Increasing robustness of pairwise methods for effective connectivity in Magnetic Resonance Imaging by using fractional moment series of BOLD signal distributions

Natalia Z. Bielczyk[1,2*], Alberto Llera[1,2], Jan K. Buitelaar[1,2],
Jeffrey C. Glennon[1,2], Christian F. Beckmann[1,2,3]

(1) Donders Institute for Brain, Cognition and Behavior, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

(2) Radboud University Nijmegen Medical Centre, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, the Netherlands

(3) Oxford Centre for Functional MRI of the Brain, University of Oxford, Oxford OX3 9DU, United Kingdom

∗ Correspondence: natalia.bielczyk@radboudumc.nl

August 15, 2017

Abstract

Estimating causal interactions in the brain from functional magnetic resonance imaging (fMRI) data remains a formidable task. Multiple studies have demonstrated that essentially all current analytical approaches to determine effective connectivity perform poorly even when applied to synthetic fMRI datasets. Recent advances in this field include methods for pairwise inference, which involve creating a sparse connectome in the first step of the inference, and then using a classifier in order to determine the directionality of every connection in the second step. In this work, we propose an advance to the second step of this inference procedure, by building a classifier based on all moments of the BOLD distribution combined into cumulants. The classifier is trained on a dataset generated under under the Dynamic Causal Modeling (DCM) generative model [12, 23]. First, we evaluated the performance of this classifier on synthetic benchmark datasets. Secondly, we simulated the DCM generative model with natural confounds present in the fMRI datasets such as varying signal magnitudes and noise variances between the upstream and downstream regions. Our approach outperforms other methods for effective connectivity research when applied to the benchmark datasets, but crucially, it is also more resilient to known confounding effects such as differential noise level across different areas of the connectome. This suggests that the classifier proposed in this work can be recommended for further validation on the human fMRI datasets.

Keywords: causal inference, effective connectivity, functional Magnetic Resonance Imaging, pairwise causal inference

1 Introduction

In the context of fMRI research, effective connectivity refers to the process of estimating causal interactions between distinct regions within the brain. Several characteristics of fMRI data impose severe restrictions on the possibility of estimating such effective connectivity using MRI. Firstly, the temporal resolution of the image acquisition in MR imaging is low (sampling rate $< 1[Hz]$). Furthermore, BOLD activity is delayed with respect to neuronal firing, with a delay of $3 - 6[s]$ in the adult human brain [1]. The delayed haemodynamic response can also induce spurious cross-correlations between two BOLD time series [20, 4]. Both subject-to-subject and region-to-region variability in the shape

of hemodynamic response [9] provide a general limitations to the methods for effective connectivity research in fMRI: when the HRF in one region is faster than in another, the temporal precedence of the peak of the hemodynamic response can easily be mistaken for causation. Secondly, fMRI data is characterized by a relatively low signal-to-noise ratio. Within grey matter and at field strengths of $3.0[T]$, the task-induced signal changes range within 2-3% of the mean signal depending on the task [17]. Furthermore, the stochastic noise in the brain has been shown to have a scale-free spectral characteristic [14, 3, 8] which additionally hinders identifiability of causal structures derived from fMRI data [4]. Moreover, typical fMRI protocols involve relatively short time series of a few hundred samples, which the possibility of improving signal-to-noise ratios through averaging across samples.

Multiple methods were proposed to estimate effective connectivities from fMRI data [11]. In the computational study by Smith et al. [23], a range of methods for effective connectivity were tested on synthetic datasets derived from a Dynamic Causal Modeling forward model (DCM, [12]). In this study, most methods for estimating causal interaction interactions remained at chance level. One method highlighted as relatively successful at identifying causal links is based on Patel's tau measure (PT, [19]). PT entails a two-step approach where the first step involves identifying the (undirected) connections by means of functional connectivity, and is achieved on the basis of correlations between the time series in different regions. This step results in a binary graph of connections, where all other possible edges identified as absent are disregarded from further considerations. The second step determines the directionality in each one of the previously detected connections. In this step, effective connectivity boils down to a two-node Bayesian network. The concept is based on a simple observation: if there is a causal link $X \to Y$, Y should get a transient boost of activity every time X increases activity. And vice versa: if there is a link $Y \to X$, X should react to the activation in Y by increasing activity. Therefore, one can threshold the signals $X(t)$, $Y(t)$, and compute the difference between conditional probabilities $P(Y|X)$ and $P(X|Y)$. Three scenarios are possible:

1. $P(Y|X)$ equals $P(X|Y)$: it is a bidirectional connection $X \leftrightarrow Y$ (since empty connections were sorted out in the previous step)

2. the difference between $P(Y|X)$ and $P(X|Y)$ is positive: the connection $X \to Y$ is likely

3. the difference between $P(Y|X)$ and $P(X|Y)$ is negative: the connection $Y \to X$ is likely

More recently, the Pairwise Likelihood Ratios (PW-LR, [15]) approach was proposed. It builds on the concept of PT. The authors improved on the second step of the inference by analytically deriving a classifier to distinguish between two models $X \to Y$ and $Y \to X$, which corresponds to the LiNGAM model [22] for two variables. The authors compare the likelihood of these two competitive models derived under LiNGAM's assumptions [16], and provide a cumulant based approximation to the likelihood ratio. This approach clearly outperforms all the previously tested methods on the synthetic benchmark datasets [15]. However, a single high-order moment of the BOLD distribution (either skewness or kurtosis, as proposed in PW-LR) is not robust as lower moments can also account for possible differences in local signal-to-noise ratio (SNR). The SNR magnitude can differ with respect to the various features of the underlying neuronal time series, and since the BOLD response might be impossible to deconvolve into the neuronal dynamics [21], representing the signal hidden in the BOLD by any single moment of a distribution is suboptimal.

Therefore, we further expand on the concepts of PT and PW-LR by proposing a classifier based on complex cumulants derived from multiple, possibly fractional, moments of the distribution of BOLD recordings. We compare the performance of our approach on

synthetic benchmark datasets [23] relative to other methods for effective connectivity used in fMRI research. Furthermore, we compare performance of the methods using simple two-node simulations generated from the DCM model but with varying signal magnitudes and noise variance between the upstream and the downstream node. We demonstrate that our classifier gives higher robustness against these confounds than other methods, and can therefore be recommended for validation on experimental fMRI datasets.

In section 2.1, we introduce the concept of complex moments of the distribution. In section 2.4, we describe the validation on synthetic benchmark datasets [23], and in section 3.1, we present the results of this validation on a representative simulation. Furthermore, in section 2.4, we describe an additional validation performed with use of the DCM generative model, but in presence of confounds such as a background noise and variability in hemodynamic responses between upstream and downstream region. In section 3.2, we present the results from this validation step, and in section 4, we critically discuss the results.

2   Materials and Methods

2.1   Fractional moment series of BOLD signal distributions

In this work, we propose to estimate causal links from BOLD recordings by analyzing the dependence between an expanded set of (fractional) moments of the BOLD distribution. We keep the same two-stage scheme for the causal discovery as Hyvärinen and Smith [15]. However, we improve on the second step of the causal inference - a two-node classification problem - by utilizing all moments of the BOLD fMRI distributions, and combining them into cumulants.

Fig. 1 A presents a graphical representation of a dynamical system with just two nodes. In this problem, one region ('upstream') is sending information to another region ('downstream') through a connection of weight A_{12}. Both regions receive region-specific signal $u_i(t)$ This signal can both relate to experimental input, as well as input from other regions. Both nodes also are influenced by a background neuronal noise $\sigma_i(t)$ that influence local signal-to-noise ratios.

The BOLD dynamics of such a simple two-node networks can be simulated with the DCM generative forward model [12]. Given the inputs to the network, the connection strength and sets of parameters characterizing local hemodynamic response within the nodes, the DCM generative model makes prediction on the BOLD dynamics in both nodes. The ensuing BOLD time series can then be normalized, and characterized in terms of its central moments:

$$M_k = \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i^k \tag{1}$$

where $k \in \mathcal{Q}$, $\tilde{x}$ - normalized time series, N - the length of BOLD time series. The novelty in this approach is that a discrete set of moment orders k typically used to characterize a distribution (in terms of mean, variance, skew etc), is converted into a (pseudo-) *continuous* dimension, by sampling moment order k from sets of rational numbers. These fractional moments of the distribution are not isomorphic with the moment generating function [5] as we do not convert the distribution of BOLD values into a probability density function at any stage [1]. Since the original BOLD time series is normalized to mean 0, it contains negative values and therefore, the fractional moments ($k \notin N$) will become complex. Since

---

[1]This conversion would require first reshaping the BOLD distribution so that it has an area of 1, which would require first converting the distribution into a histogram, and would be dependent on the binning of the histogram. Calculating fractional moments of the distribution does not have these problems as they are calculated directly from the BOLD time series

the Eq. 1 is continuous with respect to k, these complex moments will form a curve in the complex plane (Fig. 1 B).

In Fig. 1 B, we present a phase diagram for all moments in range $k \in [0.0, 5.0]$, for a long simulated BOLD fMRI time series generated in a simple two-node network in the noiseless case. The moments are computed separately for the upstream region (blue) and the downstream region (red). The curve starts at $(1, 0)$ for $k = 0$. Then, it traverses the upper half-plane and for de-meaned data arrives at $(0, 0)$ for $k = 1$. Subsequently, it goes back through the lower half-plane and comes back to $(1, 0)$ for $k = 2$ since the variance is equal to 1. Every time k becomes an integer, the curve crosses the real axis. Note that the imaginary axis characterizes the left half of the BOLD distribution, since fractional moments give nonzero imaginary part for negative values of distribution of the BOLD values.
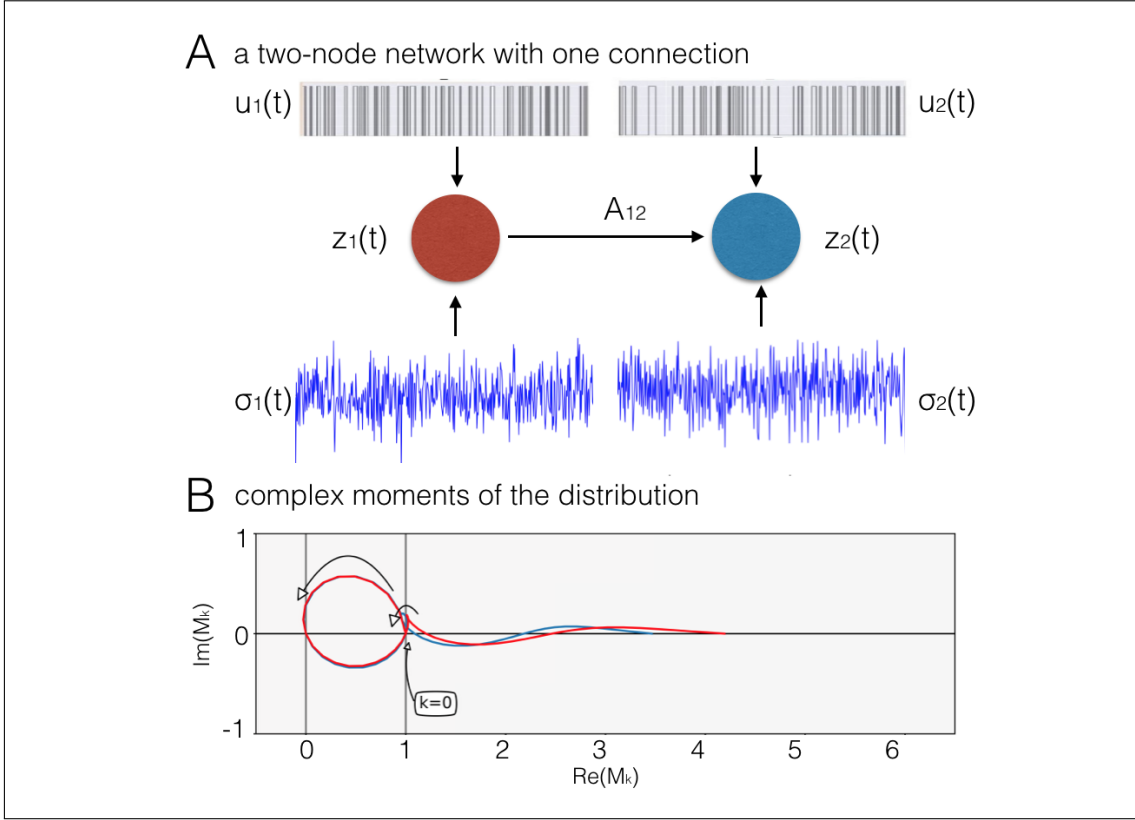

Figure 1: A two node network with one directed connection. **A**: The upstream node, $z_1(t)$, is sending information to the downstream node, $z_2(t)$, through a single connection of weight A_{12}. Both regions received a binary signal $u_i(t)$ and neuronal noise $\sigma_i(t)$. The proportion between the amplitude of $s_i(t)$ and the variance of the noise $\sigma_i(t)$ defines the SNR in the network. **B**: All the fractional moments for $k \in [0, 5]$, for the BOLD fMRI time series from a simulated 2-node network, in the noiseless case. Blue: upstream region. Red: downstream region. The curve starts from $(1, 0)$ for $k = 0$, traverses the upper half-plane and arrives at $(0, 0)$ for $k = 1$. It travels back through the lower half-plane towards $(1, 0)$ for $k = 2$ as the BOLD variance was fixed to 1 through the normalization. Every time k becomes an integer, the curve crosses the real axis.

2.2 Complex cumulants of the distribution

For two time series $x(t)$, and $y(t)$, not only the sole fractional moments but also the *asymmetry* between the moments can indicate the directionality of a connection. This

4

asymmetry can be quantified by 'fractional cumulants':

$$C_{kl} = \frac{1}{N} \sum_{i=1}^{N} (\tilde{x}_i^k \tilde{y}_i^l - \tilde{y}_i^k \tilde{x}_i^l) \tag{2}$$

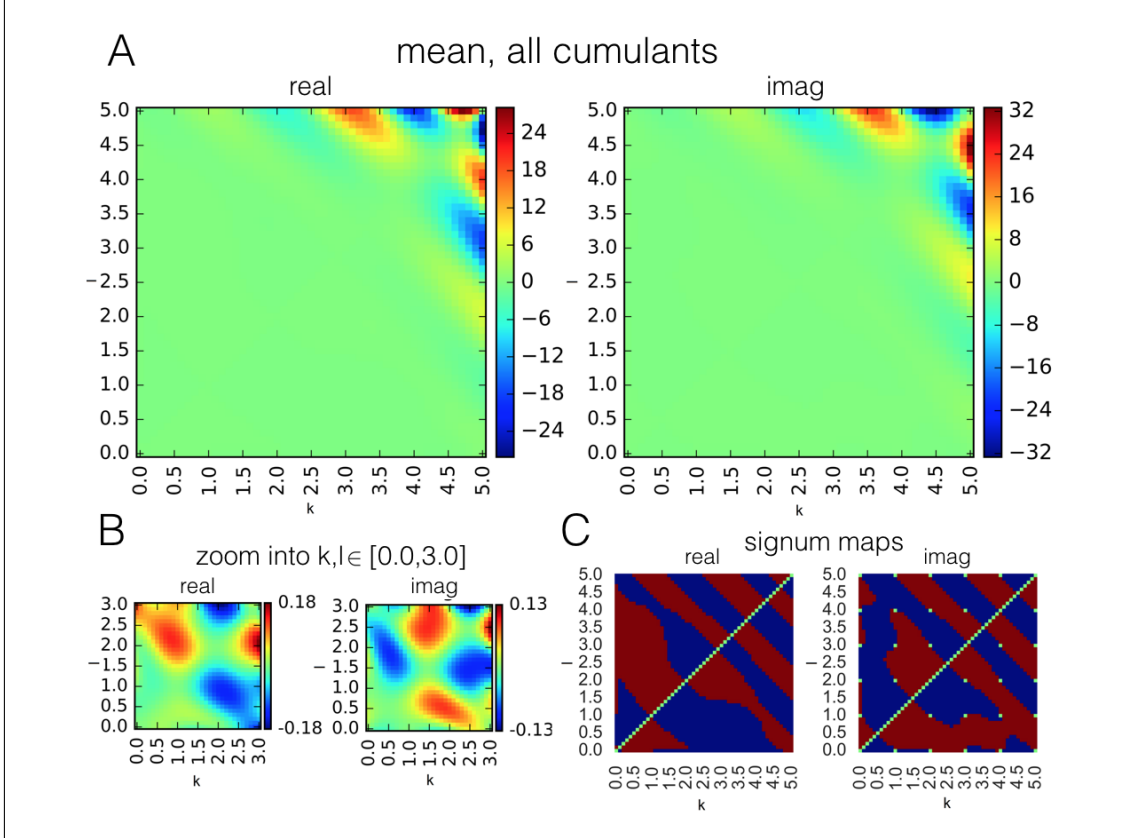where $k, l \in \mathcal{Q}$.

Figure 2: **A**: mean values for all cumulants, over 1000 simulations of a two node network with one connection (Fig. 1), for $k, l \in [0.0, 5.0]$. Since cumulants are antisymmetric with respect to indexing k, l, the heatmaps for the real and the imaginary component are also antisymmetric. **B**: a zoom of **A** into a smaller range of $[0.0, 3.0]$. The imaginary component becomes a zero for cumulants of integer orders: $k, l \in \mathcal{N}$, but exchanges the sign in the intervals between in a systematic way. **C**: sign of the cumulants in the majority of 1000 instantiations of the generative model. Red: positive. Blue: negative. Green: zero. We further denote the 'signum maps' for real and imaginary components as Sr and Si.

In this particular problem, $x(t)$, $y(t)$ denote the BOLD time series in the two-node system. In order to make a prediction ('upstream' versus 'downstream') we learn the dependencies between moment time series using simulations on the basis of the DCM generative model. We have run 1,000 2-node DCM simulations with $Fs = 200[Hz]$ for a duration of $10[min]$. In order to marginalize out the influence of the haemodynamic parameters from our results, we sampled the parameters independently for the two nodes, and from the empirical distributions [12]. In order to marginalize out the effect of different input strengths and frequencies, we also sampled the input magnitudes and frequencies (probabilities of switch from on- to off- state and vice versa) from a Gamma distribution with mean and variance of 1. The input signals driving the upstream and the downstream region were also sampled independently from each other, as trains on- and off- states governed by Poissonian processes. The background neuronal noise was set to 0 in these

5

simulations. In order to obtain a precise estimation of fractional cumulants, we did not sub-sample our synthetic BOLD every $2 - 3[s]$ as is typically done for the synthetic BOLD fMRI.

We performed this simulation twofold. Firstly, we fed in an empty connection in order to create a null distribution of cumulants. Secondly, we added a connection with a weight of 0.9 to the 2-node system. In Fig. 2, we demonstrate the *mean* values for all cumulants of indexes $k, l \in [0.0, 5.0]$ obtained from the simulations of a connection, in Cartesian coordinates.

Fig. 2 A presents the mean values for all cumulants in range $k, l \in [0.0, 5.0]$, over 1000 simulations of a two node network with one connection (Fig. 1). Fig. 2 B presents the same maps are zoomed into the range $k, l \in [0.0, 3.0]$. Finally, for every cumulant, Fig. 2 C, presents the sign of this cumulant for the *majority* of 1000 instantiations of the simulation (we further refer to these binary maps as Sr and Si).

These binary maps do not represent confidence intervals, or discriminability, for particular cumulants, as 'majority' could mean 51% as well as 100% simulations. In order to choose cumulants which can best discriminate between a 'connection' and 'no connection' case, we created distributions of cumulant values across the 1,000 simulations in the null case and compared against the distributions derived from simulations with a connection. We smoothed these distributions with kernel smoothing function, and for each cumulant, we computed the percentile of samples falling beyond 95th percentile of the null distribution (in case the mean for the given cumulant is negative as in Fig. 2 C, we took samples falling lower than the bottom 5th percentile of the null distribution, and higher than the 95th percentile otherwise).

The results of the discriminability analysis in Cartesian coordinates are shown in Fig. 3. We can observe that whenever one of the indexes k, l equal to zero, i.e., the cumulant reduces to a simple moment, it has lower discriminative value than the full cumulants. Therefore, we will disregard moments from further analysis and fully concentrate on full cumulants, i.e., asymmetry between moments $(k, l > 0)$.
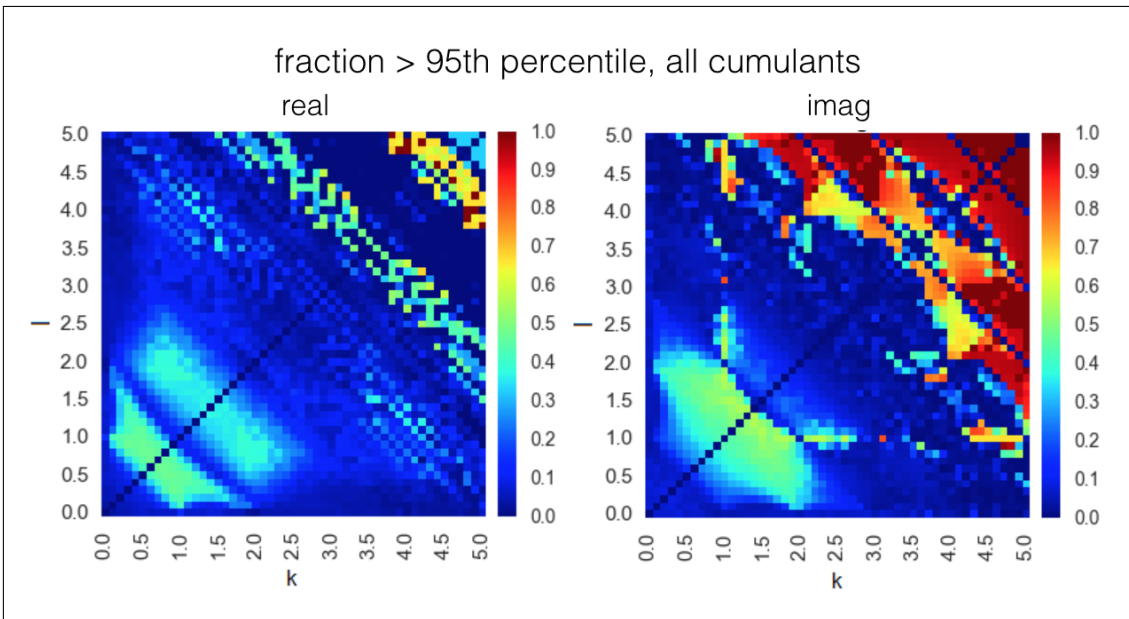

Figure 3: Discriminative power for all cumulants in range $k, l \in [0.0, 5.0]$, in the ideal case of a very long BOLD time series and no background neuronal noise.
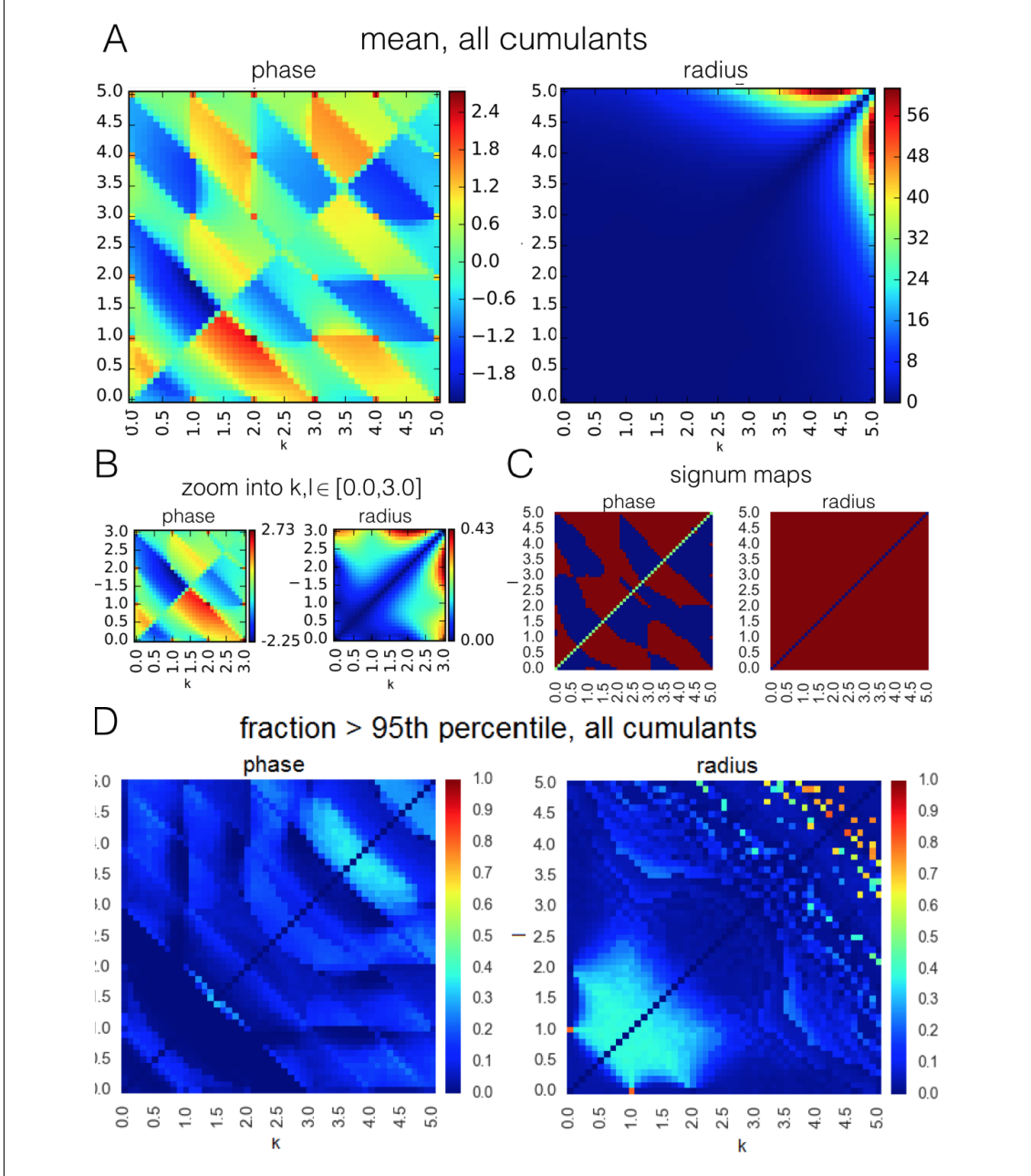
Figure 4: Cumulants in polar coordinates. **A**: mean values for all cumulants, over 1000 simulations. The radius and the phase are computed with respect to point $(0,0)$, otherwise the (anti-)symmetry would be lost. Since cumulants are antisymmetric with respect to indexing k, l, the heatmaps for the phase are antisymmetric. The radius is always positive. **B**: zoom into a smaller range of $[0.0, 3.0]$. **C**: signum of the cumulants. Red: positive. Blue: negative. Green: zero. **D**: Discriminative power for all cumulants in range $k, l \in [0.0, 5.0]$, in polar coordinates (in the ideal case of a very long BOLD time series and no background neuronal noise). Cumulants in polar coordinates are less informative. Phase is uninformative since the discriminative power is relatively high only in the high-moment regime - and high moments are hard to estimate on the short time series. Radius is an uninformative variable as it is always positive.

Interestingly, the range of high discriminability is different for real and imaginary components. For the real components, all the important information seems to be contained within the low range of $k + l < 2.0$, whereas maximal discriminability in the imaginary cu-

mulants exists in the range of $k + l > 5.0$. This different characteristic along the imaginary axis illustrates that moving from integer to fractional moments of the distribution provides with additional predictive power. Note also that, as we are deriving the discriminability maps from the DCM generative model, these maps are fingerprints of the particular problem of effective connectivity in fMRI; When derived from another generative model simulating another dataset, the maps would be different.

We also performed analogous analysis in polar coordinates, however it did not give a substantial improvement to the classification performance. The results of the analysis are presented in Fig. 4.

Now, in order to investigate how the performance of classification based on single cumulants changes when a single connection is embed in a bigger network, we evaluated their success rate in estimating connectivity for benchmark synthetic datasets [23]. Fig. 5 presents the grand mean success rate achieved by every cumulant *separately*, across all 28 benchmark simulations.

The success rate for each of the 28 separate simulations is presented in Fig. 2.2. The maps of simulation-dependent success rate relate to the maps of discriminative power (Fig. 3), but they are not identical and differ between simulations. One difference is that for the cumulants of high indexes $k + l > 4.0$, the success rate is not as high as the discriminability presented in Fig. 3 would suggest. This is because Fig. 3 represents the limit of a system of two isolated nodes with infinite SNR, and a very long BOLD time series, whereas benchmark simulations refer to more realistic case when for each pair of nodes, the time series is short, there are confounding signals from other nodes in the network, and there is certain degree of noise in the communication. Altogether, these factors cause that the high moments are hard to estimate in practice.
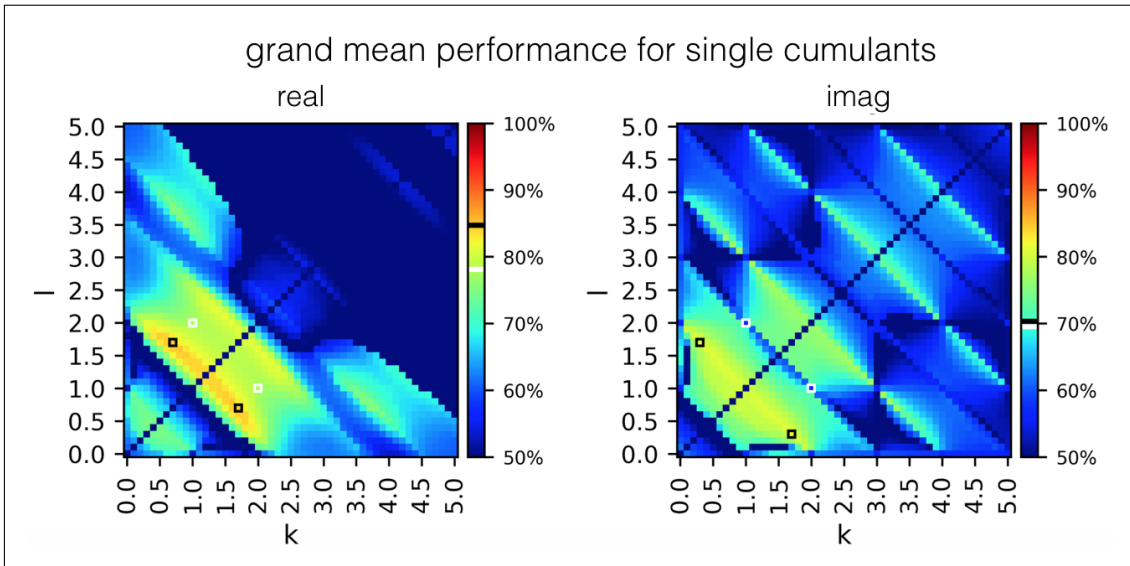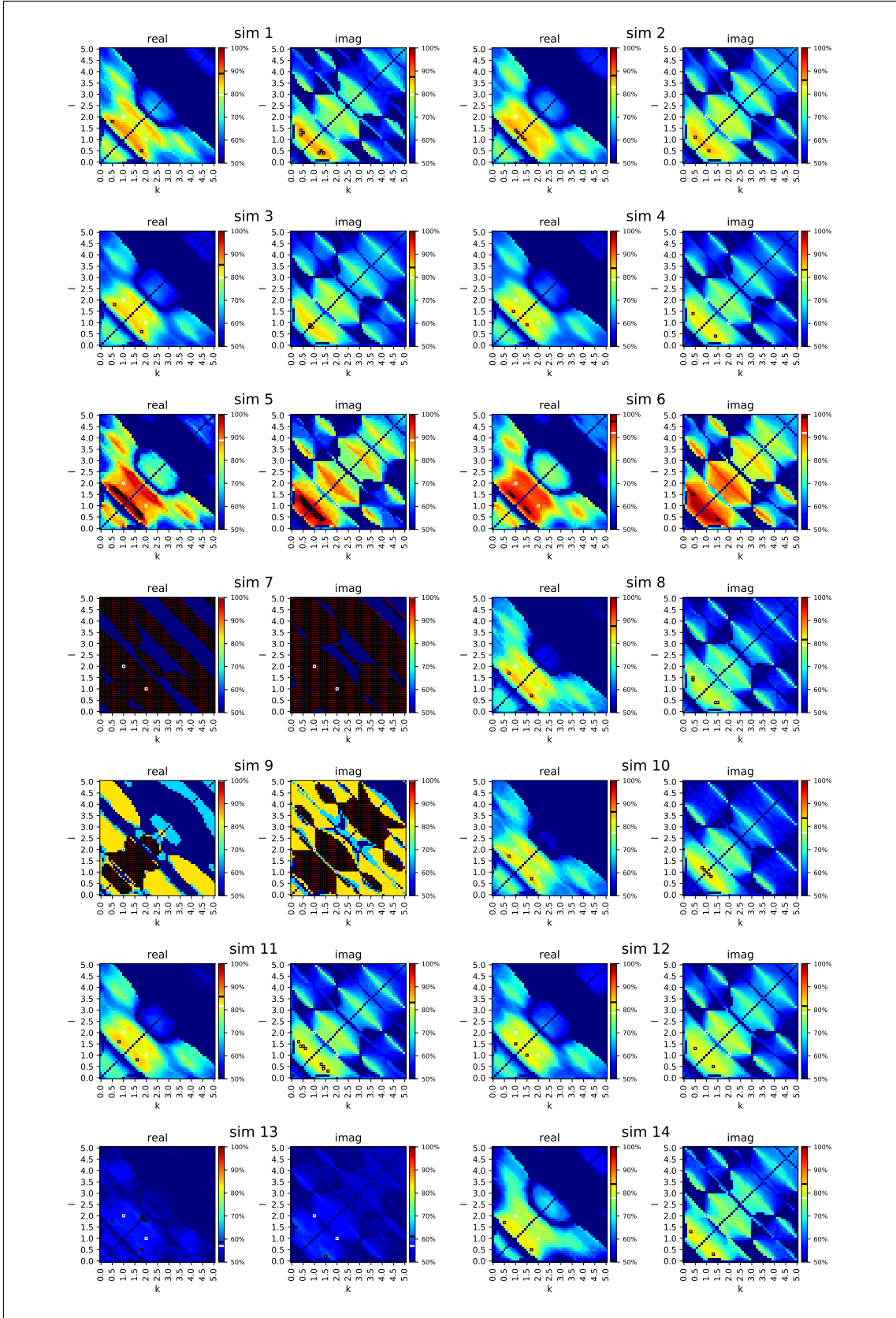

Figure 5: Success rate for all the individual cumulants, averaged over 28 simulations from the synthetic benchamrk datasets [23]. White-edged square denotes a single cumulant used by Hyvärinen and Smith [15]. The performance of this cumulant is shown in the colorbar, white band. Black-edged squares denote cumulants which give the highest performance on this dataset. Their performance is presented in the colorbar, black band. The high success rate is not preserved for high-indexed cumulants which achieved high discriminability on 2-node simulations. The maximal grand mean performance equals 0.847 for the real components, and 0.814 for imaginary components.
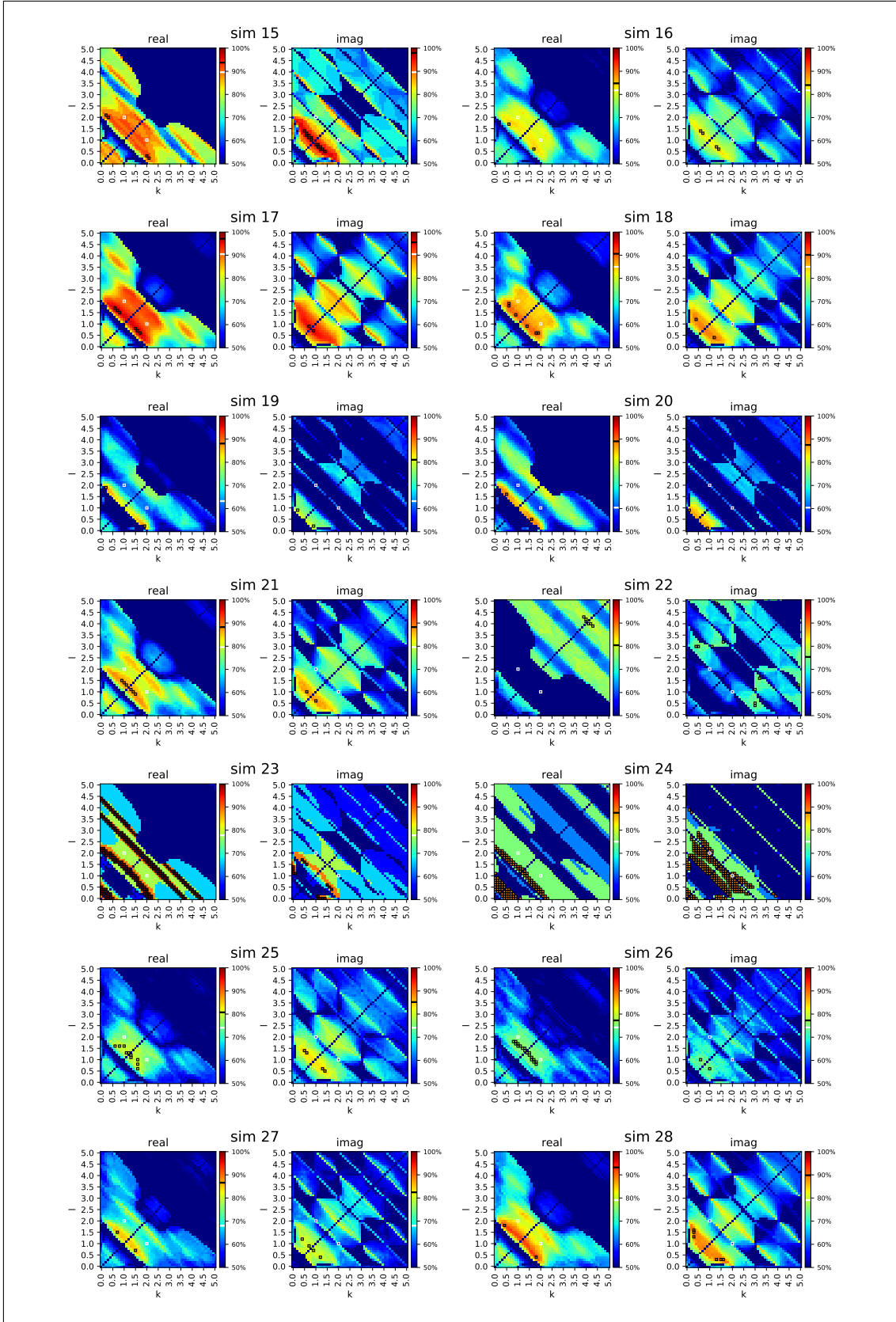
Figure 6: Success rate for all the individual cumulants, for the 28 simulations in benchmark datasets. In every case, there are cumulants performing better than the cumulant (2,1). Simulation no 7 is particularly easy for the causal inference whereas simulation no 13 is particularly misfortunate.

2.3 Combining fractional cumulants into a classifier

We propose to combine information contained in multiple cumulants by building the classifier based on a 'voting' between the cumulants. This classifier determines whether the map of cumulants obtained for a pair of time series $X(t)$, $Y(t)$ is closer to the benchmark maps presented in Fig. 2 A (which is an evidence for a connection $X \rightarrow Y$), or their inverse (which is an evidence for a flipped connection $Y \rightarrow X$). Each of the cumulants C_{kl} votes due to sign $Sr_{k,l}$, $Si_{k,l}$ (Fig. 2 C). Namely, if the sign of the cumulant is the same as in Fig. 2 C, it adds to the evidence for a connection $X \rightarrow Y$, and against this connection otherwise. Since in realistic conditions (short datasets, large TRs), high index cumulants, $k + l > 3$, yield the aforementioned estimation problem, we discount their contribution in the voting by using a nonlinearity of a form[2]:

$$f(x)=\log(\cosh(\max(x,0)))$$ (3)

Therefore, the final classifier yields:

$$\begin{cases} X \rightarrow Y & \text{if } \sum_{k,l}[\text{Sr}_{k,l}\text{f}(\text{Cr}_{k,l}) + \text{Si}_{k,l}\text{f}(\text{Ci}_{k,l})] \geq 0 \\ Y \rightarrow X & \text{otherwise} \end{cases}$$

2.4 Supervised learning using synthetic benchmark datasets

We know that (1) cumulants differ with respect to discriminability (Fig. 3), (2) the success rate of cumulants differs depending on the range $k + l \leq \text{Ind}_{max}$ (Fig. 5). Therefore, we optimize the performance of the classifier with respect to these two dimensions, by finding a combination which gives the best grand mean performance across the 28 simulations from the synthetic benchmark datasets, as they represent the variety of experimental conditions encountered in real-life fMRI setups.

Firstly, we fix $\text{Ind}_{max} = max(k,l) = 3.1$, and consider cumulants on a triangle $k, l \geq 0.1$, $k + l \leq \text{Ind}_{max}$. We then choose only cumulants of discriminability exceeding a particular value to be fed into the classifier. For instance, cutoff value of 0.1 means that we include the vote from all cumulants for which the discriminative value is not less than 0.1. We can then evaluate the grand mean success rate (as the mean success rate over all 28 benchmark simulations) in the function of the thresholding discriminability value. Fig. 7 A, demonstrates that including *all the cumulants* with a positive discriminative value (all cumulants except for $k = l$, for which discriminability is zero) gives the best classification performance.

Secondly, we optimize the window Ind_{max} for indexes k, l and compare between classifier with and without weighting with the discount function introduced in Eq. 3. Since discriminability is generally higher for low indexes k, l (Fig. 3), we will evaluate the grand mean performance based on cumulants of indexes between 0 and a maximum Ind_{max}, in the function of that maximum. We consider the maximal indexes along real and imaginary dimension separately. The results are presented in Fig. 7 B. The optimal performance in unweighted case equals 0.835 for $[\text{IndR}_{max}, \text{IndI}_{max}] = (2.4, 1.7)$, and 0.886 for $[\text{IndR}_{max}, \text{IndI}_{max}] = (2.1, 3.7)$ in weighted case, which exceeds both the grand mean performance of the 'PW-LR skew r' method by Hyvärinen (0.845) and the maximal grand mean performance of any single cumulant in our study (Fig. 5, the maximum of 0.847).

---

[2]similar function was proposed to discount the outliers in the BOLD time series in [15]
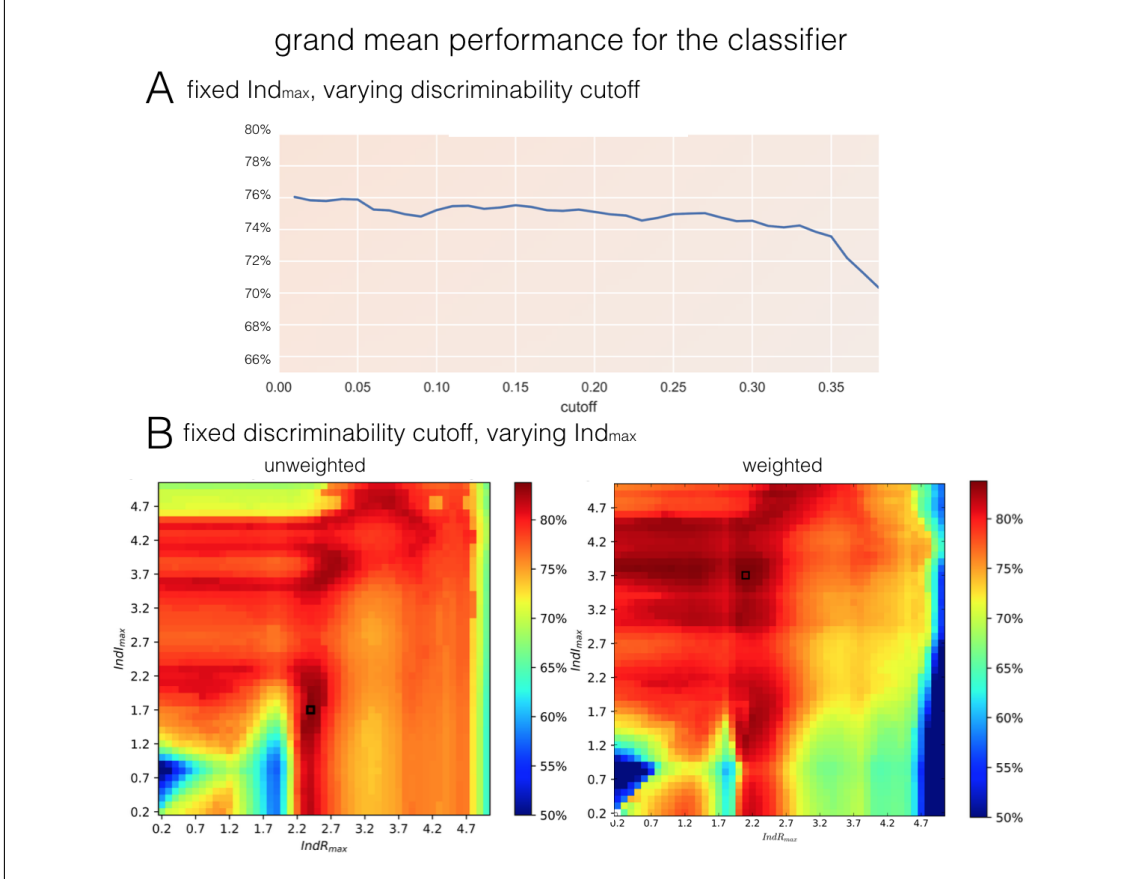
Figure 7: Dependence of the grand mean performance on synthetic datasets on the choice of cumulants. **A**: Grand mean performance for unweighted cumulants in range $k + l \leq 3.1$, in the function of the cutoff discriminative value. The higher cutoff, the less cumulants we take into account while voting for the directionality of the link. The results clearly show that in order to maximize the success rate in estimating effective connectivity, we should all the cumulants should be taken into account, except for the diagonal of $k = l$. **B**: the grand mean performance based on cumulants of indexes k, l between 0.1 and $k + l \leq \mathrm{Ind}_{max}$, in the function of that maximal index. The optimal performance in unweighted case equals 0.835 for $[\mathrm{IndR}_{max}, \mathrm{IndI}_{max}] = (2.4, 1.7)$, and 0.886 for $[\mathrm{IndR}_{max}, \mathrm{IndI}_{max}] = (2.1, 3.7)$ in weighted case.

Then, we compare the performance of this weighted classifier against the following methods:

1. Granger Causality (GC, [13, 21]), which infers effective connectivity between a pair of time series, assuming that both of them can be expressed as autoregressive processes

2. Partial Directed Coherence Partial Directed Coherence (PDC, [2]), which is effectively an equivalent to Granger causality in frequency domain. We used PDC implementation from the Extended Multivariate Autoregressive Modelling Toolbox

3. Patel's tau (PT, [19]), as described in the Introduction. Patel's tau, implemented similarly as in [23]: by recalculating each time series into the range $[0, 1]$, setting samples under the 10th percentile to 0, over the 90th percentile to 1, and linearly mapping the remaining samples to the range $[0, 1]$. Then, we infer the directionality of connection from the difference between $P(X|Y)$ and $P(Y|X)$. In addition to the

previous implementation however, we also integrate the results over all the possible thresholds in order to eliminate the thresholding problem while calculating the conditional probabilities $P(X|Y)$, $P(Y|X)$

4. Pairwise Likelihood Ratios methods (PW-LR), as described in the Introduction[3]

As in [15], we performed the first step of the inference as inverse covariance thresholded with permutation testing.

2.5   Testing robustness of the methods against confounds

In addition to evaluating our approach against the existing simulations from [23], we further evaluate the performance under additional yet typical modes of variation in the data. Specifically we are interested in characterising the discriminative performance relative to (i) more complex forms of stochastic noise in the data and (ii) unequal levels of SNR per node.

The benchmark synthetic datasets involve temporally uncorrelated, white background noise of a low magnitude on the neuronal level [23]. This type of noise is not physiologically plausible, as it is known from physiological studies that in the neuronal networks, the background noise has a scale-free power spectrum [14, 3, 8, 4]. Therefore, we simulated a two node system and introduced scale-free (pink) noise to the system. Then, we varied the variance of the noise in the range of $[0.2, 5.0]$ while keeping the amplitude of the inputs $s_i(t)$ fixed to 1.0. We performed 500 realizations of $10[min]$ simulation at high temporal resolution of $Fs = 200[Hz]$, for each configuration of the noise variances.

Furthermore, in the original version of the DCM procedure [12], as well as in most computational studies [23], equal stimulus strengths to both nodes $s_i(t)$ are assumed. This assumption might not hold true in the real fMRI datasets. Therefore, we performed another, noiseless simulation, in which we varied signal strengths between the upstream and downstream region. We performed 500 realizations of $10[min]$ simulation at high temporal resolution of $Fs = 200[Hz]$, for each configuration of input strengths in the range of $[0.2, 5.0]$.

---

[3]Hyvärinen used the cumulant $k, l = (2, 1)$ weighted with covariance for synthetic benchmark datasets. Therefore, for this comparison, we use the classifier based on fractional cumulants weighted with covariance. For our study, we chose a 'PW-LR r skew' version of the method, which involves inference based on a third cumulant with a discount for outliers [15]

3 Results

3.1 Supervised learning using synthetic benchmark datasets

The best version of the classifier was obtained for voting between cumulants in the range $[IndR_{max}, IndI_{max}] = (2.1, 3.7)$, with the discount for high moment indexes (Eq. 3). The comparison of this classifier against four other methods, GC, PDC, PT and 'PW-LR r skew', on the benchmark simulation no 2 is presented in Fig. 8. The violin plots denote the distribution of the z-scores for connections as compared to the null distribution. Blue dots denote the percentage of correct assignments for the true connections, as in [23]. In most of the other 27 benchmark datasets, our classifier outperforms all the other methods (Fig. 3.1). As in the original study by Smith et al. [23], lagged methods, GC and PDC, perform worse than the structural methods. 'PW-LR r skew' and fractional cumulants both outperform PT, most probably because PT is based on the thresholded signal and therefore contains a free parameter. In general, in the benchmark synthetic datasets, fractional cumulants outperform all other techniques in almost all cases, although the difference between performance of the fractional cumulants and PW-LR methods is small.
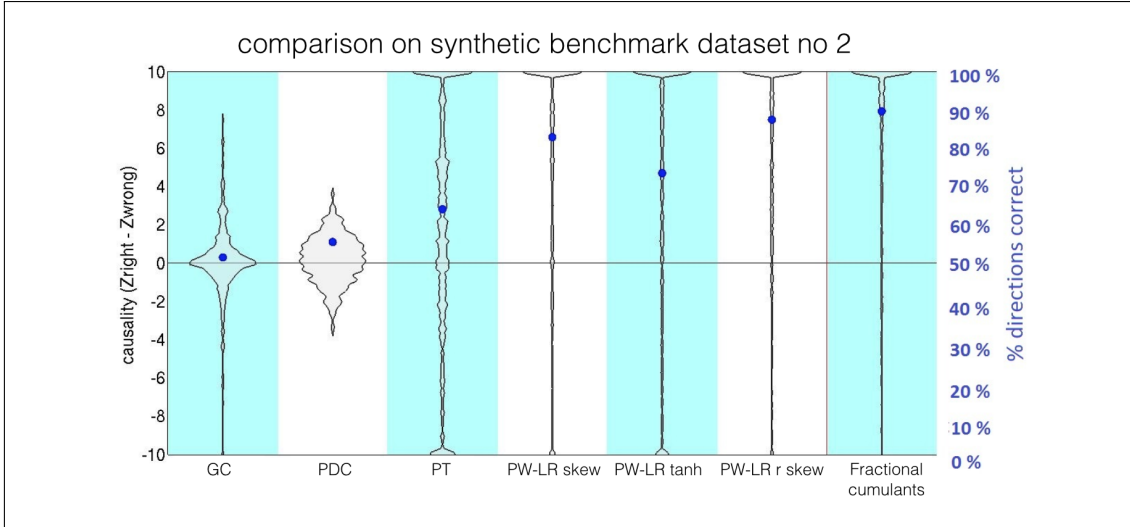

Figure 8: Comparison between the classifier based on the fractional cumulants and several other methods, on benchmark simulation no 2. The violin plots denote the distribution of the z-scores for connections as compared to the null distribution. Blue dots denote the percentage of correct assignments for the true connections [23]. The difference in performance between the classifier based on fractional cumulants and 'PW-LR r skew' [15] is small.
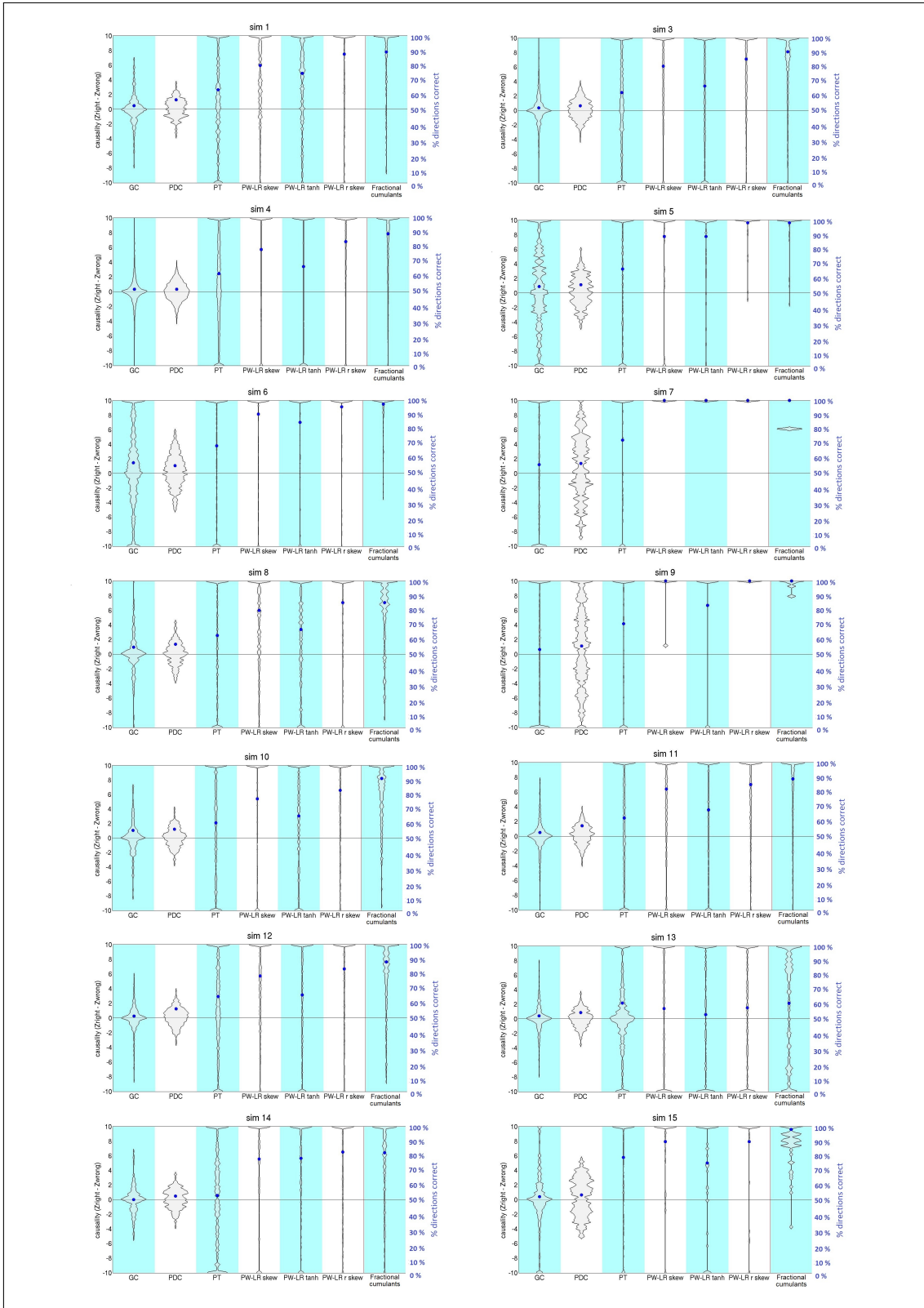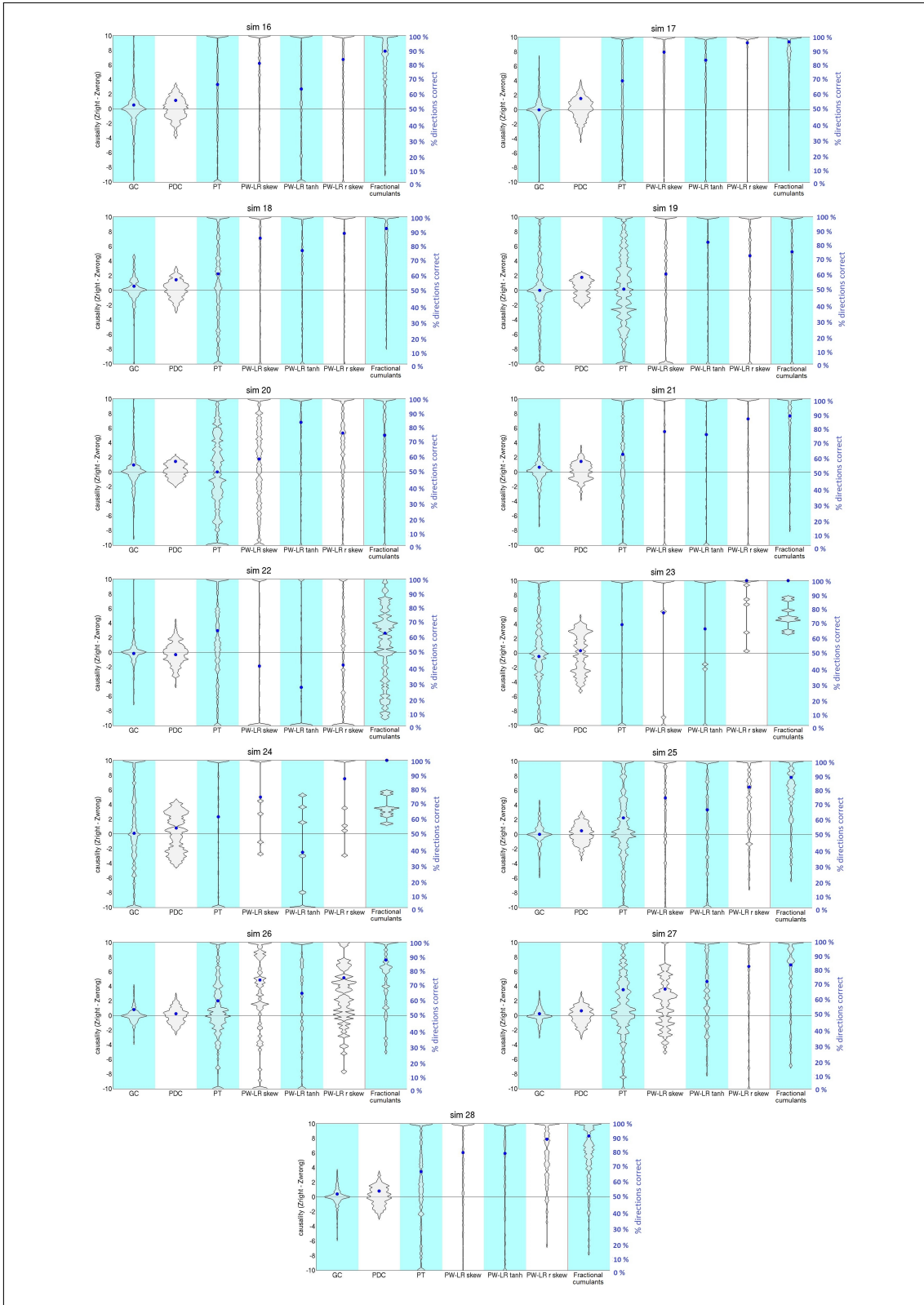
Figure 9: Overall success rate of Fractional cumulants against other methods. In most of the simulations, we achieved a slight improvement with respect to the main competitor, 'PW-LR r skew'. In simulations 7, 8, 9, 14, 17, 20 and 23 the performance is roughly the same.

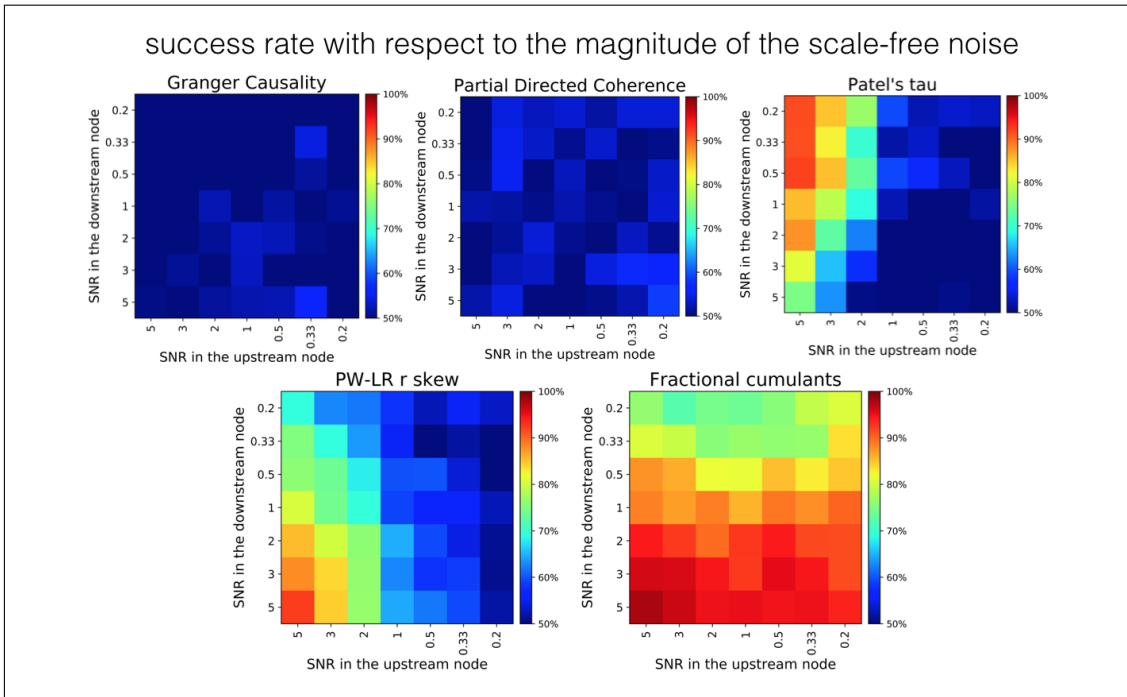## 3.2 Robustness of the methods with respect to confounds

Figure 10: Robustness of the methods against the background scale-free neuronal noise. The variance of the background noise $\sigma(t)$ differs between upstream and downstream region in the range of $[0.2, 5.0]$. As signal magnitude is constant and equal to 1 in these simulations, the signal-to-noise ratio (SNR) was calculated as $\frac{1}{\mathrm{var}(\sigma)}$. Only the classifier based on fractional cumulants gives a performance higher than chance across the whole parameter space.

Fig. 10 presents the comparison between fractional cumulant classifier and various other methods on a 2-node simulation with varying levels of physiologically plausible, scale-free neuronal noise. The results suggest that all previously tested methods show low levels of robustness towards such additional sources of variability in the data. GC as well as PDC are at the lowest performance, and give results on a chance level across the whole parameter space. PT seems to be fully resilient to the background noise in the downstream node, but not in the upstream noise. The performance of 'PW-LR r skew' drops down to the chance level with respect to both the noise level in the upstream and downstream region, whereas GC and PDC are performing almost equally poorly under any combination of noise variances (probably because the variance of the noise from the fitted autoregressive model is used to establish the directionality of the causal influence in GC).

Fig. 11 presents the comparison between the classifier based on fractional cumulants and other methods, given noiseless simulation and varying signal magnitudes. The classifier based on fractional cumulants is the only method where the performance does not descent towards chance level within the chosen parameter space. GC and PDC give performance around the chance level across the whole parameter space, whereas 'PW-LR r skew' and PT exhibit certain resilience towards this variability in the inputs. However, the performance breaks down towards the chance level at for the disproportion of the inputs higher than 3.0.
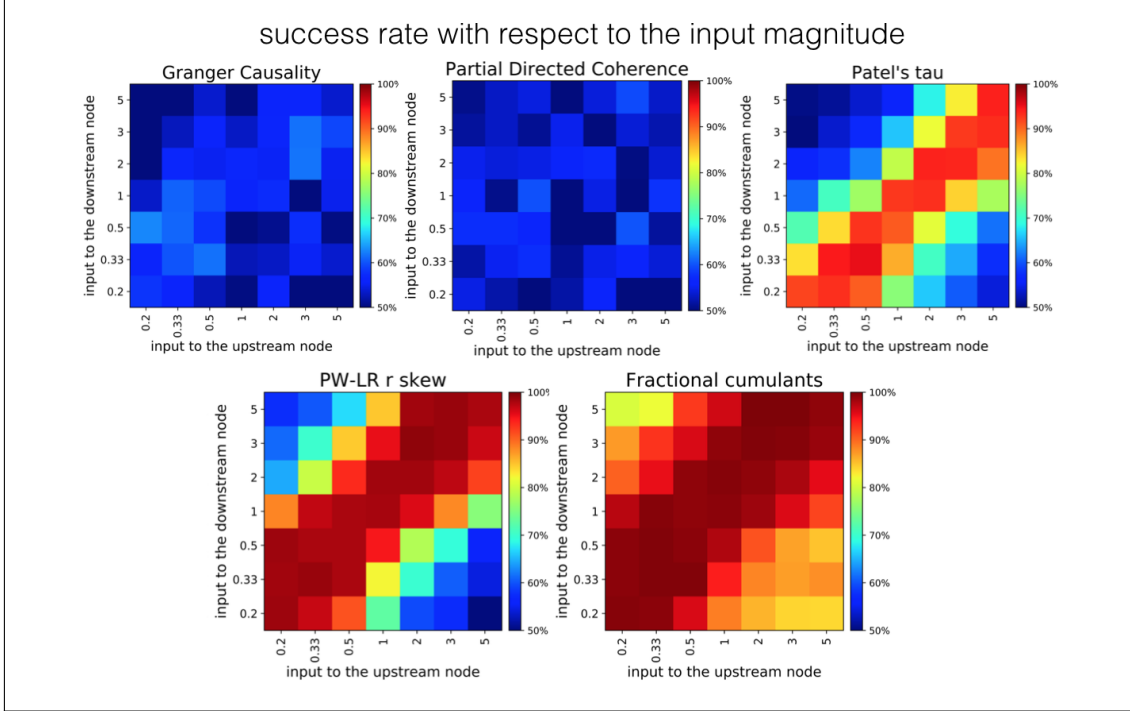
Figure 11: Robustness of different methods to the change in signal strengths. The variance of the signal differs between upstream and downstream region, both in the range of $[0.2, 5.0]$. GC and PDC give performance around the chance level across the whole parameter space, whereas 'PW-LR r skew' and PT exhibit certain resilience towards this variability in the inputs. However, the classifier based on fractional cumulants is the only method whose performance does not fall towards the chance level within the parameter space.

4 Discussion

This work provides an advance to the effective connectivity research in functional Magnetic Resonance Imaging with use of pairwise methods, by utilizing the additional information contained in the BOLD time series with use of fractional moments of the BOLD distribution combined into cumulants. Usage of this additional information (embedded within a classifier) significantly increases the robustness towards plausible sources of variability in fMRI, namely presence of physiologically realistic (scale-free) background noise as well as disproportion in the inputs strength, either due to differences in the amount of neuronal activity locally induced and/or due to effective differences induced by e.g. regional variations in the coil sensitivity profiles. This is where the value of added information coming from fractional cumulants reveals, as, among the methods tested in this work, only the classifier based on the fractional cumulants gives a performance better than chance across the whole parameter space.

Although fractional moments of the distribution, as a mathematical concept, were studied [10, 18], this concept was not applied to natural sciences to date. One reason for this lack of applications might be that the fractional moments become complex numbers for the normalized time series, and that subsequently, the features characterized by these moments cannot be conceptualized as easily as the features characterized by the integer moments (e.g., skewness can be interpreted as a measure of 'asymmetry' of the distribution, and kurtosis can be interpreted as its 'flatness'). In this work, we demonstrate that these features of the distribution provide important additional information about the distribu-

tion. We first perform supervised learning of the classifier on the set of benchmark synthetic datasets, and then validate the classifier on 2-node simulations with biologically realistic confounds. We believe that confounding factors such as a physiological background noise of a magnitude varying between the nodes is an important to overcome for any method for causal inference in fMRI. This is because every network in the brain is embedded in larger networks, therefore the background activity of other interconnected networks can be interpreted as 'noise' [7]. We demonstrate that our approach can increase the robustness of the methods for pairwise inference in fMRI to the main sources of variability in BOLD fMRI.

In this work, we performed the inference on the full BOLD response, without deconvolving the BOLD time series into the neuronal time series. Our previous theoretical research suggests that deconvolution is not necessary in effective connectivity research in fMRI if the used method is not lag-dependent [4]. This is because under the assumption that the signal is in the low frequency range, the hemodynamic response does not affect the signatures of different connectivity patterns present in the data.

Unlike the previous methods for pairwise inference in fMRI [15], the classifier built in this study is informed by the Dynamic Causal Modeling generative model, therefore it incorporates the priors derived from the neurophysiological studies [6]. Deriving benchmark signum maps for the classifier from the multiple instantiations of the DCM generative model allows for marginalizing out all the parameters unimportant for the effective connectivity research: the classification procedure focuses only on classifying a pair of regions into upstream and downstream instead of fitting all the hyperparameters as is done in the classic DCM inference procedure. Therefore, this approach is a *reduction* of the problem of effective connectivity in a large network to a two-node classification problem on one hand, and an *extension* of the feature space from integer to fractional moments on the other hand.

One crucial limitation of our approach as well as previous methods such as pair-wise likelihood ratios is that these techniques only retrieve the *net* connectivity. In case where two brain regions have a bi-directional communication (as is probably true in most of the connections in the brain), it will retrieve the *stronger* out of the two connections. Since in the fist step of the inference, we pick only strong functional links for the further classification, we can interpret ambiguous output of the voting as a bidirectional connection. One interesting direction for the method development would be also classification between excitatory and inhibitory connectivity, which is missing in this study as we focus on the connections of a positive sign only.

In the context of fMRI research, increasing the granularity of moments in order to better characterize the distribution is an especially useful application because the experimental fMRI datasets are short (a few thousands samples at most), therefore the estimation error for the high order integer moments of the distribution becomes high. However, the subsequent cumulants contain information redundant to a certain extent, as they are correlated for any given time series $x(t)$. We chose the granularity which gives smooth patterns of discriminability (Fig. 3), which is $\Delta_k = 0.1$. Choosing the optimal moment resolution is a subject to future research, although we believe that increasing index resolution to substantially less than 0.1 would not be beneficial, yet it would substantially increase the computational cost for the method.

Despite these limitations, our approach presents a significant improvement over previous approaches for pair-wise estimation of functional causal interactions: the robustness against known sources of variability significantly increases due to the simultaneous incorporation of multiple aspects of the associated BOLD distributions. We believe that, as this approach based on fractional moments of a distribution increases resilience of the methods for pairwise connectivity to potential confounds in the experimental data, it can become

a generic method to increase the power of causal discovery studies, both in cognitive neuroimaging and beyond.

References

[1] T. Arichi, G. Fagiolo, M. Varela, A. Melendez-Calderon, A. Allievi, N. Merchant, N. Tusor, S. J. Counsell, E. Burdet, C. F. Beckmann, and A. D. Edwards. Development of BOLD signal hemodynamic responses in the human brain. *NeuroImage*, 63(2):663–73, 2012.

[2] L. A. Baccalá and K. Sameshima. Partial Directed Coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–74, 2001.

[3] C. Bédard, H. Kröger, and A. Destexhe. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Physical Review Letters*, 97:118102, 2006.

[4] N. Z. Bielczyk, A. Llera, Jan K. Buitelaar, J. C. Glennon, and C. F. Beckmann. The impact of haemodynamic variability and signal mixing on the identifiability of effective connectivity structures in BOLD fMRI. *Brain and Behavior*, 2017.

[5] P. Billingsley. *Probability and Measure*. New York: Wiley, 1995.

[6] R. B. Buxton, E. C. Wong, and L. R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. *Magnetic Resonance in Medicine*, 39(6):855–64, 1998.

[7] G. Deco, V. K. Jirsa, and A. R. McIntosh. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56, 2011.

[8] N. Dehghani, C. Bédard, S. S. Cash, E. Halgren, and A. Destexhe. Comparative power spectral analysis of simultaneous elecroencephalographic and magnetoencephalographic recordings in humans suggests non-resistive extracellular media. *Journal of Computational Neuroscience*, 29(3):405–21, 2010.

[9] I. M. Devonshire, N. G. Papadakis, M. Port, J. Berwick, A. J. Kennerley, J. E. Mayhew, and P. G. Overton. Neurovascular coupling is brain region-dependent. *NeuroImage*, 59(3):1997–2006, 2012.

[10] I. M. Dremlin. Fractional moments of distributions. *Journal of Experimental and Theoretical Physics Letters*, 59(9):585–588, 1994.

[11] K. J. Friston. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36, 2011.

[12] K. J. Friston, L. Harrison, and W. Penny. Dynamic Causal Modeling. *NeuroImage*, 19(4):1273–302, 2003.

[13] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrika*, 37(3):424–38, 1969.

[14] B. Y. He. Scale-free brain activity: past, present, and future. *Trends in Cognitive Neurosciences*, 18(9):480–87, 2014.

[15] A. Hyvärinen and S. Smith. Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14(1):111–52, 2013.

[16] A. Hyvärinen, K. Zhang, S. Shimizu, and P. O. Hoyer. Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity. *Journal of Machine Learning Research*, 11:1709–31, 2010.

[17] G. Krüger, A. Kastrup, and G. H. Glover. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, 45(4):595–604, 2001.

[18] M. Matsui and Z. Pawlas. Fractional absolute moments of heavy tailed distributions. *arXiV preprint*, 2014.

[19] R.S. Patel, F. Dubois Bowman, and J.K. Rilling. A Bayesian approach to determining connectivity of the human brain. *Human Brain Mapping*, 27(3):267–76, 2006.

[20] J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R.A. Poldrack, and C. Glymour. Six problems for causal inference from fMRI. *NeuroImage*, 49(2):1545–58, 2010.

[21] A. K. Seth, A. B. Barrett, and L. Barnett. Granger Causality Analysis in Neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–7, 2015.

[22] S. Shimizu, P. O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–30, 2006.

[23] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich. Network modelling methods for fMRI. *NeuroImage*, 54(2):875–91, 2011.