A Comparison of Effective Connectivity Methods Using Different Performance Metrics*

Su Kyoung Kim[1,2], Suraj Kumar Sanga[3] and Elsa Andrea Kirchner[1,2]

Abstract— **Different analysis methods have been developed to determine brain connectivity patterns. To select suitable methods depending on application contexts, it is essential to evaluate different methods using suitable performance metrics. We propose three application-oriented metrics which enable to measure multivariate causality qualitatively. Using the proposed metrics, the most used analysis methods (Directed Transfer Function, Partial Directed Coherence, Granger-Geweke Causality) are compared on synthetic electroencephalographic data with a predefined causality structure. Furthermore, the performances obtained by using all metrics are evaluated. The results allow us to select the most stable analysis method and the optimal metric by estimating the similarity between the performance obtained by using each metric and the graphically displayed predicted network.**

I. INTRODUCTION

Different connectivity methods based on Granger causality [1] have been developed to determine brain connectivity patterns that allow to present the causality between nodes in networks and thus to provide the direction of connections in networks. To analyze such causality in networks with multiple nodes (i.e. effective connectivity), model-based approaches are necessary. The most used model is the multivariate autoregressive (MVAR) model. Several analysis methods based on MVAR model have been developed [2]–[5] and tested on both synthetic and real neurophysiological data [4]–[7]. To select a suitable method for different application contexts, it is desirable to evaluate analysis methods.

For this reason, there are some studies that have compared analysis methods [7]–[11]. In [9], it is unclear how the performance is exactly calculated. In [7], [8], the metric to evaluate performances of different methods is calculated as the ratio of correct connections and all connections in the ground truth network, i.e. the ratio of predicted connections that are present in the ground truth network and all connections in the ground truth network. This metric represents the true positive rate (TPR = TP/(TP+FN)), where true positives (TPs) denote connections in the predicted network which are present in the ground truth network, and false negatives (FNs) denote missing connections in the predicted network compared to the ground truth network. Thus, such simple metric considers missing connections (TNs) alone and ignores additional connections in the predicted network which are not present in the ground truth network (false positives: FPs).

In few studies [10], [11], both types of wrong connections, missing connections (FNs) and additional connections (FPs), are taken into account in performance measures. In the study of [11], the ratio of additional wrong connections (FPs) and correct connections (TPs) is calculated. Such ratio (FP/TP) can serve as metric to compare different methods, but is not suitable to intuitively demonstrate achieved performances of predicted network compared to the ground truth, since the maximum possible value in such metric is not obvious. In study of [10], both the rate of missing connections, i.e. false negative rate (FNR = TN/(FN+TN)) and the rate of additional connections, i.e. false positive rates (FPR = FP/(FP+TN)) are used for performance measures. However, correctly missing connections, i.e. true negatives (TNs) are not relevant for brain connectivity research. Rather, we can achieve a high value of true negative rate (TNR = TN/(TN+FP)) or FPR, since the amount of not existing connections are greater than the amount of existing connections in the ground truth (i.e. TPR and TNR are highly unbalanced). In the same context, a high *accuracy* ((TP+TN)/(TP+TN+FP+FN)) [12] is achieved, considering that TNs are included for calculation. Even using the balanced accuracy (0.5*(TPR+TNR)) the achieved performance is still very high.

In this paper, we propose two performance metrics considering both types of wrong connections, i.e. missing and additional connections (FNs, FP), but not correctly missing connections (TNs). For comparison, the analysis methods are also evaluated by using the existing metrics including TN (e.g. *Accuracy*, *balanced Accuracy*) and excluding TN (e.g. *F-measure*) [12]. Furthermore, we propose one metric that measures correctly predicted indirect connections. Using different metrics, we compare the most used effective connectivity methods. First, all metrics are evaluated to find out the most stable analysis method. Thus, the best method can be found by comparing performances in all proposed metrics. Second, the similarity between the performance of each metric and the graphically displayed predicted network is estimated to find out the optimal metric.

II. METHODS

A. Multivariate autoregressive model (MVAR)

Effective connectivity methods based on the MVAR model allow to calculate multivariate causality. A multi-sensor

*This work was supported by the German Ministry of Economics and Technology (grant no. 50 RA 1011 and grant no. 50 RA 1012).

[1]Su Kyoung Kim and Elsa Andrea Kircher are with the Research Group Robotics, Faculty of Mathematics and Computer Science, University of Bremen, Germany. sukim, ekir@informatik.uni-bremen.de

[2]Su Kyoung Kim and Elsa Andrea Kirchner are with the Robotics Innovation Center (RIC), German Research Center for Artificial Intelligence (DFKI) GmbH, Robert–Hooke–Str. 5, 28359, Bremen, Germany.

[2]Suraj Kumar Sanga is with the Department of Electrical Engineering, Bremen University of Applied Science, Bremen, Germany. ssanga@stud.hs-bremen.de

TABLE I

SIMULATION OF 10-CHANNEL NETWORK WITH TWO DIFFERENT NETWORK COMPLEXITIES BASED ON THE MVAR MODEL

Equation for 10-channel MVAR model with **5** indirect connections	Equation for 10-channel MVAR model with **10** indirect connections
$X_1(t) = 0.54X_1(t-1) + E_1(t)$	$X_1(t) = 0.54X_1(t-1) + 0.5X_2(t-3) + E_1(t)$
$X_2(t) = 0.6X_2(t-2) + E_2(t)$	$X_2(t) = 0.6X_2(t-2) + E_2(t)$
$X_3(t) = 0.5X_2(t-3) + 0.7X_3(t-3) + 0.7X_4(t-3) + E_3(t)$	$X_3(t) = 0.5X_2(t-1) + 0.7X_3(t-3) + 0.7X_4(t-3) + E_3(t)$
$X_4(t) = 0.55X_4(t-1) + E_4(t)$	$X_4(t) = 0.55X_4(t-1) + E_4(t)$
$X_5(t) = 0.55X_5(t-1) + 0.75X_7(t-1) + E_5(t)$	$X_5(t) = 0.35X_3(t-1) + 0.55X_5(t-1) + 0.75X_7(t-1) + E_5(t)$
$X_6(t) = 0.55X_6(t-1) + 0.75X_8(t-1) + 0.35X_3(t-1) + E_6(t)$	$X_6(t) = 0.55X_6(t-1) + 0.75X_8(t-1) + 0.35X_3(t-1) + E_6(t)$
$X_7(t) = 0.55X_7(t-1) + 0.75X_9(t-1) + 0.85X_{10}(t-1) + E_7(t)$	$X_7(t) = 0.55X_7(t-1) + 0.75X_9(t-1) + 0.85X_{10}(t-1) + E_7(t)$
$X_8(t) = 0.55X_8(t-1) + 0.45X_1(t-1) + E_8(t)$	$X_8(t) = 0.55X_8(t-1) + 0.65X_{10}(t-1) + 0.45X_1(t-1) + E_8(t)$
$X_9(t) = 0.55X_9(t-1) + E_9(t)$	$X_9(t) = 0.55X_9(t-1) + E_9(t)$
$X_{10}(t) = 0.65X_{10}(t-1) + E_{10}(t)$	$X_{10}(t) = 0.65X_{10}(t-1) + E_{10}(t)$

signal X_t can be defined as $X_t = [X_1(t), \ldots, X_m(t)]^T$, where t denotes the time index and m denotes the number of sensors. A multivariate autoregressive process can be represented as:

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + E_t \tag{1}$$

where A_j are the MVAR model coefficients (i.e., $m \times m$ coefficient matrix), j is the time delay, p is the MVAR model order that determines the number of past time points, which is selected by the Akaike Information Criteria (AIC) [13], and E_t is a vector of prediction error.

By rearranging (1), we obtain (2), where A_0 = -I (identity matrix):

$$\sum_{j=0}^{p} A_j X_{t-j} = E_t \tag{2}$$

We obtain (3), i.e. the Yule-Walker equation [14] by multiplying $X_{t-r}^T (r = 1, ..., p)$ to both sides of (2),

$$\sum_{j=1}^{p} A_j R(j-r) = 0 \tag{3}$$

where $R(n)$ is the covariance matrix of X_t with lag n. The Levinson-Wiggins algorithm [14] is used to obtain the coefficients of the MVAR model.

B. Effective connectivity measures

1) Directed Transfer Function (DTF): The DTF [2], [15] is a measure of direction of information transfer between nodes. To obtain the causal information flow in the frequency domain, the DTF uses the transfer function $H(f)$ calculated as the ratio of input $E(f)$ and output $X(f)$.

$$E(f) = A(f)X(f) \Longleftrightarrow X(f) = A(f)^{-1}E(f) = H(f)E(f) \tag{4}$$

where

$$A(f) = -\sum_{j=1}^{p} A_j e^{-i2\pi f j} \tag{5}$$

The DTF, $\gamma_{ij}(f)$ represents causal influence of node j on node i at frequency f by calculating the ratio of inflow of node j to node i and all inflows to node i.

$$\gamma_{ij}(f) = \frac{|\, H_{ij}(f) \,|^2}{\sum_{m=1}^{k} |\, H_{im}(f) \,|^2} \tag{6}$$

where H_{ij} is the inflow from node j to node i and H_{im} is all the inflows to node i.

Accordingly we obtain the power spectral matrix $S(f)$:

$$S(f) = X(f)X(f)^* = H(f)V(f)H^*(f) \tag{8}$$

where $V(f)$ is the covariance matrix of $E(f)$ and * denotes the transposition and complex conjugation.

2) Partial Directed Coherence (PDC): The PDC [3], [16] represents the relative coupling strength of the interaction of a given signal source j to i. The PDC, $\pi_{ij}(f)$ is calculated as the ratio of outflow from the source node j to node i and all the outflows from the source node j.

$$\pi_{ij}(f) = \frac{A_{ij}(f)}{\sqrt{\sum_{x=1}^{N} A_{xj}(f)^H A_{xj}(f)}} \tag{7}$$

where A_{ij} is the outflow from the source node j to the node i and A_{xj} is all the outflows from the source node j.

3) Granger-Geweke Causality (GGC): The Granger causality (GC) [1] measures the causality between nodes in time domain. The GC can also be measured in the frequency domain, i.e. Granger-Geweke causality [4].

$$G_{x \to y}(f) = -\ln(1 - \frac{\sum_{xx} - \sum_{xy}^2 / \sum_{yy} \,|\, H_{yx}(f) \,|^2)}{S_{yy}(f)} \tag{8}$$

where x and y are time series of node x and y and $\sum$ denotes covariance matrix.

III. EVALUATION

A. Data generation

A 10-channel network with two different network complexities (i.e. two different causal structures) was generated based on MVAR modeling (see Table I). The error term $E(t)$ of the MVAR model consisted of random numbers in the range of [-4, 4]. The obtained 10-channel networks with different network complexities served as ground truth network.

B. Performance metrics

To measure the performance of predicted networks, we used two performance metrics containing both missing and additional connections (FN, FP). As mentioned earlier, TNs are excluded in the proposed metrics, since TNs were not

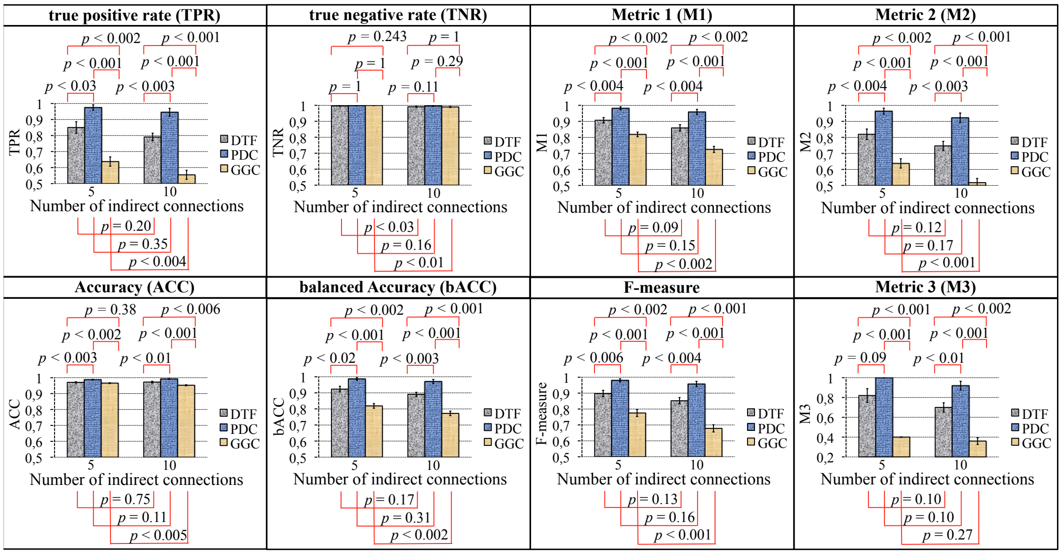


Fig. 1. The performances of different effective connectivity methods based on different types of performance metrics (TPR, TNR, M1, M2, M3, ACC, bACC, $F_{measure}$) are compared for each network complexity (5, 10 indirect connections). Mean and standard error are depicted.

relevant for our case and the TPR and TNR were highly unbalanced. Thus, in the proposed metric, missing and additional connections (FNs and FPs) were included, but not the contribution of TNs (e.g. TNR or FPR). The first metric (M1) is the arithmetic mean of recall (TP/(TP+FN)) and precision (TP/(TP+FP)). The second one (M2) is the ratio of correct connections (TPs) and the sum of correct connections (TPs) and all types of wrong connections (FNs, FPs). The third metric (M3) is the ratio of amount of indirect connections in the predicted network (I_y) and the amount of all connections in the ground truth network (I_x).

$$M1 = 0.5 * ((TP/(TP+FN)) + (TP/(TP+FP))) \quad (10)$$

$$M2 = TP/(TP+FN+FP) \quad (11)$$

$$M3 = I_y/I_x \quad (12)$$

Compared to the proposed metrics, the existing metrics *accuracy (Acc)*, *balanced accuracy (bACC)*, and the harmonic mean between precision and recall, i.e. *F-measure* [12] were used. These metrics are represented as:

$$ACC = (TP+TN)/(TP+TN+FN+FP) \quad (14)$$

$$bACC = 0.5 * (TPR+TNR) \quad (14)$$

$$F_{measure} = 2 * (precision * recall)/(precision + recall) \quad (15)$$

C. Evaluation Procedure

To investigate the influence of randomly generated noise ($E(t)$) on performance, the network was predicted 30 times for each metric and each network complexity and thus we obtained 30 performances for each metric and each network complexity. This allowed us to obtain the stability of performance for each predicted network. In the end, we obtained 30 performance values for each performance metric (M1, M2, M3, TPR, TNR, *accuracy*, *balanced accuracy*, *F-measure*), each set of network (5, 10 indirect connections), and each effective connectivity method (DTF, PDC, GGC).

D. Statistical analysis

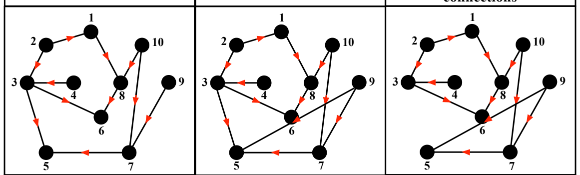
To compare the different effective connectivity methods for each network complexity, the obtained performances were analyzed by repeated measures ANOVA with the complexity of network (indirect connections: 5, 10) and the type of effective connectivity method (DTF, PDC, GGC) as within-subjects factors. For comparison of the different performance metrics, the type of metric (M1, M2, M3, TPR, TNR, *accuracy*, *balanced accuracy*, *F-measure*) was used as an additional within-subjects factor. If necessary, the Greenhause-Geisser was applied and for multiple comparisons, Boferroni correction was applied.

IV. RESULTS

Fig. 1 shows the performance of different effective connectivity methods for each network complexity (indirect connections: 5, 10) and each type of performance metric (M1, M2, M3, TPR, TNR, *accuracy*, *balanced accuracy*, *F-measure*).

For both types of network complexity, the highest performance was achieved with the PDC followed by the DTF and GGC. Such highest performance of PDC was shown for all types of performance metrics except for TNR. The GGC was the worst method for all types of performance metrics,

Ground truth	1 additional connection	1 additional and 1 missing connections



performance based on graph	91 %	82 %
true positive rate (TPR)	100 %	91 %
true negative rate (TNR)	99 %	99 %
metric I (M1)	96 %	91 %
metic II (M2)	92 %	83 %
metric III (M3)	100 %	73 %
Accuracy (ACC)	99 %	99 %
balanced accuracy (bACC)	99 %	95 %
F-measure (Fm)	96 %	91 %

Fig. 2. The performances obtained by using the proposed and the existing metrics are compared. The performance based on the graph serves as a baseline. The ground truth of the 10-channel network with 10 indirect connections (see Table I) has 11 information flows (i.e. 1→8, 2→1, 2→3, 3→5, 3→6, 4→3, 7→5, 8→6, 9→7, 10→8, 10→7). The performance based on the graph was calculated as the difference of all connections in the ground truth network (TP+FN) and all types of wrong connections (FP+FN) divided by all connections in the ground truth network (TP+FN): (TP+FN)-(FP+FN)/(TP+FN).

especially in case that the performance was measured based on indirect connections (see M3 in Fig. 1).

Furthermore, the performance of GGC was affected by the network complexity (i.e. higher performance for the less complex network compared to the more complex network). Such pattern was not shown for DTF and PDC (statistical values, see Fig. 1).

Especially, the PDC provided a stable performance, in which the difference in performance among all types of performance metrics was less compared to the DTF and GGC. The PDC showed a stable performance for all types of performance metrics (above 90%), whereas the performance of DTF and GGC varied among the type of performance metric (Details, see Fig. 1).

Fig. 2 illustrates the comparison of different performance metrics. Among the different types of metrics, M2 was the one that was closest to the graphically displayed predicted network. That means, the performance using M2 was similar to the performance based on the graph. In contrast, a higher performance was obtained using M1 and *F-measure* compared to the performance based on the graph. Such higher performances compared to the performance based on the graph was more characteristic for the metric including TNs, e.g. *ACC* and *bACC*. M3 showed a high value in case of additional connections.

V. CONCLUSION

In this paper, different performance metrics for measuring performances of different methods of effective connectivity are proposed and evaluated. The proposed metrics take into account all types of wrong connections and enable to demonstrate performances intuitively. The evaluation of the proposed metrics allows to measure the stability of the achieved performance and contributes to the selection of the best methods based on performance stabilities. Furthermore, the results of metric evaluation allows to find out the qualitatively better metric by estimating the similarity between the performance of each metric and the graphically displayed predicted network. In real electroencephalographic (EEG) data with larger amount of channels (e.g. 128 channels), the distance between channels may play a critical role in interpreting performances, since additional connections between closely positioned channels have qualitatively different meanings compared to additional connections between channels with large spatial distances (e.g. wrong connections between frontal and occipital channels). The issue of metrics considering the connection distance will be addressed in future work.

REFERENCES

[1] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug 1969.

[2] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological Cybernetics*, vol. 65, no. 3, pp. 203–10, 1991.

[3] K. Sameshima and L. A. Baccala, "Partial directed coherence: a new concept in neural structure determination," *Journal of Neuroscience Methods*, vol. 94, no. 1, pp. 93–103, 1999.

[4] F.-H. Lin, K. Hara, V. Solo, M. Vangel, J. W. Belliveau, S. M. Stufflebeam, and M. Hämäläinen, "Dynamic granger-geweke causality modeling with application to interictal spike propagation," *Human Brain Mapping*, vol. 30, no. 6, pp. 1877–1886, 2009.

[5] Y. Chen, S. L. Bressler, and M. Ding, "Frequency decomposition of conditional granger causality and application to multivariate neural field potential data," *Journal of Neuroscience Methods*, vol. 150, no. 2, pp. 228–237, 2006.

[6] R. Kus, M. J. Kaminski, and K. J. Blinowska, "Determination of EEG activity propagation: Pair-wise versus multichannel estimate," *IEEE Trans Biomed Engineering*, vol. 51, no. 9, pp. 1501–1510, 2004.

[7] L. Astolfi, F. Cincotti, D. Mattia, S. Salinari, C. Babiloni, A. Basilisco, P. M. Rossini, L. Ding, Y. Ni, B. He, M. G. Marciani, and F. Babiloni, "Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution eeg," *Magnetic Resonance Imaging*, vol. 22, pp. 1457–1470, 2004.

[8] L. Astolfi, F. Cincotti, D. Mattia, M. Marciani, L. Baccala, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. C. Edgar, G. Miller, B. He, and F. Babiloni, "Comparison of different cortical connectivity estimators for high-resolution eeg recordings," *Human Brain Mapping*, vol. 28, pp. 143–157, 2007.

[9] M.-H. Wu, R. E. Frye, and G. Zouridakis, "A comparison of multivariate causality based measures of effective connectivity," *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1132–1141, 2011.

[10] E. Florin, J. Gross, J. Pfeifer, G. Fink, and L. Timmermann, "Reliability of multivariate causality measures for neural data," *Journal of Neuroscience Methods*, vol. 198, pp. 344–358, 2011.

[11] A. Fasoula, Y. Attal, and S. D, "Comparative performance evaluation of data-driven causality measures applied to brain networks," *Journal of Neuroscience Methods*, vol. 215, pp. 170–189, 2013.

[12] I. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, 2005.

[13] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, 1974.

[14] S. M. Kay, *Modern Spectral Estimation*. Prentice Hall, 1999.

[15] M. J. Kaminski, M. Z. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, no. 2, pp. 145–57, 2001.

[16] K. Sameshima and L. A. Baccala, "Using partial directed coherence to describe neuronal ensemble interactions," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.