A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis

Richter, C.G.; Thompson, W.H.; Bosman, C.A.; Fries, P.

DOI
10.1016/j.neuroimage.2015.04.040

Publication date
2015

Document Version
Final published version

Published in
NeuroImage

Link to publication

Citation for published version (APA):
Richter, C. G., Thompson, W. H., Bosman, C. A., & Fries, P. (2015). A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis. *NeuroImage*, *114*, 57-70. https://doi.org/10.1016/j.neuroimage.2015.04.040

NeuroImage 114 (2015) 57–70

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A jackknife approach to quantifying single-trial correlation between covariance-based metrics undefined on a single-trial basis

Craig G. Richter [a,b,*], William H. Thompson [a,c], Conrado A. Bosman [d,e], Pascal Fries [a,e]

[a] Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, 60528 Frankfurt, Germany
[b] Laboratoire de Neurosciences Cognitives, École Normale Supérieure, 75005 Paris, France
[c] Department of Clinical Neuroscience, Karolinska Institute, 171 76 Stockholm, Sweden
[d] Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, 1098 XH Amsterdam, Netherlands
[e] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525 EN Nijmegen, Netherlands

ARTICLE INFO

Article history:
Received 27 June 2014
Accepted 20 April 2015
Available online 24 April 2015

Keywords:
Single-trial correlation
Jackknife
Functional connectivity
Coherence
Granger causality
Spectral analysis

ABSTRACT

The quantification of covariance between neuronal activities (functional connectivity) requires the observation of correlated changes and therefore multiple observations. The strength of such neuronal correlations may itself undergo moment-by-moment fluctuations, which might e.g. lead to fluctuations in single-trial metrics such as reaction time (RT), or may co-fluctuate with the correlation betwe'en activity in other brain areas. Yet, quantifying the relation between moment-by-moment co-fluctuations in neuronal correlations is precluded by the fact that neuronal correlations are not defined per single observation. The proposed solution quantifies this relation by first calculating neuronal correlations for all leave-one-out subsamples (i.e. the jackknife replications of all observations) and then correlating these values. Because the correlation is calculated between jackknife replications, we address this approach as jackknife correlation (JC). First, we demonstrate the equivalence of JC to conventional correlation for simulated paired data that are defined per observation and therefore allow the calculation of conventional correlation. While the JC recovers the conventional correlation precisely, alternative approaches, like sorting-and-binning, result in detrimental effects of the analysis parameters. We then explore the case of relating two spectral correlation metrics, like coherence, that require multiple observation epochs, where the only viable alternative analysis approaches are based on some form of epoch subdivision, which results in reduced spectral resolution and poor spectral estimators. We show that JC outperforms these approaches, particularly for short epoch lengths, without sacrificing any spectral resolution. Finally, we note that the JC can be applied to relate fluctuations in any smooth metric that is not defined on single observations.

Introduction

Brain activity exhibits a very high degree of moment-to-moment variability. Activity fluctuations in one brain area are often correlated to fluctuations in other areas. These inter-areal correlations themselves most likely also undergo moment-to-moment fluctuations in strength, and it is an intriguing question whether those fluctuations are related to fluctuations in behavior, in the activity of other brain areas, or in the strength of correlation between other brain areas. Consider the following example: Areas A and B might show beta-band coherence, and at the same time areas B and C might show gamma-band coherence. This might lead us to wonder if the interaction between areas A and B is related to the interaction between B and C. Determining such a relation is highly desirable for neuroimaging applications where the correlation between elements of large-scale networks is an issue of great interest (Park and Friston, 2013; Turk-Browne, 2013). Yet, this is difficult to achieve, because determining the strength of correlation already entails the observation of changes in one signal and related changes in another signal. Thus, determining correlation requires multiple observations and therefore, the strength of correlation cannot be determined on a single observation, i.e. it cannot be determined on a moment-by-moment basis. So, is it impossible to relate fluctuations in the correlation strength between two areas to fluctuations in other parameters?

Here, we present an approach that achieves this: The Jackknife Correlation (JC). JC builds on the work of Stahl and Gibbons (2004), which extended the jackknife method of Miller et al. (1998) to the case of quantifying correlations between brain potentials and behavioral variables. They demonstrate that correlating jackknife estimates of the lateralized readiness potential to personality metrics is superior to single-subject based approaches. JC transfers this rationale to the case of correlations involving covariance-based metrics, which are strictly not defined for single observations. JC enables the correlation of these

[*] Corresponding author at: Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, 60528 Frankfurt, Germany.
E-mail address: craiggrichter@gmail.com (C.G. Richter).

http://dx.doi.org/10.1016/j.neuroimage.2015.04.040
1053-8119/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

metrics to other metrics, like RT (that are defined on single observations), but crucially, JC also allows the correlation of these metrics to each other like in the above example of correlating the A–B beta-band coherence to the B–C gamma-band coherence. Thereby, it is an important new tool for the investigation of functional connectivity.

The jackknife technique successively leaves out each observation once. Each time one observation is left out, this results in an all-but-one ensemble of observations, called a jackknife replication. Thereby, for N observations, there are N jackknife replications. Each jackknife replication contains N-1 observations, and thereby allows quantifying the correlation strength across those N-1 observations. These correlation strengths fluctuate across the N jackknife replications. Because each jackknife replication leaves out only one observation, the variance across jackknife replications is small. Yet because each jackknife replication leaves out precisely one observation, the variance across jackknife replications is a precise transform of the variance across the original observations. Because correlation is driven solely by covariance and normalized for the variances of the correlated signals, the correlation between jackknife replications is in fact identical to the correlation between the original observations. We will demonstrate this first for simulated data that are defined for each single observation. We propose that this is an answer to the abovementioned question, namely that the same approach can be taken for testing whether fluctuations in correlation are related to other parameters, even though it may not be possible to determine the value of either variable on a moment-by-moment basis, as is the case for ensemble metrics such as coherence. We support the proposal by simulating data with an autoregressive model such that the correlation was dependent on a fluctuating pre-determined control parameter. This pre-specified relation between the control parameter and the correlation was then successfully recovered through JC.

Alternative approaches to computing correlation upon covariance-based metrics

Approaches to this problem can be divided into two classes. The first seeks to determine a value for the covariance-based metric or each single trial. The second approach estimates the ensemble metric over subgroups of trials formed by decomposing the total number of trials into subensembles. Consider the following example: Suppose we wish to investigate the trial-by-trial correlation between reaction time (RT) and inter-areal gamma-band coherence. While RT is defined on each trial, coherence is not. Coherence quantifies the consistency of phase relations across multiple trials, which renders it undefined at the level of a single trial. The first approach would attempt to determine the coherence of each single trial by subdividing each trial into multiple epochs and computing the coherence over each of these sub-segments (Welch, 1967; Lachaux et al., 2000). Alternatively, one could achieve the same single-trial estimate by applying multiple data tapers over the single epoch (Mitra and Pesaran, 1999). Yet, both methods are limited by the nature of brain dynamics in general, where periods of interest are often present for only brief instances, such that single trials are typically too short to derive multiple spectral estimators, or apply large numbers of tapers. Another approach to estimating coherence on a single-trial basis, which is in fact closely related to JC, is the use of jackknife pseudovalues (Womelsdorf et al., 2006). The jackknife pseudovalue is an estimate of the single-trial value of a statistic that is based on the difference between 1) the statistic calculated across all trials (weighted by n) and 2) the statistic calculated on all-but-one trial (weighted by n-1). A problem with the pseudovalue approach can arise e.g. from the following combination of facts: 1) the difference between the all-trial and the leave-one-out estimate is very small, and 2) many interesting metrics, like coherence, carry a sample-size dependent bias (Maris et al., 2007), i.e. the coherence bias will be slightly larger for the leave-one-out estimate than for the complete estimate. While the bias from point 2) is small, also the difference from point 1) is small, and this combination can lead to problems with the single-trial estimate, that necessitate complicated solutions. These problems are fully avoided by JC, because it calculates the correlation directly between the jackknife replications of the statistic without attempting to estimate the statistic on a single trial. If one nevertheless wants to estimate coherence on the single trial level, e.g. for illustration purposes, then the pseudovalue approach might be used together with a bias-free metric of interaction strength, like the recently introduced pairwise phase consistency metric, PPC (Vinck et al., 2010). To summarize, single-trial estimation approaches all suffer from either reduced accuracy of the estimate, or excessive computational complexity, thus it is most desirable to work with coherence estimates computed over multiple trials.

The second approach does just this and we address it as sorting-and-binning. Sorting-and-binning can only be used if one of the variables is defined on the basis of single observations. The observations are sorted and binned according to this (single-observation-defined) variable. For this variable, the mean per bin is computed. For the other variable, which is not defined on a single observation, the covariance-based metric is calculated separately per bin, across the multiple observations within each bin. Finally, the correlation between the two metrics is computed across the bins. This approach can be found in a large number of studies ranging beyond neuroscience. See Liang et al. (2002),Hanslmayr et al. (2007),Womelsdorf et al. (2007) andvan Elswijk et al. (2010) as examples of the technique. It's important to note that if neither quantity over which we wish to perform the correlation is defined on a single-observation basis, then this method cannot be applied, since sorting cannot be performed. JC is not limited in this way since neither variable need be defined for a single trial. We will further investigate the process of sorting-and-binning to illustrate some often overlooked statistical pitfalls of this technique while in parallel developing the mechanics of JC.

The sorting-and-binning approach

The sorting-and-binning approach proceeds in the following manner: Suppose we have 1000 trials. We can sort these according to RT, bin them into 20 bins of 50 trials, calculate the mean RT per bin, calculate coherence per bin across the 50 trials in the bin, and finally calculate the correlation between RT and coherence across the 20 bins. With this approach, the coherence per bin can be computed, because each bin comprises 50 trials. We will demonstrate below that such a binning strategy carries substantial statistical costs. Suppose we have only 200 trials. We do not want to bin them into 20 non-overlapping bins of 10 trials each, because 10 trials will result in poor coherence estimates. On the other hand, non-overlapping 50-trial bins will result in only 4 bins, which is a very low n for useful correlation. Thus, we might consider overlapping our bins. If the 50-trial bins are overlapped by 40 trials, this furnishes us with 16 bins. We will demonstrate that the combination of binning with overlap incurs further costs.

To simplify this demonstration, we begin with two correlated random variables, of 1000 trials, that are both defined on a single-trial basis, such as e.g. the gamma-band power of two brain areas. We use the mean as the statistical operation we apply to each bin. The variables were generated with a covariance of 0.1, which leads to a Pearson correlation coefficient of $r(998) = 0.1$, $p < 0.0018$. The r-value and p-value surfaces (Fig. 1) were computed using a grid of combinations between overlap percentages and bin sizes, which was selected so as to include only those combinations that used all of the data with no remainder, i.e. the final bin terminated on the final data sample. This grid is irregularly spaced. For a maximum bin size of 250 trials, this resulted in 1286 bin/overlap combinations, which each resulted in a Pearson product moment-correlation coefficient r and significance level p. To establish statistical stability, these values were evaluated 10,000 times and averaged. Correlation coefficients were converted to t-values and assessed for significance using Student's t-distribution (Rahman, 1968). The resulting irregular grid of r- and p-values was interpolated to an even

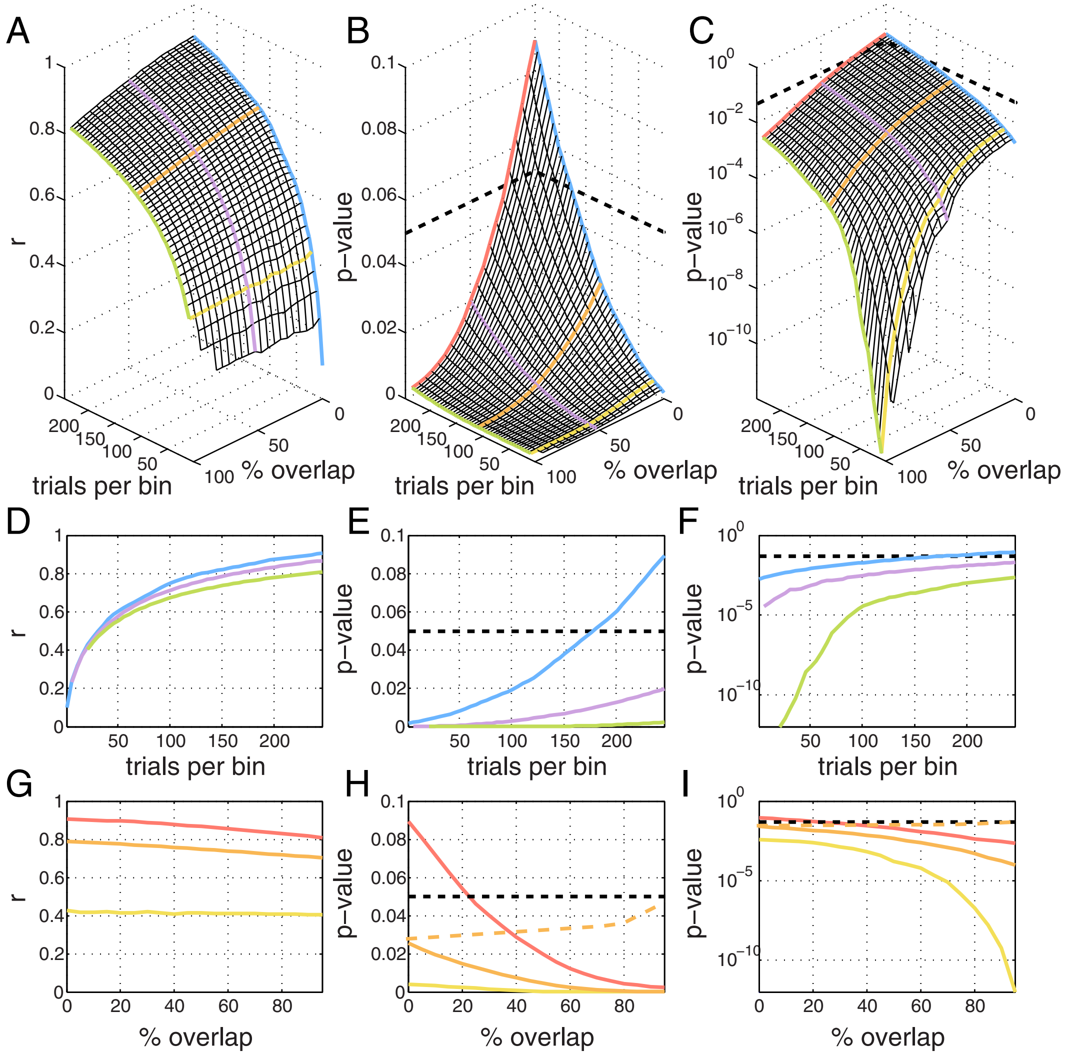

Fig. 1. Parametric examination of the effect of sorting-and-binning on Pearson's product moment correlation coefficient (r), and its statistical significance. A. Parametric surface depicting the effect on r as the bin size and degree of overlap are varied. B. Corresponding map of the parametric change in statistical significance (p-value) for the r-values shown in A. C. Same as B, but with logarithmic p-value axis. D. The effect of bin size on r shown for three overlap parameters. G. The effect of overlap on r shown for bin sizes of 20 (yellow), 120 (orange) and 245 (red). E, H. Corresponding p-values for D and G. These p-values are false and are depicted only for illustration purposes (see main text). F, I. Same as E and H, but with logarithmic p-value axis. The dashed black line marks the 0.05 significance threshold in all p-value plots. The dashed orange line in H and I shows the correct (Monte-Carlo based) p-values for the middle bin size.

grid with a spacing of 5 trials using Delaunay triangulation. All simulations were performed using MATLAB (The MathWorks, Inc.).

We can now examine the various binning/overlap parameterizations (Fig. 1). For the case of zero overlap, Fig. 1 demonstrates that an increasing number of trials per bin results in a correlation coefficient that increases from the true trial-by-trial value of 0.1 to a value close to 1 (blue curve in Fig. 1A). This can be explained by examining the scatter plots in Fig. 2(E, F, I, and J), which show that the residuals (the distance of each point from the line of best fit) decrease as the bin size increases. This effect is due to the averaging out of random variation in the data. While the large r is not incorrect if considered in the context of its calculation, and it might appeal to a scientist looking for a clear effect, there are several points that have to be considered: 1) When the r-value is computed between the single-trial variables, then the squared r-value gives the variance in one variable explained by the variance in the other, i.e. r and r-squared can be used directly as metrics of an effect size (Cohen, 1988). After binning, the (squared) r-value cannot anymore be interpreted in this way. Readers need to take this into account when interpreting r-values obtained with binning. 2) The resulting r-value will depend on the original r-value and also on the amount of binning. When binning differs, e.g. between different studies, this renders the r-values incomparable. 3) The amount of binning affects the p-value. Fig. 1 reveals that the increase in the correlation coefficient is mirrored by an increase in the p-value, which is simply explained by the decrease in n (the number of bins) as the bin size increases. As a consequence, with increasing bin size, more and more tests will fail to reach significance. We illustrate this with the power analysis shown in Fig. 3. The curve in Fig. 3 corresponds to the zero-overlap tests shown by the blue curves in Figs. 1 and 2. Statistical power is the probability of correctly identifying an experimental effect, and thus quantifies the sensitivity of a test. To establish the statistical power as a function of bin size we performed the following Monte-Carlo simulation (keeping the type I error rate fixed at 0.05). We first simulated (using the MATLAB function mvnrnd) two random variables of sample size n and expected correlation r, and computed the observed correlation r in that sample. This was repeated 1000 times, leading to a randomization distribution for r. Second, we generated two random variables of sample size n and correlation r of zero, and computed the correlation r_0, again 1000 times, leading to a randomization distribution for r_0. From this latter distribution, we determined the 95th percentile. Finally, we determined the proportion of the randomization distribution of r that exceeded the 95th percentile of the randomization distribution of r_0. This proportion was taken as the power for detecting a significant correlation, given r and n. For a true r of 0.1, Fig. 1A shows the estimates of r

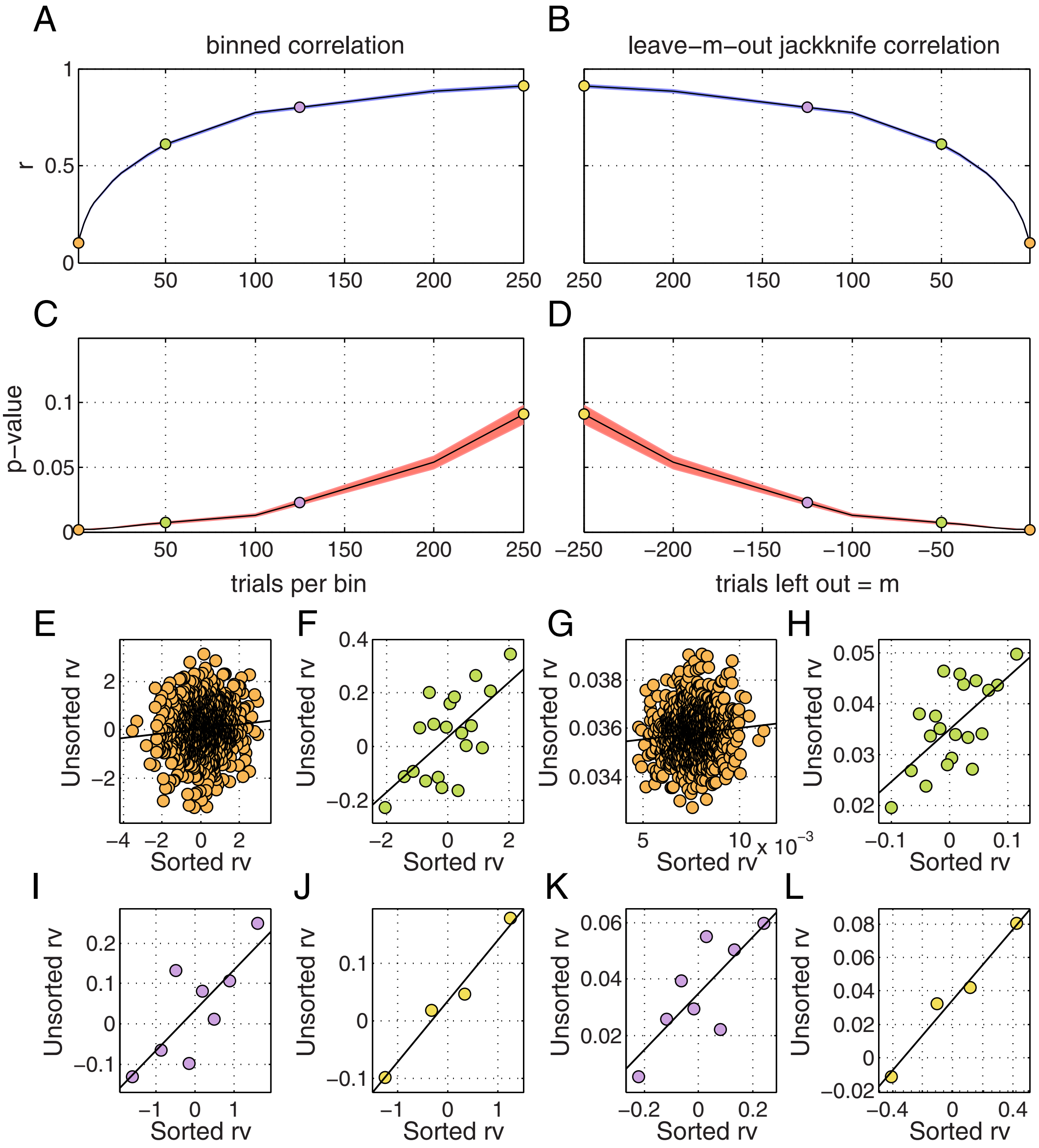

Fig. 2. Equivalence of binned correlation with leave-*m*-out jackknife correlation. A, B. Pearson's *r* as a function of bin size (A), and as a function of the number of trials left out (B). C, D. Corresponding p-values for panels A and B. E, F, I, J. Scatter plots of the bin averages of the sorted and unsorted random variables (rv) for the 4 correspondingly colored points shown in A and C. G, H, K, L. Scatter plots of the leave-*m*-out averages of the sorted and unsorted random variables (rv) for the 4 correspondingly colored points shown in B and D.

for different overlap percentages and bin sizes. For zero overlap and a representative subset of bin sizes, we determined the number of bins *n* and the estimate of *r* (the latter by reading from Fig. 1A). For those pairs of *r* and *n*, we determined the statistical power, as explained above, 100 times and show the average statistical power across those 100 repetitions in Fig. 3.

As suspected based on the behavior of the p-value with increasing bin size, the statistical power systematically decreases as the bin size is increased from 1 trial to 250 trials. As bin size increases, the number of bins decreases, thus though the correlation coefficients increase (Figs. 2E, F, I, and J), this increase is countered by a decrease in the number of observations, which results in a net loss of statistical power.

To summarize, the binning of data (even without overlap between bins), results in an increase in the correlation coefficient due to smoothing of the data, and a decrease in the statistical power of the test. Thus this analysis indicates that the optimal approach is to not use a binning strategy, such that statistical power is maximized and the *r* and r^2 values can directly be taken as metrics of effect size.

Let's now consider the effects of overlapping the bins. It is apparent from the *r* surfaces/lines in Figs. 1A, D, and G, that binning with overlap leads to the same inflation of the correlation coefficient that occurs without overlap. It is also clear that overlapping the bins also leads to a marginal decrease in the *r*-value, but more worrisome is the massive decrease in the p-value as overlap is increased. Following the colored curves (red, orange, yellow) in Figs. 1B, C, H, and I, we can see that the p-value dramatically decreases as overlap is increased. The result appears attractive, since large *r*-values are achieved in combination with impressively small p-values, but these results are false, because the data points entered into the correlation analysis are not independent due to the bin overlap. With an increasing degree of overlap, the bins become less independent, which effectively inflates the degrees of freedom (*df*), such that from the point of view of the test, there are far more observations than were in fact there. This is basic statistics, but the issue should be kept in mind since in more complex designs, this violation may be more difficult to spot. If conditions demand that overlap must be used, the statistical inflation may be corrected by applying the following Monte Carlo approach to computing the p-value:

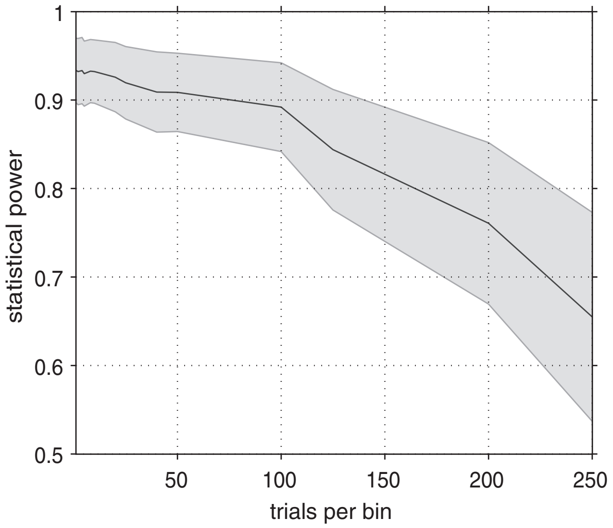

Fig. 3. Statistical power as a function of the number of trials per bin. Gray shaded region specifies +/− 1 standard deviation of the mean.

1) Randomly pair the data of the unsorted random variable with the sorted random variable, such that the first variable remains sorted.
2) Recompute *r* over the bins as before.
3) Repeat steps 1 and 2 hundreds to thousands of times to produce a distribution of chance values for *r*.
4) Determine the p-value from the proportion of the surrogate values that the empirical correlation coefficient exceeds.

This method should not be confused with another Monte Carlo approach used for assessing the statistical significance of the correlation coefficient. In the context of our example, this would involve randomly re-pairing the bins, and computing a surrogate distribution based on these random re-pairings. This is a much faster approach since the bins do not need to be recomputed, as they are in step 2 of the algorithm above, but because of this, it will fall victim to the same decrease of the p-value that is shown for the parametric case in Fig. 1. Fig. 1 (H, I) depicts, for 120 trial bins, the deflated parametric p-values as orange lines, and the correct non-parametric p-value as dashed orange lines. These values reveal that when fairly assessed, the slightly increasing p-values (Figs. 1H and I) parallel the slightly decreasing *r*-values (Fig. 1G). Thus it is apparent that, when properly computed, overlap conveys a disadvantage since the *r*-value is always decreased relative to the zero overlap case, so such a procedure should only be employed when proceeding without overlap is impossible, and great care should be taken to establish a legitimate assessment of statistical significance.

To summarize, we were initially motivated by the example of determining the correlation between RT and interareal gamma-band coherence. Since coherence is not defined for a single trial, we postulated a sorting-and-binning approach, with or without overlap, as a potential solution. Using simulated data, we then demonstrated the undesirable properties of binning, with and without overlap, which are deficiencies that extend to both variables that are defined or undefined on a single-trial basis. Thus, the sorting-and-binning approach has brought us no further than where we began, since we now have even more reason to aim for quantifying correlation at the single-trial level, i.e. 1 trial per bin, due to the following desirable properties of this approach:

1) It has the maximal statistical power over all binning strategies.
2) The correlation coefficient is most representative of the true underlying correlation coefficient and can directly be used as metric of effect size.
3) It allows for the correct assessment of statistical significance using conventional methods.

Yet, despite these desirable properties, we are still barred from performing a single-trial correlation by the lack of definition of coherence on a single trial. In the following section we will introduce a method designed to overcome this issue: the jackknife correlation (JC).

Jackknife correlation

We will begin the explanation of JC by reviewing the fundamental technique underlying the method: the jackknife. The jackknife technique, originally proposed by Quenouille (1949) and extended by Tukey (1958), is a method designed to assess the standard error of an estimator without underlying parametric assumptions (Parr, 1985). The procedure involves computing a statistic of interest iteratively over all the combinations of the data where one sample, or trial in our case, has been left out of the calculation. This is known as the leave-one-out jackknife replication (or just "jackknife replication") of the statistic, and is defined as follows:

$$S_i = S(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

where *S* is the statistic of interest calculated over the samples *x*. In terms of our example, x_i is a single trial, and *S* is the coherence. So practically, this operation entails computing the coherence *n* times, as each of the samples *x* is systematically left out. This results in *n* jackknife replications of *S*, each referred to as S_i.

We propose that the jackknife rationale offers an attractive solution for the single-trial correlation of covariance-based metrics. The logic for this solution begins with establishing the equivalence between ordinary correlation and JC.

Following the approach of Stahl and Gibbons (2004), correlation can be expressed as the expectation of the product of the standard scores of two variables *x* and *y*:

$$r_{xy} = E(z_x z_y),$$

where *E* is the expectation operator. We wish to establish that:

$$r_{xy} = r_{x_{jk} y_{jk}},$$

where x_{jk} and y_{jk} are the jackknife replications of *x* and *y*. We thus need to determine that the following relation is true:

$$E(z_x z_y) = E\left(z_{x_{jk}} z_{y_{jk}}\right).$$

This can be easily shown, due to the fact that the z-score of a jackknife replication is simply the z-score of the original value multiplied by −1 (Stahl and Gibbons, 2004). If we make this substitution in the above equation, then we can see that the equivalence between ordinary correlation and JC must be true.

$$E(z_x z_y) = E(-z_x - z_y)$$

Therefore,

$$E(z_x z_y) \equiv E\left(z_{x_{jk}} z_{y_{jk}}\right).$$

The above relation offers a unique avenue for dealing with the correlation of covariance-based quantities since even though they cannot be adequately defined for single trials; covariance-based quantities are defined over jackknife replications. Thus the single-trial correlation of covariance-based quantities may be determined as the correlation of their jackknife replications. To illustrate this more intuitively, we may compare a leave-*m*-out jackknife strategy to the sorting-and-binning approach. When sorted data is binned, the statistical operation *S* is performed on the *m* samples that compose the *n* bins. The correlation is then performed on these *n* results. For the leave-*m*-out jackknife, the correlation is performed on the n results of the function S applied to the data remaining after each bin of m trials has been left out once. The symmetry of these methods is apparent in Figs. 2A–D, where leave-m-out JC has been applied to the numerical data used to investigate the consequences of sorting-and-binning, i.e. data for which single-trial estimates are available. Figs. 2A–D show the comparison of correlations based on binning without overlap (Figs. 2A and C), and leave-m-out JC (Figs. 2B and D). What is immediately obvious from the topmost panels is that the correlation functions are mirror symmetric. Comparison of the leftmost point of the binned correlation (Figs. 2A and C), corresponding to a bin size of 1; with the rightmost point of the leave-m-out JC (Figs. 2B and D), which corresponds to the leave-one-out JC, reveals precisely the same r-values and p-values. Thus, as dictated by the mathematical proof, conventional correlation is equivalent to JC.

JC has a particular strength that should be noted. Since the method does not require the sorting of any variables, neither variable involved in the correlation needs to be defined on a single-trial basis. This means that the method may be used to assess the correlation between two variables that are both not defined on the level of a single trial.

Note that the JC entails an inversion and compression of the sample distributions by the jackknife method. This can be seen by comparing the binned versus the leave-m-out jackknife scatter plots in Figs. 2E–L. The leave-m-out jackknife scatter plots are up/down and left/right mirror reversals of the scatter plots that result from binning. Furthermore, the JC scatter plots contain smaller values, because they essentially represent only the small changes in the function F when a single trial out of 1000 trials is left out. Both the compression and the double inversion are irrelevant for correlation analysis. The compression is compensated by the fact that the correlation coefficient normalizes the covariance by the product of the variances of each of the two variables. The inversion is irrelevant, because it occurs in both variables, and correlation is invariant to the sequence of the paired variables. Yet, if instead of linear correlation metrics, non-linear fits are to be performed, or non-linear effects are qualitatively assessed visually from the data, one must be careful to provide the correct interpretation. To illustrate this point, Fig. 4 shows the effect of the JC for two example variables with a non-linear relation. This makes the inversion very apparent and the potential for misinterpretation quite obvious.

It must also be noted that the jackknife technique in general, and thereby also the JC, should only be used in combination with statistics whose underlying distributions are smooth (Miller, 1974; Efron, 1979; Parr, 1985). An example of a statistic that is not smooth is the median. The jackknife replications of the median of a distribution are the middle two values of the distribution. These two values do not capture the variance contained in the full distribution, which we attempt to capture with the jackknife method. Most functional connectivity metrics, such as coherence, are based on an averaging operation over the trial dimension, and thus should be suitable for use with the JC.

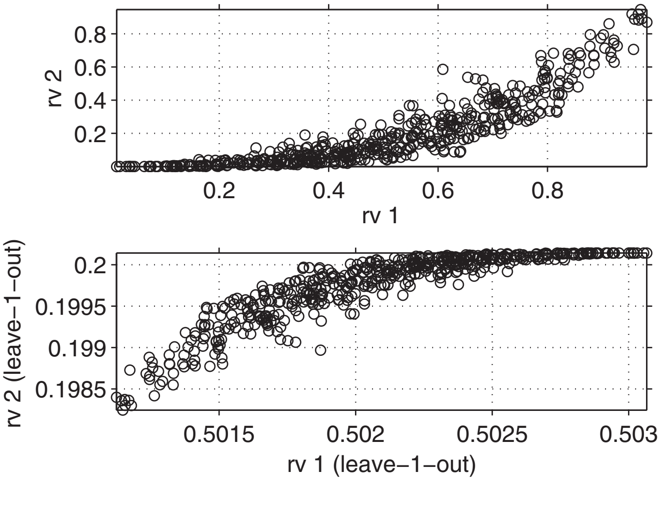

Fig. 4. The jackknife procedure causes an inversion of the distribution and a variance compression of random variable 1 (rv1) and random variable 2 (rv2).

Numerical investigation of JC and application to simulated data: methods

We have demonstrated in Fig. 2 that, for parameters that are defined on a single trial, the JC is identical to the conventional single-trial correlation. Ideally, we would want to show that the same holds for metrics of interactions like coherence. Yet, this is problematic since the coherence is not defined for single spectral estimates, and therefore, the JC for coherence cannot be compared against a ground-truth estimate. While we cannot estimate coherence for single spectral estimates, we may generate simulated data epochs with a coherence that is proportional to a coupling parameter c using a generative model (Brovelli, 2012). If we further vary c across multiple instantiations of this generative model, we can expect that c will be correlated with the single-trial coherence. We employ such a setup to test whether trial-by-trial correlation generated in this way can be recovered by the JC technique.

We constructed the following autoregressive (AR) model to generate appropriate data to test the JC technique in a system exhibiting inter-areal coherence and unidirectional GC:

$$\begin{aligned} x_t &= 0.95x_{t-1} - 0.8x_{t-2} + \varepsilon_{x,t} \\ y_t &= 0.8y_{t-1} - 0.5y_{t-2} + cx_{t-1} + \varepsilon_{y,t}, \end{aligned}$$

where x_t and y_t represent time series from two brain areas. x_t is a function of its own values at one and two time steps in the past, i.e. x_{t-1} and x_{t-2}, weighted by some chosen coefficients (0.95 and −0.8), plus $\varepsilon_{x,t}$, i.e. an noise term (also called the innovation). The situation is similar for y_t, except that it is a function not only of its own past values plus the noise term, but also of y_{t-1}, weighted by c. This model is a bivariate (two signals) autoregressive (the signals depend on their own past) model of order two (they depend on two time steps into the past), i.e. it is an AR(2) model. We chose an AR(2) model specifically, because it is the model of minimal complexity that generates synthetic data with band limited power, coherence, and GC spectra. The crucial aspect of the model is that it exhibits unidirectional coupling and thereby coherence that is determined by the parameter c. This will allow us to vary c from trial to trial and thereby generate fluctuations in coherence that are perfectly rank-correlated with the fluctuations in c. While c might translate into the coherence magnitude in a non-linear way, it does translate in a monotonic and smooth way and this guarantees that the expected rank correlation has a value of one. Fifty time-series pairs were generated for the bivariate AR(2) process with each time-series modulated by a unique coupling parameter value c chosen randomly from a uniform distribution between 0 and 0.1. Each time series was 25,600 samples long at a sampling rate of 500 Hz, resulting in 51.2 second segments. AR models cannot only generate simulated time series, but they can also be fit to experimental or to simulated data in order to quantify the spectral properties, like power, coherence or GC. For all parametric analyses of the generated data, AR modeling was performed using software developed by Steven Bressler and Mingzhou Ding. The model parameters were determined using a vectorized implementation of the algorithm of Morf et al. (1978), generously provided by Anil Seth and Lionel Barnett (Barnett and Seth, 2014). Bivariate AR(2) models were fit to the synthetic time-series pairs using portions of the data truncated to varying lengths (see Fig. 6). In the case of conventional single-trial correlation, models were fit to the data from each trial pair, while in the case of JC, models were fit using all trials minus one for all leave-one-out possibilities. Coherence and Granger causality spectra (Granger, 1969; Geweke, 1982) were derived from the fitted AR(2) models, and the max of the spectrum was selected. For these max values we then determined their correlation with the coupling coefficient c. This procedure was repeated 1000 times with the mean taken over the resulting ensemble of Spearman's rho correlation coefficients to achieve sufficiently smooth estimates. All p-values were determined parametrically.

Spectral estimates were also obtained using non-parametric analysis. All non-parametric spectral connectivity analyses were performed using Fieldtrip (Oostenveld, et al., 2011). Spectral estimates were computed via a Fourier transform using the multi-taper method (Mitra and Pesaran, 1999; Thomson, 1982) with a spectral smoothing of +/− 10 Hz. We compared the JC approach to a non-JC approach based on epoch subdivision of single trials. In the subdivision-based approach, for data windows less than 400 ms in length, the cross-spectral density (CSD) was computed as the mean of the estimates deriving from the multiple data tapers. For trial lengths longer than 400 ms, 400 ms windows with 300 ms overlap were employed. In these cases, the CSD was computed as the mean over tapers and windows. For the JC-based approach, the CSD was determined via jackknifing, which entails taking the mean CSD resulting from all trials minus one, for each window, for all leave-one-out combinations. Coherence spectra were derived from the power and CSD, while GC spectra were determined using non-parametric spectral factorization, which is a method of obtaining GC from non-parametric spectral estimates such as the Fourier or wavelet transform (Dhamala et al., 2008a,b; Wilson, 1972). This procedure was performed for each 400 ms window and following Welch's method (1967) the coherence and GC spectra for each jackknife replication were determined as the average over the windows. The peak value of the coherence and GC spectra was determined, and the mean of the frequency band of +/− 30 Hz around this peak was used for subsequent JC and conventional single-trial correlations. As for the parametric case, this procedure was repeated 1000 times to achieve smooth estimates of Spearman's rho and parametrically determined p-values.

Numerical investigation of JC and application to simulated data: results

We fit an AR(2) model to the first 10 s of the simulated data to inspect the spectral properties of the simulated data (Fig. 5). The use of a long 10-second segment allows us to capture the parameters of the model almost perfectly, and clearly see the effects of modulating parameter *c*. It is apparent that in all spectra, where there is spectral energy, that there is a peak between 70 and 90 Hz, corresponding to the gamma band, as specified by the model parameters. It is also apparent that the single-trial power of x_t (Fig. 5A) is not modulated by *c*, which is evident since the progression of color from the smallest to largest peak does not follow the color scheme corresponding to parameter *c*, whereas the power of y_t (Fig. 5B), though lower in overall value, shows modulation by *c* based on the color progression. This is due to the unidirectional flow of power from x_t to y_t that is modulated by *c*. The coherence shown in Fig. 5C shows clear modulation by *c*, and the GC (Figs. 5D and E) shows unidirectional coupling also modulated by *c*. We determined the JC between *c* and coherence or GC influence and confirmed that it approaches a value of one for long epoch lengths (Fig. 6A for coherence, B for GC influence). These JCs are shown in blue for spectral estimates derived from fitting AR models, and in gold for spectral estimates derived non-parametrically. For shorter epoch lengths, the JC decays. Short epochs realize the properties of the generative model only in an imperfect way. Thus, the decay in JC correlation away from the value of one is not necessarily due to an imperfect estimation of the underlying correlation, but it is likely due to the fact that the correlation between, on the one hand *c*, and on the other hand the short-epoch coherence is actually low. In order to substantiate this claim, we turn again to a case in which we can quantify the metric of interest on data epochs of arbitrary length. We generated 50 Gaussian random signals with zero mean and unit variance. We then added a random offset *o* to each of these signals, drawn from a uniform random distribution between zero and one. We then correlated the *o* to the means calculated over data epochs of variable length (randomly subsampled from the full-length signal). The result is shown in Fig. 7. As the data epochs get shorter, the correlation between the epochs' means and the values of *r* falls off in a way that is very similar to the drop-off in correlation seen in Fig. 6. This effect is solely due to error between the subsample means and the original offsets (which are equal to the full-length data means). Thus, this is not due to an error in estimating the mean, but rather it is due to the failure of the shorter data segment to express the expected mean of the process. With this in mind, we can go back to Fig. 6. With the imperfect expression of the model parameters in short epochs explaining an overall drop-off, we can turn to the differences between different approaches to estimating the correlation between short epochs across single trials. In Figs. 6A and B, we see that indeed as the trial length decreases, so does the correlation coefficient. The critical test for JC is to determine if, for a given epoch length, the method is providing superior estimates of the correlation coefficient in comparison to conventional approaches to this problem. We compare the JC against conventional single-trial correlations, which attempt to estimate the Fourier transform from single, short data epochs using either overlapping data windows (Fig. 6, green lines) or single-trial AR(2) models (Fig. 6, red lines). We see in Figs. 6A and B that at the longest epoch length, the correlation coefficient based on either conventional single-trial metrics approached one, like the JC correlations. Critically, as the epoch length decreased, the JC estimates of the correlation coefficients remained above the correlation coefficients based on single trial parametric estimates (Fig. 6, red lines) and non-parametric estimates (Fig. 6, green lines). This is even more evident if we plot the percentage difference between the estimator with the largest correlation coefficient at each trial length and each of the different metrics, as shown in Figs. 6C and D, where we see similar performance between the JC on parametrically and non-parametrically derived estimators when the trial length exceeds 300 ms. The small superiority of parametric JC versus non-parametric JC, particularly for short window lengths, is likely due to two effects: 1) data windowing effects in the non-parametric approach, which are exacerbated at short epoch lengths and 2) the fact that the data had been generated with a parametric model and therefore might be fitted particularly well with a parametric model. As will be shown in the following section, both metrics show

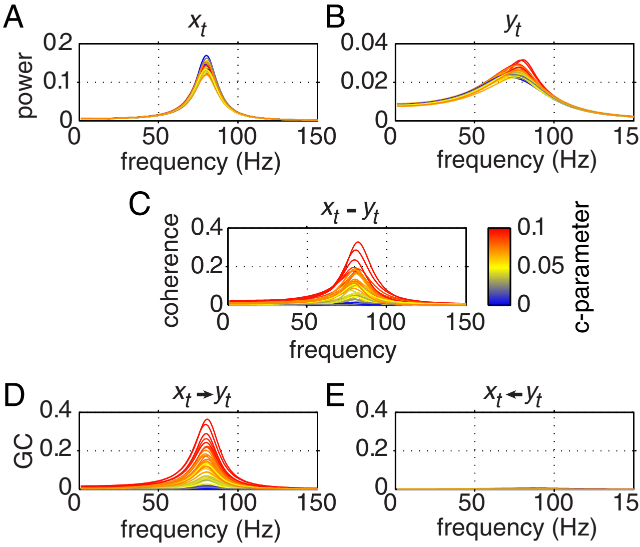

Fig. 5. Spectral properties of the bivariate autoregressive model for various values of the coupling parameter *c*. A. Power of variable X_t for the 50 simulated time series. B. Power of variable Y_t for the 50 simulated time series. C coherence between X_t and Y_t for the 50 simulated time series. D. Granger causality of X_t to Y_t for the 50 simulated time series. E. Granger causality of Y_t to X_t for the 50 simulated time series. Colorbar denotes the magnitude of the coupling coefficient *c*.

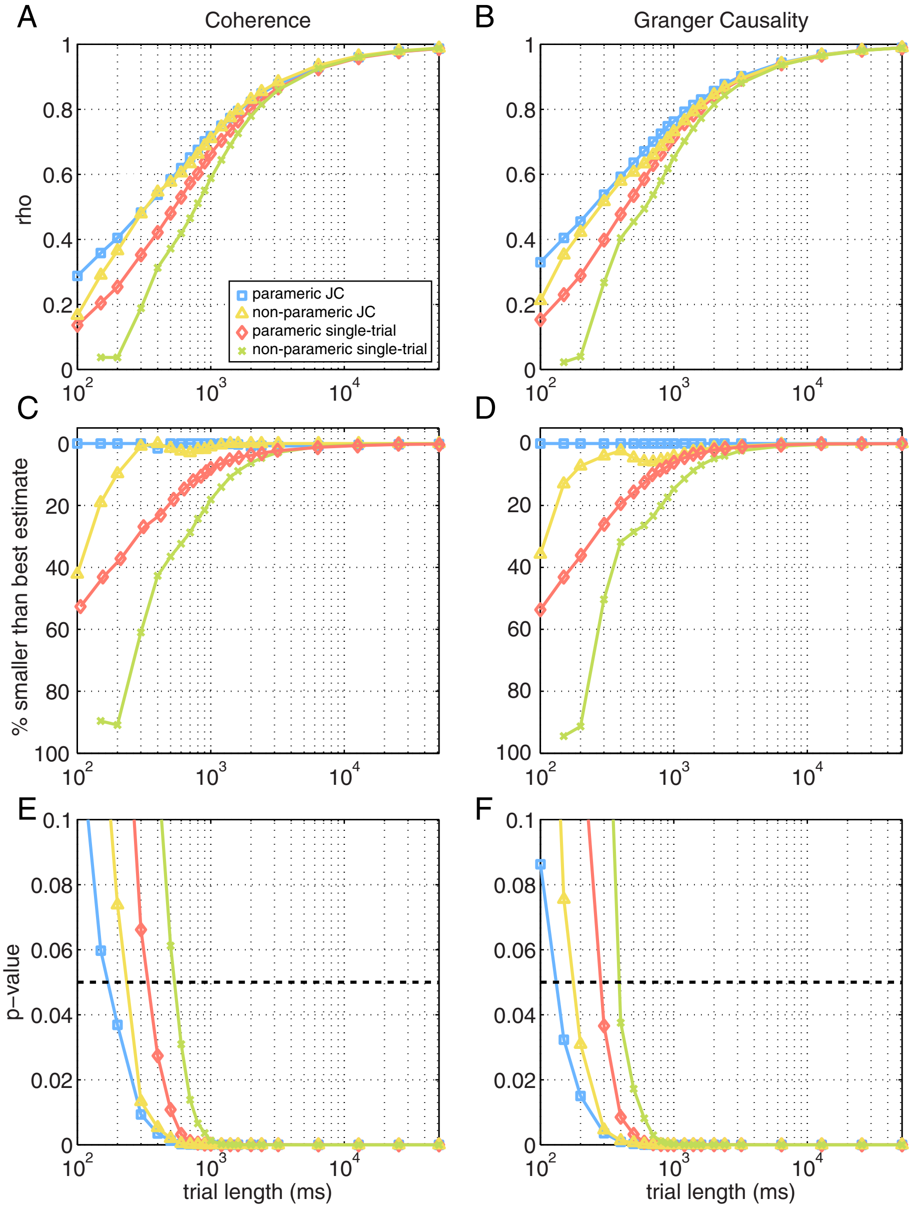

Fig. 6. Comparison of JC with conventional spectral correlation methods as a function of trial length on simulated time series. A, B. Spearman's rank correlation of the coherence (A) and GC (B), with the coefficient c for the 50 coupled simulated time series X_t and Y_t as a function of trial length. Blue squares depict parametric JC, gold triangles depict non-parametric JC, red diamonds depict single-trial parametric estimation, while green crosses depict a window-based non-parametric approach. C, D. Percentage that each correlation metric is smaller than the metric yielding the greatest rho-value at each trial length. E, F. p-Values corresponding to each rho-value shown in A and B with dashed lines marking the 0.05 significance threshold.

similar performance on empirical data, with the non-parametric estimator showing moderately increased performance. Panels E and F of Fig. 6 display the corresponding p-values to demonstrate the average epoch length necessary before a significant result is obtained. This illustrates that the JC discovers a significant effect with shorter data epochs, which is of great advantage in neuroimaging analysis. Another effect of note is the poor performance of the single-trial non-parametric window-based method on short epochs. Based on these simulations and the empirical analyses that will be shown below, it is advisable that this method be avoided for short data epochs in favor of JC.

In conclusion, these numerical examples demonstrate the superior performance of JC in recovering simulated correlations compared to conventional single trial estimates. In the following section, we will investigate whether this holds for experimental data.

Application to neurophysiological data: methods

Electrophysiological recordings and experimental paradigm

We compared the JC with single-trial approaches for both parametric and non-parametric spectral estimators applied to neurophysiological data to evaluate the performance of JC. All experimental procedures were approved by the ethics committee of the Radboud University Nijmegen (Nijmegen, The Netherlands). For details of the experimental methods and recording techniques see Bosman et al. (2012), except that stimuli were positioned in opposite hemifields with only one stimulus co-activating recording sites in areas V1 and V4.

Two rhesus monkeys (*Macaca mulatta*) were trained to perform a covert visual spatial attention task. We show data in this paper from monkey K. Two grating stimuli were presented, one in the lower right visual hemifield, and one in the upper left visual hemifield. The gratings

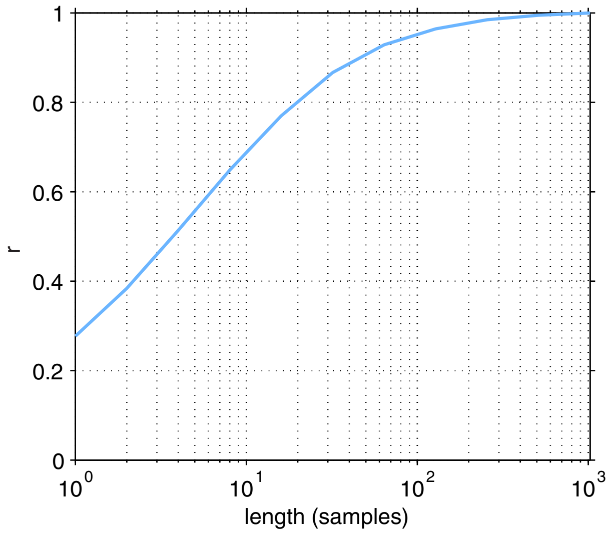

Fig. 7. The effect of trial length (samples) on the correlation (r) between two perfectly correlated random variables.

were isoluminant and iso-eccentric drifting sinusoidal gratings with a diameter of 3° visual angle, a spatial frequency of 0.66 cycles/degree, drift velocity of 1.2°/s, with a resulting temporal frequency of 0.8 -cycles/s and 100% contrast. The two gratings had orientations that were always 90° away from each other and, when they were moving, inconsistent with the interpretation of a chevron pattern seen through two apertures. For a given session, two orientations were chosen, and on a given trial, the orientation shown contralateral to the ECoG grid was chosen from those two orientations pseudorandomly. The stimuli were presented on a CRT monitor with a 120 Hz refresh rate non-interlaced. For each trial, one stimulus was randomly tinted yellow, and the other blue. Local field potentials (LFP) were recorded from the left hemisphere with a subdural electrocorticographic (ECoG) grid consisting of 252 electrodes (1 mm diameter), spaced 2–3 mm apart. Signals from immediately neighboring electrodes were subtracted to remove the common recording reference, because otherwise the common reference leads to artifactual coherence/GC influence. We refer to the bipolar derivative resulting from the subtraction of two neighboring electrodes as a "site". For coherence and GC influence analysis, we investigated interactions between primary visual cortex (V1) and extrastriate visual cortex (V4). The assignment of electrodes to brain areas was based on macaque brain atlases. The current analysis examined 29 sites recorded from area V1 and 17 sites recorded from area V4, resulting in 493 V1–V4 site pairs.

The covert spatial attention task consisted of three successive epochs. 1) The prestimulus period where the monkey had achieved fixation. 2) The pre-cue period where the stimulus gratings had appeared and the monkey waited for a color cue, and 3) the cue period, where the fixation point changed color to indicate to which stimulus the monkey should attend, and respond to (the target). While the stimuli were present, either the target or distractor could change shape at any time. The monkey was rewarded for responses to a change of the shape of target stimulus.

For the examples presented in this paper, trials were selected from data segments that spanned at least 2 s from cue onset prior to any stimulus change. The first 0.4 s of this segment were discarded to avoid transients generated by the cue change, leaving 1.6 s for analysis. Correlations were then assessed for 8 different pairs of data windows of varying length from 0.1–0.8 s. Each pair consisted of an early and late segment such that the early windows always terminated one sample prior to 1.2 s, whereas the later windows always commenced at 1.2 s. 352 trials were selected from 9 sessions. Trials were included where attention was directed to either visual hemifield, i.e. data were pooled across attention conditions.

Spectral estimation

The spectral properties of the data were determined both parametrically via AR modeling, and non-parametrically via Fourier analysis so the performance of JC could be compared for both techniques. Parametric spectral estimates were computed in the following way. Data was resampled to 250 Hz. The coherence, and Granger Causal (GC) influence in the "bottom-up" direction (V1 to V4) were then obtained by fitting bivariate autoregressive (AR) models with model order 9, computed for each V1–V4 pair of sites. The model order was determined via the minima of the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC), between model orders 1 and 25. When the model was used for more than one data epoch, e.g. for JC on all-but-one trial, the fit was simultaneously to the ensemble of epochs (Ding et al., 2000), where a separate model was constructed for each leave-one-out jackknife replication. For the non-parametric JC analyses, Fourier transforms were computed using the multi-taper method (Mitra and Pesaran, 1999; Thomson, 1982) on de-meaned data segments with +/− 10 Hz smoothing. For the JC approach, all the data segments (8 variable lengths from 0.1–0.8 s) were zero-padded to 1 s, resulting in a consistent 1 Hz spectral grid, with the CSD for each jackknife replication derived as the mean CSD over trials after one trial had been left out. Spectral coherence was computed for each jackknife replication, while GC was determined via non-parametric spectral factorization of each replication. Identical to the parametric method, coherence was analyzed between V1 and V4 channels and the "bottom-up" direction was analyzed from V1 to V4 for the GC. The single-trial approach followed that of Brovelli (2012), where a 250 ms window (zero-padded to 1 s), was moved at 5 ms steps throughout each single trial, to construct multiple estimates of the CSD, where coherence and GC were determined from the average of these CSD estimates. To ensure that at least ten CSD estimates were averaged before computing coherence or GC, only trials with lengths of at least 300 ms were analyzed.

A standardized peak frequency was employed for all analyses where activity was assessed at a single spectral maximum. To establish this, coherence and GC were estimated over the entire 1.6 s of data. This was done for both parametric and non-parametric implementations. A Hann taper was used for the non-parametric estimation, otherwise all the spectral estimation parameters were identical to those outlined above. The peak GC and coherence were found to lie in the gamma band at 74 Hz for the parametrically derived estimates and 75 Hz for the non-parametric technique, which were subsequently used throughout, for the parametric and non-parametric analyses, respectively.

Statistical analysis

We determined the correlation between coherence (or GC influence) from two neighboring within-trial epochs, across trials, for each frequency–frequency combination. Correlations were computed either between conventional single-trial estimates or using the JC. This was done for all frequency–frequency combinations between 51 and 100 Hz. For the assessment of statistical significance of correlations, a Monte Carlo approach was employed and was identical for both the coherence and GC, and parametric and non-parametric cases. The test statistic used was the mean Spearman's rho computed over the jackknife replications from the early and late epochs for all the possible V1–V4 pairs. To construct the surrogate distribution, the JC between the early and late epochs was determined after the jackknife replications had been randomly paired, i.e. the trial order of the early epoch was randomized with respect to the late. Note that the random trial reordering was identical for each V1–V4 site pair. This was repeated 1000 times to form a null distribution of mean Spearman's rho values, which functions to disrupt the empirical relationship between the early and late epoch single trial pairs, so that their empirical degree of correlation can be compared with the distribution of correlation coefficients that occurred due to chance. When we computed JC on eight neighboring windows ranging from 0.1 to 0.8 s, this procedure was repeated for each of the eight epoch lengths, resulting in eight null distributions. When testing for cross-frequency interactions, the issue of multiple comparisons needed to be addressed. To correct for the multiple comparisons over the 50 × 50 frequency combinations, the largest absolute value of the correlation across all frequency–frequency combinations was selected for each of the 1000 permutations, resulting in a distribution of maximal test statistics (Nichols and Holmes, 2002). Empirical test statistic values were considered significant at $p = 0.05$, two-tailed, if their absolute value was larger than the 975th percentile of the distribution. Where p-values smaller than 0.05 are shown for visualization purposes, the tail of the distribution above the 975th value was extrapolated with a Generalized Pareto distribution function, an appropriate distribution for modeling the extreme values of a distribution.

A parametric approach was used to assess the statistical significance of a representative single channel pair over the eight neighboring epochs. Here we wished to show the precise p-value that corresponded with each rho-value, which was not feasible using a non-parametric approach since the p-values are sufficiently small that a Monte Carlo method is not computationally tractable to estimate these values. We used the standard approach, where the rho-value and number of trials are used to derive a t-statistic, which in combination with the corresponding degrees of freedom yields the p-value from Student's t-distribution (Rahman, 1968).

Application to neurophysiological data: results

Bosman et al. (2012) have established that areas V1 and V4 show robust gamma band coherence and bottom-up GC during sustained attention. It is well known that the correlation between neighboring time-points in a trial dissipates as the temporal distance between them increases (autocorrelation). We capitalize on this property to compare JC with single-trial methods for both parametric and non-parametric spectral estimators of the strength of V1–V4 gamma coherence and bottom-up GC influence. The logic is that neighboring windows should show correlated coherence and GC, which we can assess using JC. To achieve this, we calculated the correlation between the magnitude of gamma band coherence (and bottom-up GC influence) from two neighboring within-trial analysis windows, across trials. Fig. 8 shows the same characteristic pattern that resulted from the numerical simulations (Fig. 7), where the correlation coefficients increase as the data window is increased in length. As mentioned above, the two data windows were neighboring within a given trial, and one might therefore be concerned that longer windows included data temporally more adjacent, and therefore more correlated. To counter this potential effect, the windows were designed such that the end point of the first window coincided with the starting point of the second window, which results in longer windows possessing data that are temporally more distant. In agreement with the simulations, the JC curves for the average over all V1–V4 site pairs (Figs. 8A and B, parametric: blue lines, non-parametric gold lines) show a considerable improvement over conventional single trial approaches (Figs. 8A and B, parametric: red line, non-parametric: green lines). Fig. 9 shows an example V1–V4

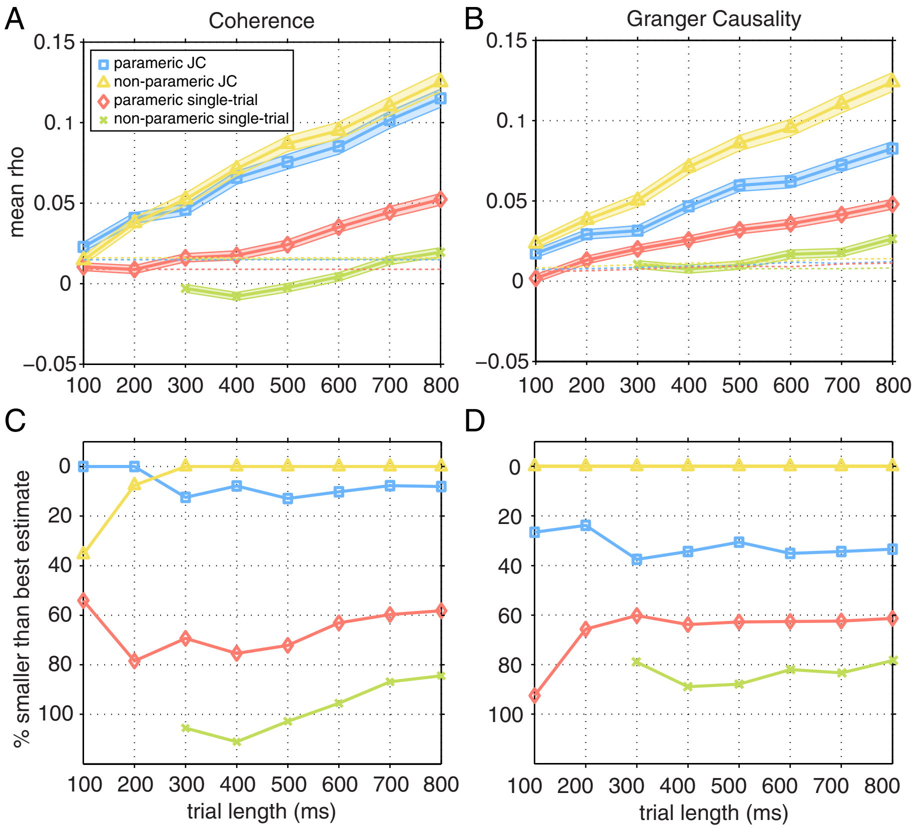

Fig. 8. Comparison of JC with conventional spectral correlation methods as a function of trial length for V1–V4 channel pairs. A, B. Spearman's rank correlation of the coherence (A) and bottom-up GC (B) between two neighboring time windows, averaged over V1–V4 site pairs. The shaded area indicates the standard error of the mean (s.e.m). Dashed lines denote the mean rho-value at which the statistical significance is 0.05, with each color corresponding to the respective correlation metric. C, D. Percentage that each correlation metric is smaller than the metric yielding the greatest rho-value at each trial length. Gold triangles depict non-parametric JC, blue squares depict parametric JC, red diamonds depict single-trial parametric estimation, while green crosses depict a window-based non-parametric approach.

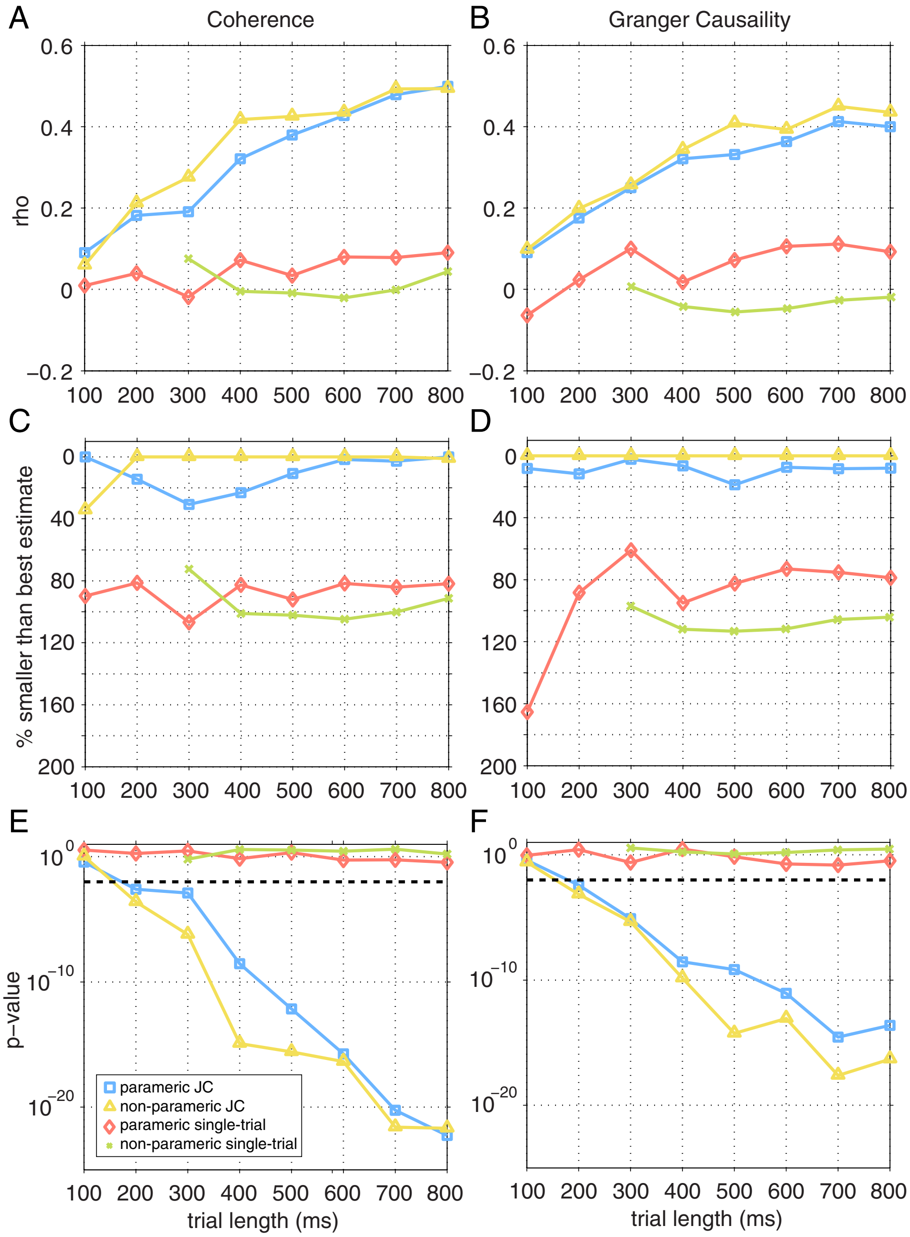

Fig. 9. Comparison of JC with conventional spectral correlation methods as a function of trial length for a selected V1–V4 channel pair. A, B. Spearman's rank correlation of the coherence (A) and bottom-up GC (B) between two neighboring time windows for a selected V1 channel and V4 channel pair. C, D. Percentage that each correlation metric is smaller than the metric yielding the greatest rho-value at each trial length. E, F. p-Values corresponding to each rho-value shown in A and B. Gold triangles depict non-parametric JC, blue squares depict parametric JC, red diamonds depict single-trial parametric estimation, while green crosses depict a window-based non-parametric approach.

pair of sites employing the same plotting conventions a Fig. 6. As in the group average, the JC correlation curves (Figs. 9A and B) show a marked increase over the conventional single-trial approaches. Figs. 9E and F reveal that JC is much more sensitive for revealing correlations. While conventional correlation approaches do not reach significance, JC is significant for window lengths of 200 ms and beyond, for both coherence and GC influence. These results demonstrate that for biologically/behaviorally interesting window lengths, the JC method substantially outperforms conventional approaches. It is also apparent that the parametric and non-parametric JC approaches provide similar results. Parametric JC was slightly superior for coherence on the shortest windows. Non-parametric JC was slightly superior for coherence at all other window lengths and considerably superior for GC at all window lengths (Fig. 8D).

For the shortest data window of 100 ms, we now apply the JC approach for all frequency–frequency combinations. Fig. 10 reveals significant correlation of the coherence for a range of frequencies surrounding the gamma band peak, both when determined non-parametrically (Fig. 10A) and parametrically (Fig. 10C). The precise spectral extent of the peaks is due to the specific choices of the parametric and non-parametric spectral estimation, i.e. the model order and the number of data tapers. The JCs of GC (Fig. 10B for non-parametric JC and D for parametric JC) show similar results. Fig. 10E shows a small significant region for parametrically determined single-trial coherence, while Fig. 10F shows no significant cross-frequency correlation for single-trial parametric estimates of GC, consistent with the numerical simulations (Fig. 6) and single-frequency analyses (Figs. 8 and 9).

Taken together, the empirical results demonstrate that over all window lengths tested, JC substantially outperforms conventional single-trial methods.

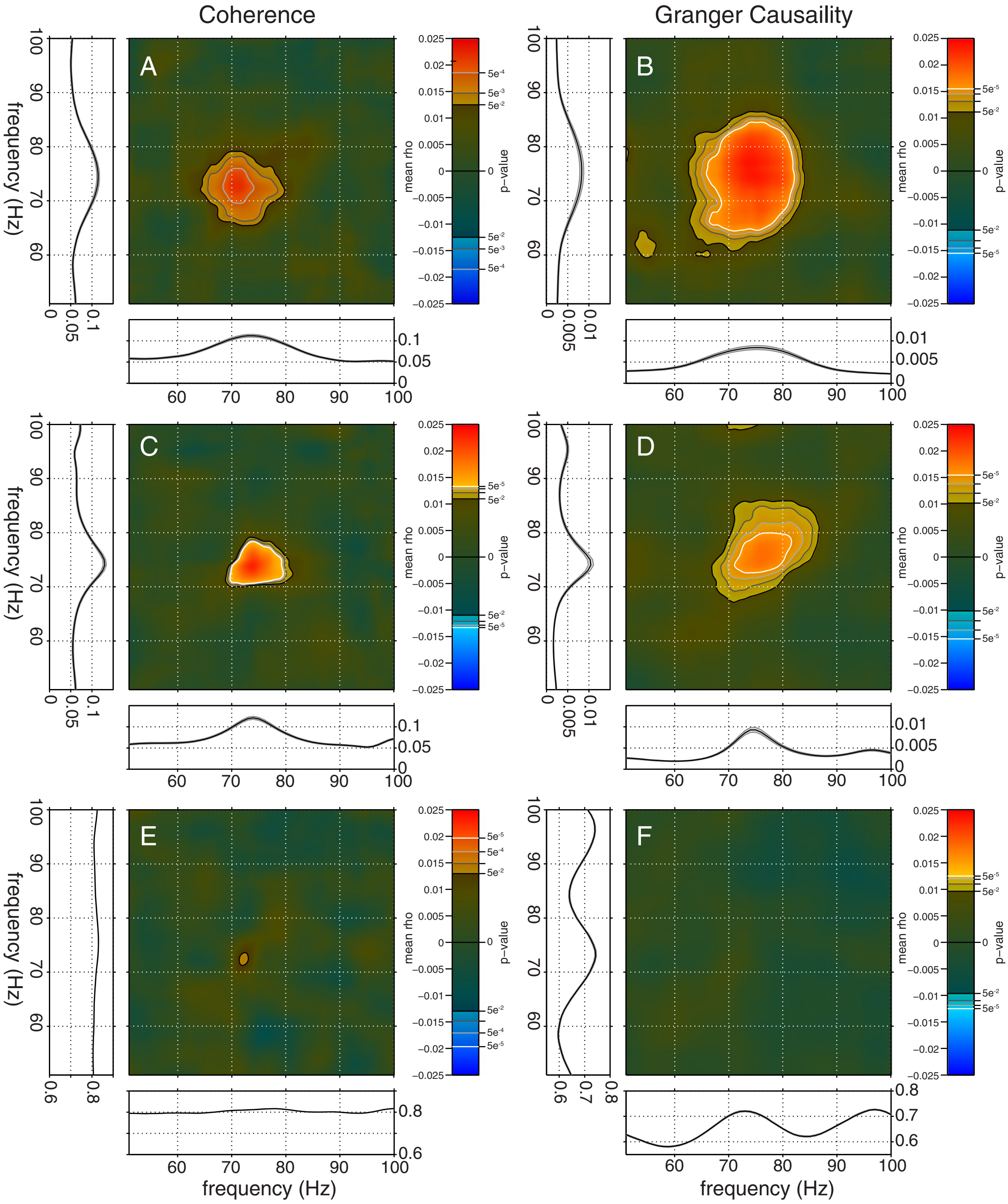

Fig. 10. Average correlation between two 100 ms neighboring windows over all V1–V4 pairs. A, B. Mean JC for coherence (A) and GC (B) computed from non-parametric estimates. C, D. Mean JC for coherence (C) and GC (D) computed from parametric spectral estimates. E, F. Conventional correlation computed from parametric single-trial estimates. For each panel (A–F), the central plot depicts the mean correlation over channel pairs for each frequency–frequency combination, with different shadings reflecting p-values of 0.01, 0.001, 0.0001, and 0.00001. Contour lines correspond with shading transitions, with the gray-scale value corresponding to the p-value, where white indicates the lowest p-value and black the largest. Spectral plots to the left of the frequency–frequency map correspond to the average spectrum over V1–V4 pairs of the earlier time window, while spectral plots below the map correspond to the average spectrum of the later time window. Colorbars indicate the mean correlation over pairs and the corresponding p-values, with shading and gray-scale lines following the same convention as the shaded areas and contour lines of the frequency–frequency maps.

Discussion

To summarize, we presented jackknife correlation (JC) that allows the relation of moment-by-moment fluctuations in correlation strength to other parameters, even though either correlation metric may not be defined on a moment-by-moment basis, i.e. on the basis of a single observation. We started out by investigating an approach that has been commonly used in the case of assessing correlation between a single-trial defined variable and an undefined variable, namely the sorting-and-binning approach. In this case, the single-observation-defined variable allows the sorting-and-binning, which in turn allows the calculation of the single-observation-undefined metric over the multiple observations in each bin. The sorting-and-binning approach is often used with overlapping bins in order to cope with limited numbers of observations. We demonstrated that the sorting-and-binning approach leads to correlation coefficients that depend on the choice of bin size and bin overlap and therefore can only be interpreted with these parameters in mind, which makes them difficult to compare across studies. Furthermore, we found that statistical power was actually maximal when correlations were determined across single observations, rather than across binned data. Since sorting-and-binning may be considered a form of factorial design, where bin is considered a factor, our numerical results support the arguments presented by Stahl and Gibbons (2004), that the correlative framework is indeed the more powerful approach. Moreover, when overlapping bins are used, a failure to control for the lack of independence between bins can lead to erroneous p-values with a dramatic overestimation of statistical significance. These difficulties and insights motivated the introduction of JC, which was shown to optimally address the above concerns.

The JC not only provides a quantitative improvement of estimation properties in comparison to the sorting-and-binning approach, but most critically, it allows for the extension of correlation to cases where neither variable is defined on the level of a single trial. While the sorting-and-binning approach always requires that one of the correlated variables be defined for single observations, the JC does not require this and therefore allows determination of the correlation between two single-observation-undefined metrics. This allows, for example, the investigation of whether the functional connectivity between brain areas A and B depends on the functional connectivity between brain areas C and D.

In the same vein, we note that the scope of the JC reaches beyond relating fluctuations in correlation strength. The JC can facilitate the investigation of relations for any metric that is defined only across multiple observations (or observation epochs) and that is a smooth function of the observations (i.e. leaving out one of many observations results in a correspondingly small change). For example, the variance is a smooth function that is defined only across multiple observations. The JC provides a straightforward approach to relating e.g. fluctuations in neuronal response variance to stimulus or task parameters, or even relating fluctuations in neuronal response variances between different brain areas. Additionally, use of the JC is not limited to electrophysiological data, but is equally applicable to all time-series analyses, such as that used in fMRI or in fields outside of neuroscience.

Here, we were particularly interested in frequency-resolved, i.e. spectral, analyses. The estimation of any spectral estimator, in order to define frequency, requires multiple observations to form an observation epoch of finite length. The epoch length in turn defines the frequency resolution of the spectral estimator. Spectrally resolved metrics of correlation, like coherence, when estimated at the maximal spectral resolution allowed by a given epoch length, are strictly not defined on the basis of a single observation epoch. This can in principle be overcome by either cutting individual epochs into multiple shorter epochs, by applying multiple orthogonal taper windows, or by fitting a parametric model with its typically relatively low order. Yet, all those approaches reduce the spectral degrees of freedom in some form, either by essentially downsampling the spectral resolution (in the case of cutting into segments), by rendering neighboring spectral estimates non-independent through spectral boxcar smoothing (multi-tapering), or by a reduction of the full spectral complexity of the data to a small number of model parameters (parametric model). Furthermore, short epochs cannot be subdivided in many sub-epochs, and metrics that require many epochs for a proper estimation will remain poorly estimated on the basis of few sub-epochs. We compared these approaches directly to the JC method. This demonstrated the superior performance of JC on data generated from a simulated system of coupled brain areas. This analysis was repeated on empirical data recorded from the macaque monkey, where again JC showed an enhanced ability to recover correlated trial-by-trial fluctuations in inter-areal connectivity metrics.

Acknowledgments

This work was supported by the Human Frontier Science Program Organization Grant RGP0070/2003 (P.F.), the Volkswagen Foundation Grant I/79876 (P.F.), the European Science Foundation (044.035.003) European Young Investigator Award Program (P.F.), the European Union (HEALTH-F2-2008-200728 to P.F.), the LOEWE program ("Neuronale Koordination Forschungsschwerpunkt Frankfurt" to P.F.), the Ministry of Economic Affairs and the Ministry of Education, Culture and Science, Netherlands (BrainGain to P.F., Robert Oostenveld. and Jan-Mathijs Schoffelen.), and the Netherlands Organization for Scientific Research Grant452-03-344 (P.F.).

References

Barnett, L., Seth, A.K., 2014. The MVGC multivariate granger causality toolbox: a new approach to Granger-causal inference. J. Neurosci. Methods 223, 50–68. http://dx.doi.org/10.1016/j.jneumeth.2013.10.018.

Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., et al., 2012. Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron 75 (5), 875–888. http://dx.doi.org/10.1016/j.neuron.2012.06.037.

Brovelli, A., 2012. Statistical analysis of single-trial Granger causality spectra. Comput. Math. Methods Med. 2012, 697610. http://dx.doi.org/10.1155/2012/697610.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. second ed. Lawrence Erlbaun Associates.

Dhamala, M., Rangarajan, G., Ding, M., 2008a. Estimating Granger causality from fourier and wavelet transforms of time series data. Phys. Rev. Lett. 100 (1), 018701.

Dhamala, M., Rangarajan, G., Ding, M., 2008b. Analyzing information flow in brain networks with nonparametric Granger causality. NeuroImage 41 (2), 354–362. http://dx.doi.org/10.1016/j.neuroimage.2008.02.020.

Ding, M., Bressler, S.L., Yang, W., Liang, H., 2000. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol. Cybern. 83 (1), 35–45.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc. 77 (378), 304–313.

Granger, C., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37 (3), 424–438.

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C.S., Bauml, K.-H., 2007. Prestimulus oscillations predict visual perception performance between and within subjects. NeuroImage 37 (4), 1465–1473. http://dx.doi.org/10.1016/j.neuroimage.2007.07.011.

Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., Varela, F.J., 2000. Studying single-trials of phase synchronous activity in the brain. Int. J. Bifurcation Chaos 10 (10), 2429–2439. http://dx.doi.org/10.1142/S0218127400001560.

Liang, H., Bressler, S.L., Ding, M., Truccolo, W.A., Nakamura, R., 2002. Synchronized activity in prefrontal cortex during anticipation of visuomotor processing. Neuroreport 13 (16), 2011–2015.

Maris, E., Schoffelen, J.-M., Fries, P., 2007. Nonparametric statistical testing of coherence differences. J. Neurosci. Methods 163 (1), 161–175. http://dx.doi.org/10.1016/j.jneumeth.2007.02.011.

Miller, R.G., 1974. The jackknife — a review. Biometrika 61 (1), 1–15. http://dx.doi.org/10.1093/biomet/61.1.1.

Miller, J., Patterson, T., Ulrich, R., 1998. Jackknife-based method for measuring LRP onset latency differences. Psychopysiology 35 (1), 99–115.

Mitra, P.P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. Biophys. J. 76 (2), 691–708. http://dx.doi.org/10.1016/S0006-3495(99)77236-X.

Morf, M., Vieira, A., Lee, D., Kailath, T., 1978. Recursive multichannel maximum entropy spectral estimation. IEEE Trans. Geosci. Electron. 16 (2), 85–94.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011. http://dx.doi.org/10.1155/2011/156869.

Park, H.-J., Friston, K., 2013. Structural and functional brain networks: from connections to cognition. Science 342 (6158), 1238411. http://dx.doi.org/10.1126/science.1238411.

Parr, W.C., 1985. Jackknifing differentiable statistical functionals. J. R. Stat. Soc. Ser. B 47, 56–66.

Quenouille, M.H., 1949. Approximate tests of correlation in time-series. J. R. Stat. Soc. B 11, 68–84.

Rahman, N.A., 1968. A Course in Theoretical Statistics. Charles Griffin and Company.

Stahl, J., Gibbons, H., 2004. The application of jackknife-based onset detection of lateralized readiness potential in correlative approaches. Psychopysiology 41 (6), 845–860.

Thomson, D.J., 1982. Spectrum estimation and harmonic analysis. Proc. IEEE 70, 1055–1096.

Tukey, J.W., 1958. Bias and confidence in not-quite large samples (abstract). Ann. Math. Stat. 29, 614.

Turk-Browne, N.B., 2013. Functional interactions as big data in the human brain. Science 342 (6158), 580–584. http://dx.doi.org/10.1126/science.1238409.

van Elswijk, G., Maij, F., Schoffelen, J.-M., Overeem, S., Stegeman, D.F., Fries, P., 2010. Corticospinal beta-band synchronization entails rhythmic gain modulation. J. Neurosci. 30 (12), 4481–4488. http://dx.doi.org/10.1523/JNEUROSCI.2794-09.2010.

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., Pennartz, C.M.A., 2010. The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. NeuroImage http://dx.doi.org/10.1016/j.neuroimage.2010.01.073.

Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. 15 (2), 70–73. http://dx.doi.org/10.1109/TAU.1967.1161901.

Wilson, G.T., 1972. The factorization of matricial spectral densities. SIAM J. Appl. Math. 23 (4), 420–426.

Womelsdorf, T., Fries, P., Mitra, P.P., Desimone, R., 2006. Gamma-band synchronization in visual cortex predicts speed of change detection. Nature 439 (7077), 733–736. http://dx.doi.org/10.1038/nature04258.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., Fries, P., 2007. Modulation of neuronal interactions through neuronal synchronization. Science 316 (5831), 1609–1612. http://dx.doi.org/10.1126/science.1139597.