors. By temporal derivative we mean here taking finite forward differences

\[ dx_t = x_{t+1} - x_t = A(x_t - x_{t-1}) + \xi_{t+1} - \xi_t. \tag{B.2} \]

To obtain its equal-time cross covariance matrix—which we denote by \( \dot{C} \)—we follow the same steps as outlined in the previous section, which leads to

\[ \dot{C} = 2C - C(1) + C(-1) \tag{B.3} \]

\[ = (1 - A^T)C + (1 - A)C, \tag{B.4} \]

where the first equality is general, while the second one holds for VAR(1) process only (cf. A.5); \( I \) denotes the \( N \times N \) unit matrix. Thanks to the connection between covariances of the derivatives and the raw time series, now we only need to solve (A.6), which is described in the next section. Although more convoluted, equations for \( C \) and \( \dot{C} \) can be obtained also for higher VAR orders, as well as for VARMA models.

**Appendix C. Vectorization**

Equation (A.6) provides us with a way to compute \( C \) (in the limit \( T \to \infty \)) knowing \( A \) and \( \Sigma \). The procedure is simpler if we utilize vectorization (see, e.g., Laub, 2005), which means stacking columns of a matrix into a single column vector. In the vec notation the equation takes the form

\[ \text{vec}(C) = \text{vec}(AC^T) + \text{vec}(\Sigma) \tag{C.1} \]
where in the second line we used the property \( \text{vec}(A_1A_2A_3) = (A_1^T \otimes A_1)\text{vec}(A_2) \), which expresses the matrix multiplication as a linear transformation on matrices with the help of the Kronecker product \( \otimes \). Solving for \( \text{vec}(C) \) is now straightforward:

\[
\text{vec}(C) = (1 \otimes I - A \otimes A)^{-1}\text{vec}(\Sigma). 
\]  
(C.3)

When one of the eigenvalues of \( A \) is equal to one, the matrix in the parenthesis is not invertible (which may be an onset of non-stationarity of time series).

References


