IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 5, 2020, accepted January 5, 2021, date of publication January 18, 2021, date of current version January 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051985

Description Length Guided Unified Granger Causality Analysis

ZHENGHUI HU, FEI LI, XUEWEI WANG, AND QIANG LIN
College of Science, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Zhenghui Hu (zhenghui@zjut.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701400, and in part by the Public Projects of Science Technology Department of Zhejiang Province under Grant LGF20H180015.

ABSTRACT In this article, we propose a description length guided unified Granger causality analysis (uGCA) framework for sequential medical imaging. While existing efforts of GCA focused on causal relation design and statistical methods for their improvement, our strategy adopts the minimum description length (MDL) principle in the GCA procedure where the MDL principle offers a unified model selection criteria for deciding the optimal model in the sense of description length. Under this framework, we present different description length forms of linear Granger representations under several coding schemes that all achieve the lower bounds on redundancy, thus producing valid MDL model selection criteria. The efforts are validated using a 5-node network synthetic experiment, illustrating its potential advantage over conventional two-stage approach. The subtle distinction between the performance of different uGCA forms is investigated as well. More importantly, the proposed approach gives a more similar network topology than conventional approach in a challenging fMRI dataset, in which neural correlates of mental calculation elicited by visual and auditory stimulation (respectively) in the same task paradigm, allowing one to evaluate the performance of different GCA methods.

INDEX TERMS Description length, Granger causality analysis (GCA), minimum description length (MDL), model selection.

I. INTRODUCTION

Nervous system of the brain is characterized by complex temporal and spatial changes, which has drawn increasing attention in describing its dynamic space-time network [1]–[6]. Until now, from macro to micro levels, a growing body of research demonstrates that the brain works together through distributed neural subsystems (functional integration) [7]–[12], rather than specialized independent systems (functional specialization) [13]–[15]. To investigate how this distributed dynamic connection network is integrated, causal connectivity analysis, also known as effective connectivity, which refers to the event correlation of a nervous system effected by another one, should be essential [16]–[19].

Granger causality analysis (GCA) provides a statistical hypothesis test for investigating causal connectivities that can provide information about the dynamics and directionality of associations [20]. Certainly, Granger-causality describes the temporal relation between variables, which is played a predicting role rather than explained the causality in philosophical. Specifically, GCA indicates time series X Granger-cause Y if importing X can provide more statistically significant information about the prediction of Y than its restricted model [21].

In general, GCA is performed in the context of linear autoregressive (AR) models for stationary time series. The candidate models with different time-lags, imply a historical dependence on variables. Then, through model selection technique, the optimal model basically is selected by balancing fitting error term and penalized term. Original Granger causality does not consider potential confounding effects, nor does it capture instantaneous and nonlinear causalities. Therefore, some extensions of GCA have been developed to capture nonlinear causal relations [22]–[24]. To differentiate the causal effects of positive from negative, an asymmetric causality test was proposed [25]. Moreover, causal investigation paradigm of GCA has been also generalized in other function spaces, e.g. Fourier spaces [26]–[28], kernel Hilbert spaces [29]–[31]. In addition, considering the model with asymptotic noise distribution data, several forms of extensions, which are especially suitable in task-related fMRI researches, have been developed [32], [33]. But causal

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

investigation among regional networks still can not express the connectivities between regions of interest completely and accurately. Thus, numerous efforts have been made to scale from a small network with several nodes to a large complex network [34]–[37]. These methods are broadly applied in exploring the internal connection of the brain for which causal relationships between neuron populations are usually nonlinear and have complex statistics in view of various sources of uncertainty.

Despite these improvements, the initially mathematical core for investigating Granger causality remains in place. That is, GCA determines the time-lag of candidate model through Akaike information criterion (AIC) or Bayesian information criterion (BIC) and then obtains causal influence through F-test, which actually is a two-stage scheme [22], [27], [38], [39]. In this sense, the conventional GCA method relies on the subjective selection of confidence level, resulting in a lack of uniformity of research results, which will bring in some performance issues. Another problem brought about by F-test needs to be compared with each other through an intermediate model, which will increase algorithm complexity, especially in large networks. And selection results by pairwise F-statistics sometimes depend on initial selected model and search path heavily [40]. It is mentioning that selecting and using the F statistical value have become very careful in current scientific research, and its statistical significance has also caused extensive discussion [41]–[44]. Generally, most GCA methods still define causal connectivities by the framework of original GCA, which does not solve the inconsistency of mathematical theories, the subjective selection of confidence level, and the algorithm complexity caused by nested model. In fact, in a purely mathematical sense, these two stages are a generalized model selection paradigm, the application of two different mathematical theories will cause inherent correlation to be discontinuous in the quantitative mapping process, namely, bringing in singularity.

Based on these considerations, to make investigating causal connectivity more consistent and continuous, we argue that the generalized model selection paradigm in GCA should follow the same benchmark under one mathematical theory. In our previous study, against the conventional two-stage scheme, we have proposed a unified GCA (uGCA) method with the minimum description length (MDL) principle that model selection follows a single mathematical theory during the GCA process. Our uGCA methods can unify the dynamics of brain regions into the same framework continuously, which guarantees the correlation is not distorted or truncated.

Advocating principle of *parsimony*, with help of the algorithmic or description complexity theory of Kolmogorov [45], [46], Wallace and Boulton [47], MDL was formulated as a broad principle governing statistical modeling in general [48]. Then embracing Shannon's information theory, MDL was endowed with a rich information-theoretic interpretation. Until now, it has been developed in several forms. For original MDL, it describes the candidate models by two-part description length (or code length), which is two-part MDL. One part is the fitting error term, the other is the model complexity term. Following, distilling the stochastic complexity, MDL is developed a mixture form which adopted some *priors* to describe the nuisance parameters, named mixture form MDL. Combined with Fisher information, the normalized maximum likelihood (NML) form MDL has been developed. These are three main forms of MDL, of course, it also has other forms of description standpoint [49], [50]. We had illustrated the benefits of introducing a unified model selection approach in simulated and real fMRI experiments, in term of two-part MDL guided uGCA [40]. These uGCA methods with different MDL forms have completely different coding schemes in terms of describing model complexity. This uGCA paradigm hopes to find a most suitable model to describe data in a description length guided framework, while, theoretically, the true model behind the data does not exist. Therefore, uGCA methods of different forms approach to the true model from different aspects to obtain their own optimal model in describing data.

The rest of article is as follows. In Section II, different forms of MDL principle have been stated in turn, the generalized formulas of several forms also have been derived with Bernoulli distribution in Markov model class. Immediately, three forms of description length guided causal investigation for ordinary linear model has been carried out. In Section III, we illustrate the advantages of several uGCA forms over the conventional two-stage GCA in two 5-node network synthetic experiments. More importantly, in a task-related fMRI dataset, uGCA methods obtained more consistent results for causality investigation of mental arithmetic networks under visual and auditory stimulus, respectively. Section IV and V demonstrate comparisons between conventional two-stage GCA and our proposal from a mathematical modeling standpoint, and discusses its following potential development.

II. MINIMUM DESCRIPTION LENGTH PRINCIPLE

The MDL provides a generic solution for model selection issue as an information criterion [51], and it regards the probability distribution as a descriptive standpoint to choose the model that gives the shortest description of data. As a broad principle, MDL has some connections with AIC and BIC, sometimes it behaves similarly to them [49], [52], [53]. But it actually represents a completely different approach for model selection relative to conventional statistical approaches. Compared with AIC/BIC, MDL fixes attention on unifying model complexity term and fitting-error term into a description length guided framework, and it does not require any assumptions about the data generation mechanism. The purpose of model selection in MDL is not to estimate an assumed but unknown distribution, but to find a more appropriate model to describe data [54], [55].

A. DIFFERENT FORMS OF MDL

In the following, we formally introduce several coding schemes that provide valid description lengths of a data string based on classes of probability models [49], [50]. To derive

formulas and explain models more intuitively, we first consider a simple parametric model class $\mathcal{M}$ comprising a family of distributions with the parameter $\theta \in R^k$. Let $x^n = (x_1, x_2, \ldots, x_n)$ denote a data string, and the model class is represented as

$$\mathcal{M} = \{f(x|\theta^n) : \theta \in \Theta \subset R^k\}.$$

We have uncovered two-part MDL in detail in previous studies [40], but still describing it briefly here in order to make a more clear comparison between several different forms of the coding scheme. Then the three descriptive schemes are introduced successively, and we derive a more general framework for each form. Built on these derivations, next section provides a rather extensive treatment of MDL for causal investigation in ordinary linear regression.

1) TWO-PART CODING SCHEME

At original MDL, it usually divides modeling for the data set into two parts, first part is to choose a subset of class $\mathcal{M}$, and then encode x^n using this distribution. From the family $\mathcal{M}$ to a candidate distribution, which is selected by an estimator $\hat{\theta}_n$, then the description length associated with a prefix code is constructed from $f_{\hat{\theta}_n}$. Ultimately, similar to AIC/BIC, the description length within two-part MDL adopts penalized fitting-error and model complexity to encode data fitting $f(x^n|\hat{\theta}_n)$ and estimated parameter value $\hat{\theta}_n$. Consequently, the most common implementation of MDL, two-part code version [40], [50], is carried out. The description length for coding x^n is then

$$\begin{aligned} L_{2p}(x^n, \theta) &= L_1((x^n|\hat{\theta}_n) + L_2(\hat{\theta}_n) \\ &= -\log f(x^n|\hat{\theta}_n) + \sum_{i=1}^{k} \log \frac{\hat{\theta}_i}{\delta} + \log(n+1). \end{aligned} \quad (1)$$

where the distribution $\hat{\theta}_n$ on parameter space was truncated to same precision $\delta = 1/\sqrt{n}$.

2) MIXTURE FORM CODING SCHEME

Different from a crude coding scheme to describe model complexity in the two-part MDL, the mixture form of MDL establishes a description of data string x^n on the basis of a distribution, which is obtained by taking a mixture of members in a family of probability density functions ω on parameters,

$$m(x^n) = \int f_\theta(x^n)\omega(\theta)d\theta \quad (2)$$

The mixture $m(x^n)$ involves integrating over model classes, to get around such difficulties, its first-order approximation to this form coincides with the two-part MDL [56]. An alternative approximation yields another form of description length known as Stochastic Information Complexity (SIC). Associating the stochastic complexity concept and creating a distribution for the data based on model class $\mathcal{M}$, an analytical approximation to the mixture MDL is obtained by Laplace's expansion when ω is smooth [49], [56]–[59]. The mixture finally arrives at, $I(\theta)$ denotes the Fisher information of θ,

$$L_g = -\log f(x^n|\hat{\theta}) + \frac{k}{2}\ln\frac{n}{2\pi} + \log\frac{|I(\theta)|^{1/2}}{\omega(\theta)} + o(1). \quad (3)$$

Indeed, the mixture form of MDL shares many formal elements with Bayesian model selection because their underlying analytical tools are the same. However, the philosophies behind each approach are much different.

3) NML FORM CODING SCHEME

In earlier two-part codes, it still remains the inherent redundancy. The NML form of MDL, taking into account Fisher information, was developed based on the coding scheme of Barron *et al.* [57] and Shtarkov [60]. In general, NML form restricted the early second part description of two-part MDL into a data region identified by parameter estimation [49]. This scheme for MDL model selection was formally introduced by Rissanen in [61], and discussed its association with minimax theory. In this case, as long as $\hat{\theta}_n$ exists for all x^n, we have

$$P^{(n)}_{nml}(x^n) = \frac{P_{\hat{\theta}_n}(x^n)}{\sum P_{\hat{\theta}_n}(x^n)}. \quad (4)$$

The sequence of distributions $P^1_{nml}, P^2_{nml}, \ldots$ constitutes minimax optimal universal model relative to $\mathcal{M}$, it tries to assign to each x^n a probability according to ML distribution for x^n [50]. And the studies were carried further by [57], [61], for sequences x^n such that $\hat{\theta}(x^n) \in \Gamma$,

$$\begin{aligned} L_n = -\log f(x^n|\hat{\theta}(x^n)) + \frac{k}{2}\ln\frac{n}{2\pi} \\ + \ln\int_\Gamma \sqrt{|I(\theta)|}d\theta + o(1) \end{aligned} \quad (5)$$

B. DESCRIPTION LENGTH GUIDED CAUSAL INVESTIGATION

Associating with above extensive treatment, the causal investigation within uGCA methods can be put into effect. Firstly, two time-series X_N and Y_N are given, to describe X_N, it has

$$\begin{cases} X_t = \sum_{j-1}^{n} a_{1i}X_{t-j} + \epsilon_{1t} \\ X_t = \sum_{j-1}^{n} a_{2i}X_{t-j} + \sum_{j-1}^{n} b_{2i}Y_{t-j} + \epsilon_{1t} \end{cases} \quad (6)$$

where ϵ_t denotes the fitting residual, which has Gaussian distribution with mean 0 and unknown variance σ^2. Thus, the residual terms ϵ_t can be a standpoint to describe the model within MDL. Distilling the sense of Granger causality, the influence from Y to X based on description length guided framework is defined by

$$F_{Y\to X} = L_X - L_{X+Y}. \quad (7)$$

where L_X denotes the minimum description length of restricted model for X_N, and L_{X+Y} denotes the minimum description length of unrestricted model for X_N after adding Y_N. If $F_{Y\to X} > 0$, it means causal influence from Y to X existed, or else there is no causal influence from

Y to X. As represented above, uGCA can unify conventional two-stage scheme into the description length guided framework to obtain causal relationship, which can avoid the inconsistency of different mathematical theories and the issue of subjective selection of confidence level.

Considering conditional Granger causality, the influence from Y to X can be investigated while controlling the effect from another node Z to X. This joint representation is

$$\begin{cases} X_t = \sum_{j-1}^{n} a_{3i} X_{t-j} + \sum_{j-1}^{n} b_{3i} Z_{t-j} + \epsilon_{3t} \\ X_t = \sum_{j-1}^{n} a_{4i} X_{t-j} + \sum_{j-1}^{n} b_{4i} Y_{t-j} + \sum_{j-1}^{n} c_{4i} Z_{t-j} + \epsilon_{4t}. \end{cases} \tag{8}$$

Thus, if $F_{Y\to X} > 0$, in the conditional concept of proposed uGCA methods, the causal influence from Y to X conditioned Z is given by the description length in (8),

$$F_{Y\to X|Z} = L_{X+Z} - L_{X+Y+Z}. \tag{9}$$

where L_{X+Z} denotes description length of the unrestricted model for X_N after adding Z_N. And L_{X+Y+Z} is the length of unrestricted model for X_N after adding X_N and Z_N. Same as above, causal influence from Z to X conditioned Y is

$$F_{Z\to X|Y} = L_{X+Y} - L_{X+Y+Z}. \tag{10}$$

Intuitively, in this unified framework, all candidate models can be described as description length and then compared freely. In other words, unlike the conventional scheme, which can only perform pairwise comparison by nested models, uGCA can freely choose the number of comparison models, which can release the algorithm complexity. Which is, if both $F_{Y\to X} > 0$ and $F_{Z\to X} > 0$ exist,

$$F_{Y,Z\to X} = min(L_{X+Y}, L_{X+Z}) - L_{X+Y+Z}. \tag{11}$$

Clearly, if $F_{Y,Z\to X} > 0$ exist, it means that both Y and Z have direct influence on X. But that will be dealt with in two cases when $F_{Y,Z\to X} < 0$. One is $F_{Y,Z\to X} = (L_{X+Y} - L_{X+Y+Z}) < 0$ existed, which means Y impacts on X directly and Z has a indirect causal influence X. The other case is $F_{Y,Z\to X} = (L_{X+Z} - L_{X+Y+Z}) < 0$, it indicates only Z has a direct causal influence X. In the unified description length framework, uGCA methods can map each candidate model into a unified mathematical space to have a straightforward comparison. Consequently, the uGCA methods are more concise and rigor, that is more in line with Occam's razor, or the principle of *parsimony*.

C. DIFFERENT FORMS OF DESCRIPTION LENGTH GUIDED uGCA

The following is model selection in the causal investigation of which different forms of uGCA guided for the linear AR model [48], [61]. Data set $x^n = \{x_1, \ldots, x_n\}$ is given,

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_k x_{t-k} + \epsilon_t. \tag{12}$$

where $t = 1, \cdots, m$, and m is more than k to keep the solution determined, n denotes data length. In order to describe x_t, turn to Gaussian distribution for ϵ_t, it arrives at

$$f(x^n|x_t, \beta, \tau) = \frac{1}{(2\pi\tau)^{m/2}} e^{-(1/2\tau)\sum_t (x_t - \sum_k \beta_k x_{t-k})} \tag{13}$$

1) uGCA-TP—CRUDE TWO-PART CODING SCHEME

Clearly, for descriptive model of x^n, its parameter vector consists of data

$$\theta = (k, \xi) \quad \text{and} \quad \xi = (\tau, \beta_1, \cdots, \beta_k),$$

where $\xi \in R^{k+1}$, $\tau = \xi_0$ is the variance-parameter of zero-mean Gaussian distribution model for ϵ_t. Let *RSS* denote the residual sum of squares corresponding to estimation in the model. Thus total description length is given as

$$L_{uGCA-TP} = m \ln\sqrt{2\pi\tau} + \frac{RSS}{2\tau} + \sum_{i=0}^{k} \ln\frac{|\xi_i|}{\delta} + \ln(k+1). \tag{14}$$

where δ is the precision, and it's optimal to choose $1/\sqrt{N}$ [49], [58], [62]. Specially, $\frac{|\xi_i|}{\delta} < 1$ should be ignored.

2) uGCA-MIX—*g*-prior FOR PARAMETER SPACE

To obtain a closed-form expression for mixture form, $\omega(\beta, \tau)$ is represented as a member of the natural conjugate family of priors for normal linear regression model, namely the *normal inverse-gamma* distributions [58],

$$\omega(\beta, \tau) \propto \tau^{\frac{-d+k+2}{2}} e^{\frac{-(\beta-b)' c\Sigma(\beta-b)+a}{2\tau}}. \tag{15}$$

where $\Sigma = X_k' X_k = mS$ ($X_k' = \{x_{i,t-k}\}$) is a $k \times m$ matrix defined by the values of regressor variables [57]. Rissanen [58] provided a special solution that $a = d = 0$, $b = (0, .., 0)$, and Zellner [63] christened a specification the g-prior. The value of Σ, provided a closed-form expression for $\hat{c}$ in [49], namely $1/\hat{c} = max(F - 1, 0)$ where $F = (m-k)(X_t' X_t - RSS)/(k \cdot RSS)$. Thus, R^2 is the usual squared multiple correlation coefficient, the mixture is given

$$L_{uGCA-MIX} = \begin{cases} \frac{m}{2} \ln\frac{RSS}{n-k} + \frac{k}{2} \ln F + \ln m, & \text{if } R^2 \leqslant \frac{k}{n} \\ \frac{m}{2} \log\frac{X_t' X_t}{m} + \frac{1}{2} \log m, & \text{otherwise.} \end{cases} \tag{16}$$

Finally, a simple approximation to this mixture form is applied to derive the SIC [49],

$$SIC = \frac{m-k-2}{2} \log RSS + \frac{k}{2} \log m + \frac{1}{2} \log det[\Sigma], \tag{17}$$

where the additive constant independent of model choice has been omitted. In this context, mixture form adapts to behave like Bayesian model selection. However, for a small n, mixture criterion will not be as sharp as the one provided by the NML form.

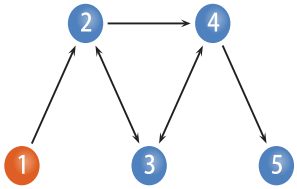


FIGURE 1. The relationships of simulation data sets in the 5-node networks in model A.

3) uGCA-NML—MINIMAX SOLUTION FOR THE INHERENT REDUNDANCY

Combining Fisher information to remove the inherent redundancy in earlier two-part codes, a sharper description length as stochastic complexity and the associated universal process are derived for a class of parametric processes [61]. And this description form is motivated by the maximum-likelihood estimate (MLE) which requires to satisfy the Central Limit Theorem (CLT) [57], [64]. Then the nonintegrability of MLE code is the key issue to be solved. For this, the Fisher information is needed, which is given by

$$|I(\beta,\tau)| = |S|/(2\tau^{k+2}),$$

and the integral of its square root dealt by [57], [61], [64] is

$$\int_{\beta' S\beta \leqslant R}\int_{\tau_0}^{\infty} |I(\beta,\tau)|^{1/2} d\tau d\beta = (2|S|)^{1/2}(\frac{R}{\tau_0})^{k/2}\frac{V_k}{k}. \quad (18)$$

where $V_k R^{\frac{k}{2}} = |S|^{-\frac{1}{2}} 2(\pi R)^{\frac{k}{2}}/k\Gamma(\frac{k}{2})$ denotes the volume of a k-dimensional ball $B = \{\beta' S\beta \leqslant R\}$. Lower bound τ_0 is determined by the precision which the data are written, then $\hat{\tau}_0 = RSS/m$ and $\hat{R} = (\hat{\beta}' X'_{t-k} X_{t-k}\hat{\beta})/m$ obtained by MLE. Thus description length associated with (5) arrives at

$$L_{uGCA-NML} = m\ln\sqrt{2\pi\tau} + \frac{RSS}{2\tau} + \frac{k}{2}\ln\frac{m}{2} - \log\Gamma(\frac{k}{2}) + \frac{k}{2}\log\frac{\hat{R}}{\tau_0} - 2\log k \quad (19)$$

III. EXPERIMENTS AND RESULTS

A. SIMULATION

To reveal the specialty of several different uGCA forms, we intended to compare the characteristics of different uGCA in simulation data of a 5-node network, seen in Fig. 1. Noise terms $\epsilon_i (i = 1, 2, \ldots, 5)$ were Gaussian distribution with mean 0, and their variance were all 0.35. This structural network was named model A, each node was given by

$$\begin{cases} x_{1,t} = 0.68x_{1,t-1} - 0.27x_{1,t-2} + \epsilon_1 \\ x_{2,t} = 0.75x_{1,t-1} - 0.32x_{1,t-2} + 0.73x_{2,t-1} \\ \quad -0.42x_{2,t-2} + 0.63x_{3,t-1} - 0.33x_{3,t-2} + \epsilon_2 \\ x_{3,t} = 0.56x_{2,t-1} - 0.42x_{2,t-2} + 0.62x_{3,t-1} \\ \quad -0.3x_{3,t-2} + 0.78x_{4,t-1} - 0.31x_{4,t-2} + \epsilon_3 \\ x_{4,t} = 0.72x_{2,t-1} - 0.27x_{2,t-2} + 0.52x_{3,t-1} \\ \quad -0.45x_{3,t-2} + 0.73x_{4,t-1} - 0.27x_{4,t-2} + \epsilon_4 \\ x_{5,t} = 0.68x_{4,t-1} - 0.24x_{4,t-2} + 0.805x_{5,t-1} \\ \quad -0.21x_{5,t-2} + \epsilon_5 \end{cases} \quad (20)$$

To ensure the stability of data, we generated 1000 data points in each node and only take the last 300 data points. And ranging the variance of noise can compare the performance of several forms uGCA more comprehensively.

First, the foothold of our model had been verified in our previous research [40]. Causal connectivities between the 5-node networks obtained by different methods showed in Fig. 2, which the connection was single edge between nodes. In conventional GCA, we chose different confidence levels in its F-test and found that the probability of false connections was indeed reduced a lot. For several different forms of uGCA method, the judgment of true connection was very accurate, the same as conventional GCA. Of course, uGCA-TP was still not accurate enough to identified false connections, but uGCA-MIX and uGCA-NML performed very well. Obviously, these two kinds of uGCA can always

	uGCA-TP					uGCA-MIX					uGCA-NML					F(0.05)					F(0.01)				
Var=0.15	0	1000	5	16	0	0	994	2	3	0	0	1000	1	2	0	0	1000	64	69	8	0	1000	17	21	2
	11	0	1000	1000	11	1	0	999	1000	2	4	0	1000	1000	1	31	0	1000	1000	44	9	0	1000	1000	11
	8	1000	0	1000	9	0	994	0	1000	2	0	1000	0	1000	2	10	1000	0	1000	71	4	1000	0	1000	18
	5	8	1000	0	1000	0	3	999	0	999	1	2	1000	0	1000	34	72	1000	0	1000	6	16	1000	0	1000
	10	1	1	0	0	0	0	0	1	0	2	0	0	1	0	44	59	5	18	0	9	22	3	2	0
Var=0.35	0	1000	5	6	2	0	1000	3	3	1	0	1000	1	1	0	0	1000	53	48	5	0	1000	14	12	2
	16	0	1000	1000	15	0	0	1000	1000	1	4	0	1000	1000	0	25	0	1000	1000	62	9	0	1000	1000	14
	5	1000	0	1000	13	0	1000	0	1000	3	0	1000	0	1000	1	12	1000	0	1000	66	7	1000	0	1000	19
	8	7	1000	0	1000	0	3	1000	0	1000	2	1	1000	0	1000	27	75	1000	0	1000	7	23	1000	0	1000
	17	3	0	0	0	0	1	0	0	0	5	1	0	0	0	62	75	4	17	0	16	20	1	4	0
Var=0.55	0	1000	1	15	1	0	1000	0	3	0	0	1000	0	0	0	0	1000	46	65	12	0	1000	12	24	2
	10	0	1000	1000	13	0	0	1000	1000	2	2	0	1000	1000	2	17	0	1000	1000	57	5	0	1000	1000	16
	11	1000	0	1000	20	1	1000	0	1000	8	2	1000	0	1000	5	14	1000	0	1000	98	8	1000	0	1000	29
	7	4	1000	0	1000	0	1	1000	0	1000	2	0	1000	0	1000	35	85	1000	0	1000	9	20	1000	0	1000
	12	0	0	0	0	0	0	0	0	0	5	0	0	0	0	58	65	7	13	0	10	16	0	2	0

FIGURE 2. The result obtained by different forms of uGCA and conventional GCA under different confidence intervals. The variance of noise ranged from 0.15 to 0.55, the length of data points is 300. The number in the sheet represented the count of identified connectivities in 1000 simulations. The row of 5*5 matrix in the figure is the exogenous nodes, the column of 5*5 matrix is the endogenous nodes. Clearly, the darker blue panes in the figure are the connected edges of each node in Fig. 1.

var=0.35

	uGCA-TP					uGCA-MIX					uGCA-NML					F(0.05)					F(0.01)				
Length=150	0	996	6	12	0	0	983	0	4	0	0	989	0	1	0	0	1000	42	40	2	0	993	13	6	0
	23	0	1000	1000	20	3	0	996	995	5	10	0	1000	1000	3	47	0	1000	1000	55	14	0	1000	1000	12
	10	1000	0	997	17	0	989	0	994	1	4	1000	0	994	1	15	998	0	1000	92	3	988	0	999	31
	13	10	1000	0	1000	1	3	996	0	999	3	2	1000	0	1000	36	74	1000	0	1000	12	16	1000	0	1000
	20	7	0	4	0	0	2	1	1	0	4	3	0	1	0	50	62	7	47	0	12	13	1	15	0
Length=200	0	1000	2	13	1	0	1000	1	5	1	0	1000	0	0	0	0	999	43	59	12	0	999	10	15	4
	22	0	1000	1000	10	1	0	1000	1000	4	6	0	1000	1000	2	41	0	1000	1000	59	8	0	1000	1000	11
	9	1000	0	1000	25	0	1000	0	1000	5	2	1000	0	1000	2	15	1000	0	1000	89	5	999	0	1000	28
	5	15	1000	0	1000	1	5	1000	0	1000	4	3	1000	0	1000	27	82	1000	0	1000	4	15	1000	0	1000
	9	5	1	0	0	0	0	0	0	0	3	1	0	0	0	43	63	6	33	0	16	11	1	9	0
Length=250	0	1000	3	8	1	0	1000	1	1	1	0	1000	0	1	0	0	1000	48	64	9	0	1000	11	14	1
	16	0	1000	1000	13	0	0	1000	1000	3	4	0	1000	1000	1	35	0	1000	1000	64	6	0	1000	1000	12
	8	1000	0	1000	26	0	1000	0	1000	5	3	1000	0	1000	1	7	1000	0	1000	101	3	1000	0	1000	30
	8	11	1000	0	1000	0	4	1000	0	1000	1	2	1000	0	1000	32	81	1000	0	1000	9	19	1000	0	1000
	13	4	0	0	0	0	1	0	1	0	1	1	0	0	0	48	71	2	27	0	12	14	2	6	0
Length=500	0	1000	3	11	0	0	1000	1	2	0	0	1000	1	0	0	0	1000	1	0	0	0	1000	1	0	0
	6	0	1000	1000	5	0	0	1000	1000	2	1	0	1000	1000	2	1	0	1000	1000	2	1	0	1000	1000	2
	3	1000	0	1000	4	1	1000	0	1000	1	1	1000	0	1000	1	1	1000	0	1000	1	1	1000	0	1000	1
	3	4	1000	0	1000	0	2	1000	0	1000	0	2	1000	0	1000	0	2	1000	0	1000	0	2	1000	0	1000
	5	1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0

FIGURE 3. The result obtained by different form of uGCA and conventional GCA under different confidence interval. The length of data ranged from 150 to 500 (Length=300 showed in Fig2). The illustration of figure is same as Fig. 2.

find true connection very accurately while rarely obtaining false connections, that is, the true positive rate(TPR) and the true negative rate (TNR) were very high. The performance of conventional GCA eliminating false connections was not very well relatively. Meanwhile, we found all the methods had relatively stable performance under different noise levels (0.15,0.35 and 0.55), except for uGCA-MIX method. At a low noise level, it seemed that uGCA-MIX cannot identify true connections very well, specifically, some true causal effects from node 1 to node 2 and node 3 to node 2 were omitted, seen in Fig. 2.

Fig. 3 showed causal connectivity obtained by uGCA and conventional GCA, in which 5 nodes were generated under the different length of data points. Firstly, for the true connections, uGCA and conventional GCA were nearly the same robust when data points ranged from 200 to 500, the six true connection edges were identified to the most extent. Conventional GCA with a high confidence level was not very accurate when the data point was 150, the same as several uGCA forms. Then, as for false connections, when data points ranged from 150 to 500, uGCA-NML and uGCA-MIX had higher TNR. The uGCA-TP method can also largely eliminate false connections, but the performance of conventional GCA was not very good when its confidence interval was 0.05, except for data length reaching 500. That meant conventional GCA can identify true connections well, but for its subjectivity or inconsistency, it seemed not self-driven enough to eliminate false connections. So far, uGCA method may be similar to conventional GCA with a high confidence level to some extent. Actually, uGCA was not equal to conventional GCA with a high confidence level at all, which had been illustrated clearly in our previous research [40]. For now, it seems that uGCA-NML is more outstanding, which is more like an enhanced version of uGCA-TP, showing same accuracy as uGCA-TP in identifying true connection, but better than uGCA-TP in eliminating false connections. This is also consistent with the different origin of formulas in the principle. Originally, the NML form is proposed to remove the redundant model description. For the mixture form, it behaves more like BIC, depending on the priors for model selection. That means, for some data sets, the performance of uGCA-MIX may be very wonderful, but its robustness is not very well.

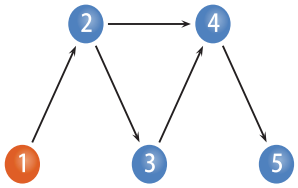


FIGURE 4. The 5-node networks of model B, it was removed two connected edges (3→2 and 4→3) of model A.

To further confirm the difference of characteristics in the several uGCA forms, we thought it necessary to further demonstrate these different approaches in other networks. Therefore, we considered that model A was removed two connected edges (3→2 and 4→3), it arrived at the other 5-node network (named model B), seen in Fig. 4. The other

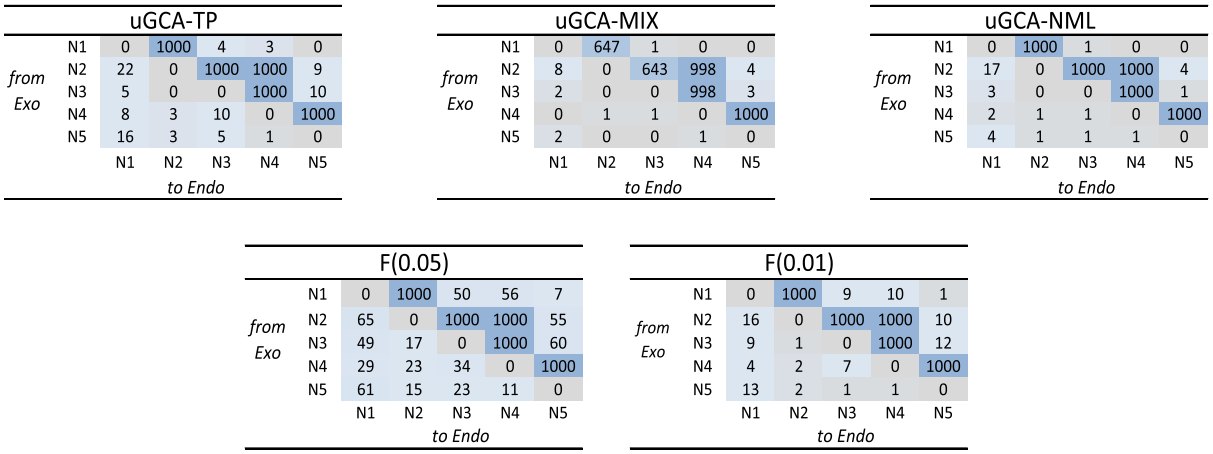


FIGURE 5. The result obtained by the different form of uGCA and conventional GCA with different confidence levels. The variance of noise was 0.15, the length of data points is 300. The *Exo* (the row of 5*5 matrix) denotes the exogenous variables, and the *Endo* (the column of 5*5 matrix) is the endogenous variable. Ni(i=1,2,3,4,5) represents the five nodes in the networks. Same as Fig. 2, the number in the sheet represented the count of identified connectivities from exogenous nodes to endogenous nodes in 1000 simulations. Namely, the darker blue panes denote the connected edges of each node in Fig.4.

setting remains the same as model A, it is given by

$$\begin{cases} x_{1,t} = 0.68x_{1,t-1} - 0.27x_{1,t-2} + \epsilon_1 \\ x_{2,t} = 0.75x_{1,t-1} - 0.32x_{1,t-2} + 0.73x_{2,t-1} \\ \quad -0.42x_{2,t-2} + \epsilon_2 \\ x_{3,t} = 0.56x_{2,t-1} - 0.42x_{2,t-2} + 0.62x_{3,t-1} \\ \quad -0.3x_{3,t-2} + \epsilon_3 \\ x_{4,t} = 0.72x_{2,t-1} - 0.27x_{2,t-2} + 0.52x_{3,t-1} \\ \quad -0.45x_{3,t-2} + 0.73x_{4,t-1} - 0.27x_{4,t-2} + \epsilon_4 \\ x_{5,t} = 0.68x_{4,t-1} - 0.24x_{4,t-2} + 0.805x_{5,t-1} \\ \quad -0.21x_{5,t-2} + \epsilon_5 \end{cases} \tag{21}$$

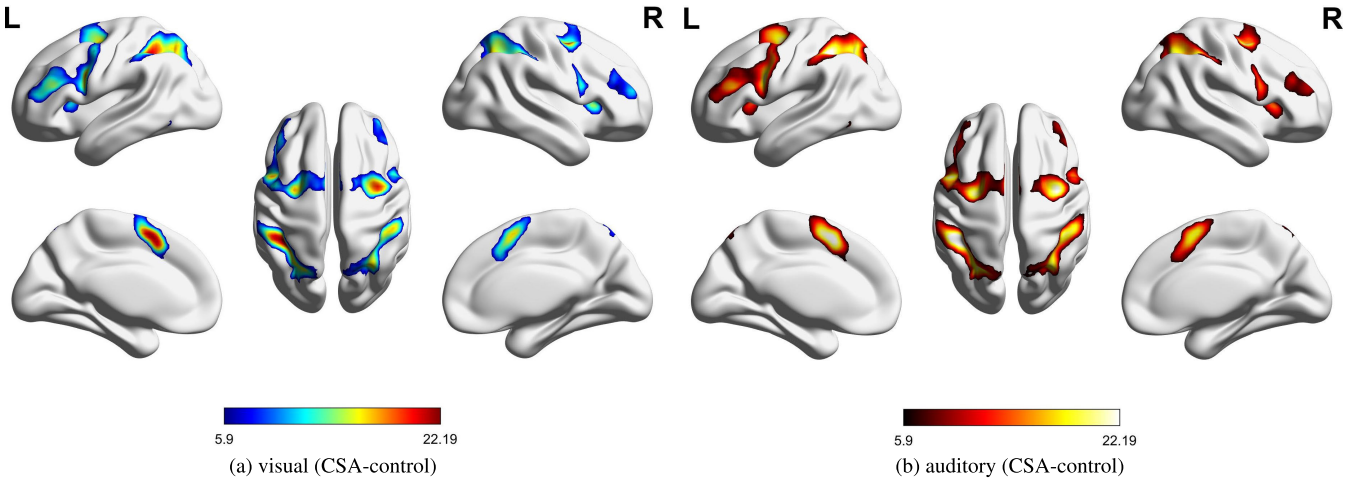
Consistent of the consequence in model A, the results of model B showed that uGCA-TP and uGCA-NML was relatively more robust in the connection networks, seen in Fig. 5. For uGCA-NML, it obtained a more sparse connection network and guaranteed a higher accuracy at the same time. But for uGCA-MIX at low noise level, its false connections can be eliminated well but the true connections were not identified accurately, seen in Fig. 5. This meant uGCA-MIX will obtain a more sparse connected matrix but that there may be some false negative for its true connections at some noise level. For conventional GCA, our previous research had executed a model similar to the structure of model B, it found that there were some true causal connectivities missed to be identified [40]. As a whole, uGCA was not equivalent to the conventional GCA with a high confidence level at all whether from the performance or their principle, and it may have more robust performance in complicated networks.

Actually, due to the subjective statistical inference process and inconsistent mathematical theories, the inherent causality between nodes cannot be guaranteed when conventional GCA identifies causal connectivity. Our previous research had demonstrated, for conventional GCA, true connections that cannot be well identified were all about the drive node, which may be related to overlapping or mutual suppression of incoming information of the other nodes, and there seemed to be more related nodes. But on the contrary, under a unified mathematical framework, uGCA methods did not have such mistakes, except for uGCA-MIX under low noise level. Furthermore, there were some interesting phenomena in the false connections from other nodes to the driver node 1 in two models. It seemed harder than other false connectivities for all methods to completely eliminate them. These misjudge may be considered to be accidental, but due to their higher TNR of uGCA-MIX and uGCA-NML, it's more reasonable to interpret it as the characteristics related to driver node. Theoretically, all incoming information of other nodes comes from the driver node, so there will be some causal aliasing in time series.

In general, several different forms of uGCA were more robust than conventional GCA in two 5-node simulations. As for uGCA-NML form, it seemed to be more admirable in investigating causality when data points were above 200. Meanwhile, at some noise level, uGCA-MIX seemed to be a reliable choice to obtain a more sparse network, and its performance is similar to that of conventional GCA with a high confidence level. For the uGCA-TP form, its overall performance was between these two forms of uGCA, which can be a good conservative choice.

B. ISOMORPHIC MAPPING IN MENTAL ARITHMETIC

The validity of causal investigation within several uGCA forms has shown in two simulated models. Next, it is necessary to verify this causal connectivity analysis of uGCA methods through real imaging data. In the study we let ten subjects perform simple one-digit(consisting of 1-10) serial addition (SSA) and complex two-digit(consisting of 1-5) serial addition (CSA) by visual stimulus and



(a) visual (CSA-control) (b) auditory (CSA-control)

FIGURE 6. **Mental arithmetic of CSA-control state under the two stimuli(visual stimulus and auditory stimulus), the activation regions were processed by SPM12, the control state meant that the sample was in rest state without mental arithmetic. (a): CSA-control state under visual stimulus. (b): CSA-control state under auditory stimulus.($P<0.0001$, uncorrected).**

simultaneously measured their brain activities with fMRI. Immediately following, the subjects were asked to perform the same serial addition arithmetic tasks by an auditory stimulus. Nine right-handed healthy subjects (four female, 24 ± 1.5 years old) and one left-handed healthy female subject (24 years old) participated. One of the subject's(a right-hand male) experimental data was removed due to excessive head motion. All subjects volunteered to participate in this study with the informal written consent by themselves.

In this experiment, different samples performed mental arithmetic tasks through visual and auditory stimuli, respectively. During tasking, these working scenarios of the brain were mental arithmetic tasks, thus working scenarios can be considered similar regardless of specific stimuli. At the same time, brain connectivity should be isomorphic when the subject performs mental arithmetic regarding a high cognitive task, that is to say, the phenomenon of isomorphic mapping should be widely present in the brain during the completion of this task. Then through SPM software, we can get the mental arithmetic activation area of the brain. In these mapped regions through statistical inference, we used several causal connectivity methods to obtain connections between networks. Finally, by comparing their similarities of mental arithmetic networks under different stimuli, the phenomenon of isomorphic mapping in processing mental arithmetic tasks was quantitatively described to demonstrate the advantages of several uGCA forms in investigating the dynamic correlation of brain regions [40], [65].

In Euclid space, for different methods, we measured the similarity of two networks under visual/auditory stimulus. This similarity was given by

$$SC = \frac{1}{1 + |A_v - A_a|}, \tag{22}$$

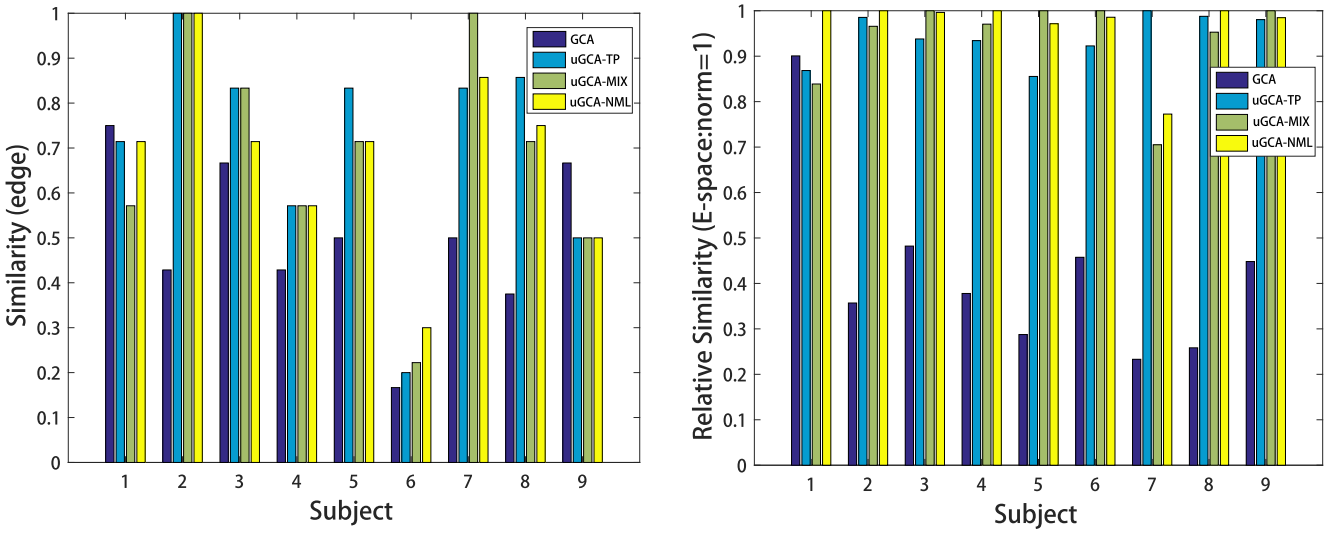
where A_v and A_a was the connection matrix under visual/auditory respectively. Specifically, in conventional GCA, the value in connection matrix was a residual ratio while it was a description length difference in several uGCA forms.

By setting the values of the connection matrix to 0 and 1, we obtain some connected networks (A'_v, A'_a) with direction and no weight. For this 0-1 connection matrix, we measured similarity by combining the same edges in obtained network. Consequently, the similarity of connected edge was given

$$SE = \frac{\sum\sum(A'_v \cap A'_a)}{\sum\sum(A'_v \cup A'_a)}. \tag{23}$$

For most subjects, their similarities of mental arithmetic networks were higher than 60%, seen in Fig. 7a. Obviously, this similarity of edges in causal connectivity networks was sufficient to show the isomorphic mapping between two stimuli during the mental arithmetic tasks in each subject. At the same time, the causal network identified by uGCA was obviously more similar than conventional GCA. Specifically, for subject 1 and subject 3, network similarities of all methods were relatively high, indicating that the brain isomorphic mapping on mental arithmetic task may be relatively more robust. In subject 2, for several forms of uGCA methods, results found that the obtained mental arithmetic network under visual/auditory is exactly identical. Furthermore, in subject 8, uGCA-MIX also obtained two identical networks.

Comparing their relative performance of different methods, this resulting similarity showed in Fig. 7b, which was divided by a maximum of SC of different methods on each subject. Intuitively, at the individual level, the similarity of connection matrix obtained by uGCA method was much higher than that of conventional GCA, which verified its excellent performance of this consistency of mathematical principle in uGCA method. Meanwhile, with help of obtained connection matrix, it found that the similarity of networks obtained by uGCA-NML was more robust, which meant that a sparse matrix obtained can guarantee higher TNR and TPR.



(a) Similarity (Edge)

(b) Similarity (Connection Matrix)

FIGURE 7. The similarity of the obtained mental arithmetic networks under two stimuli(visual stimulus and auditory stimulus). (a): The similarity of each subjects was measured by SE. (b): To compare the difference between methods, this similarity was obtained by quantifying distance in connection matrix, which obtained by a difference in description length.

To further illustrate the difference of this quantified isomorphic mapping phenomenon, causal networks obtained by different methods on individuals showed in Fig. 8 respectively. For causal network obtained by conventional GCA and uGCA-TP, their result had shown in [65]. Different from the irregularity of networks obtained by conventional GCA, causal networks obtained by uGCA-TP were very similar.

For uGCA-MIX form, causal networks of subject 2 and subject 8 were identical, seen in Fig. 8. But for subject 1, most connected edges were identical when a few edges were different. As mentioned in simulation, uGCA-MIX will obtain a more sparse network, but its networks may miss some true connections. Anyway, the identical mental arithmetic networks obtained through uGCA-MIX showed that this isomorphic mapping phenomenon of three subjects was very legible, it meant that the ability of subjects to perform mental arithmetic tasks may be more prominent.

For uGCA-NML form, its causal networks also showed in Fig. 8. In subject 1, there were two different edges, other edges were the same in networks. But we can still conclude that two networks had an isomorphic characteristic, for example, their driving node may be identical. For subject 2, whose causal networks were identical, isomorphism mapping under mental arithmetic task quantified by uGCA-NML was very legible. As for subject 8, only one edge was different, and it's clear that the two networks were almost identical. For these three subjects, mental arithmetic networks obtained through uGCA-NML all had the phenomenon of isomorphic mapping obviously. As we mentioned above, uGCA-NML can identify true connections well when eliminating the influence of false connections, then to obtain a more sparse connection matrix.

IV. DISCUSSION

In simulation experiments, uGCA methods had shown superior performance by comparing conventional GCA in 5-node networks. Whether it was true or false connections, uGCA had a greater chance to find real causal connectivities than conventional GCA. We found that uGCA-MIX preferred to find a more sparse causal network duo to its priors on parameter estimation, and uGCA-NML, no matter what the noise level was, can eliminate the influence of false connections better when found real connections, so as to get sparse connection matrix more accurately. The performance of uGCA-TP was the compromise between the other two methods.

In fMRI experiment, the isomorphic mapping involving mental arithmetic tasks in the brain was constrained into metric space to verify the performance of investigating causal connectivity for different methods. Overall, we found that the mental arithmetic networks obtained by uGCA were more similar, and isomorphic phenomenon seemed more obvious. In this research, causal networks obtained by uGCA-MIX and uGCA-NML methods showed that their mental arithmetic networks under different stimuli also existed a legible isomorphic mapping. In real imaging datasets, the requirements for comprehensive performance of these methods are much higher, due to the length of experimental data was between 200-300, advantages of uGCA-NML form are not very evident. For uGCA-MIX, misjudging a few true connections will lead to results inaccurately. As a comparison with conventional GCA, only a few subjects seemed to show clear isomorphism. We consider the difference is that uGCA integrates the conventional two-stage GCA scheme into a unified framework, and isomorphic mapping is a continuous closed process, which requires that process of quantitative isomorphism must keep the consistency of mathematical principles, otherwise there will be a breakpoint in quantitative process. In mathematics, we call it a singular point, and its operation is not closed, which its result may be a departure from the original space and become very distorted. And we know the

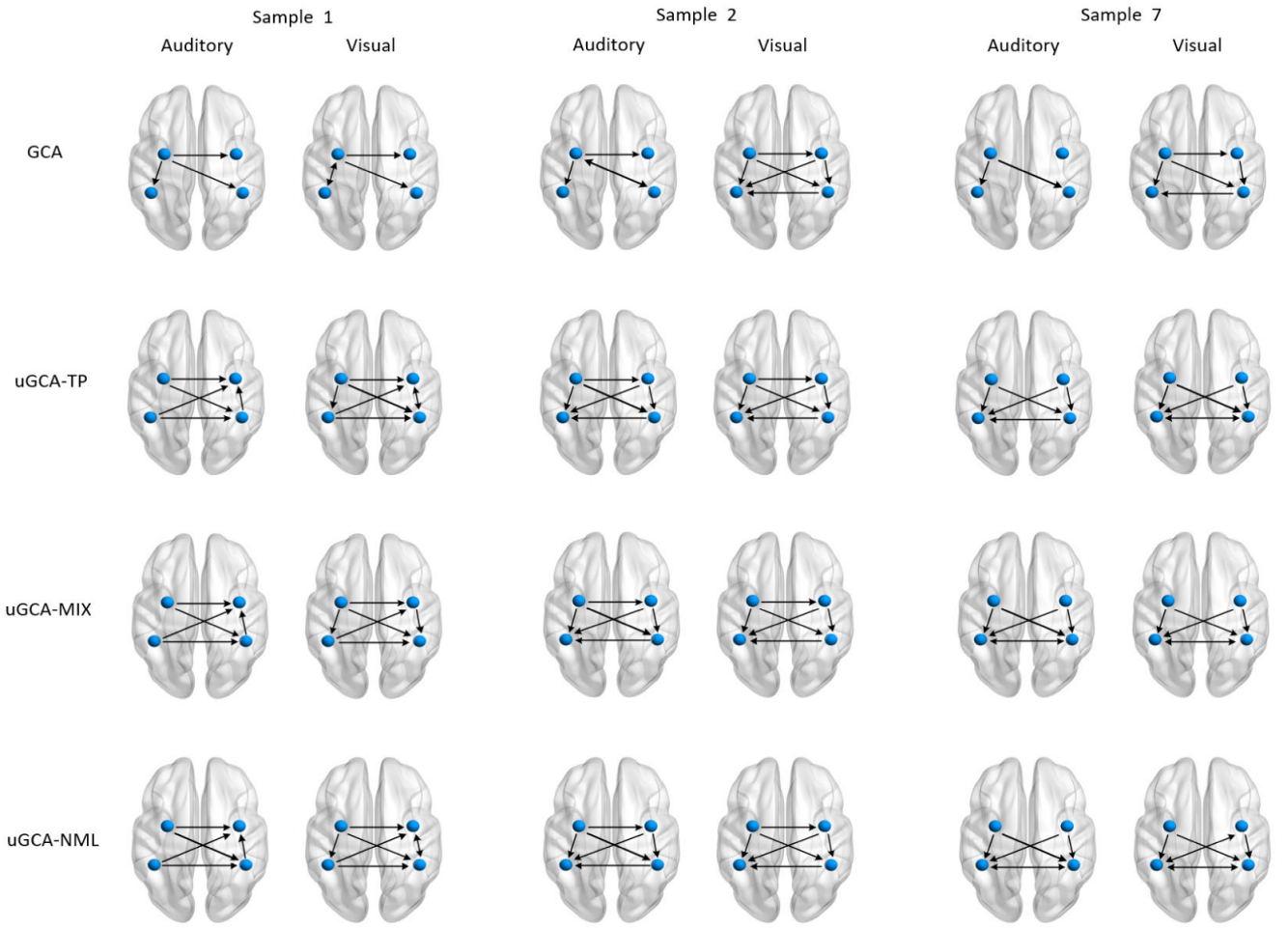


FIGURE 8. Causal network in the mental arithmetic tasks obtained by uGCA methods and conventional GCA, respectively. With the conventional GCA approach, connected edges of causal networks in two different stimuli were to a large extent distinct. In contrast, for uGCA methods, their connection networks commonly showed high similarities, only about one or two edges were different. Especially in subject 2, three uGCA all obtained two identical networks. For subject 8, uGCA-MIX and uGCA-NML also obtained two identical networks while there is only one different edge in the two networks for uGCA-TP. As for subject 1, only two connected edges were different for the networks obtained by uGCA-TP and uGCA-NML, and whose information flows were similar to a large extent. For uGCA-MIX, there were also two different edges. But because it is susceptible to noise and is not so stable, there was an opposite connected edge.

original model space can not be found at all. Thus, toward description length of the model complexity, uGCA-MIX and uGCA-NML provided some different solutions, which mapping the descriptive model into different feature spaces to approach the original model space in different aspects.

These results obtained by uGCA indicated that causal isomorphism does exist during mental arithmetic tasks. Actually, the postulation that the brain is isomorphic under similar tasks is not made up out of thin air. For decades, some studies have tried to demonstrate this capability that the brain cognizes the world by analogy [66]–[72]. And there are also some studies that use category theory to mathematically demonstrate how analogical reasoning in the human brain get rid of the spurious inferences that puzzle conventional artificial intelligence model (called systematicness) [73]–[75].Hence, it's reasonable to believe that the causal networks of brain during mental arithmetic tasks are isomorphic, even in different modes of stimuli. Specifically, this tasking network should be similar in the feature space of causal connectivity.

Intuitively, at the individual level, their causal networks under different stimuli should be similar for one subject, which means the brain activity should be isomorphic mapping. In our mental arithmetic experiment, uGCA showed the legible isomorphism of the brain under different stimuli. In contrast, the causal network obtained by conventional GCA was not very similar, or isomorphism was not clear either. At the same time, for the similarity of mental arithmetic brain network, our previous research also found a relatively legible isomorphic mapping through the method of dynamic causal modeling (DCM) [65].

For conventional two-stage approach of GCA, its causal investigating process may obscure original relationship of data from the same space and then leads to an unreliable inference. This may be one of the mathematical reasons behind the

distrust of GCA by its peers when it was introduced from economics to neuroscience. On the one hand, the imaging data space in a brain should also be closed from a mathematical viewpoint. A modeling tool that describes the brain through the imaging data should change the original closed space as little as possible, so as to ensure that the internal correlation of data is not distorted so much. Therefore, as a method of describing effective connectivities related to brain behaviors, it is critical to maintain the consistency of mathematical theory throughout the description process. Our uGCA paradigm first maps the original space into a unified description length guided space, which is also closed topologically, and then to identify the causal connectivities. Therefore, this allows data to maintain the original correlation as much as possible, thus obtaining an optimal approximate description of correlation among data in the original space. More importantly, different uGCA forms provided different sides to approach the original space, uGCA-MIX and uGCA-NML can both obtain their optimal descriptive model in different feature spaces respectively. In general, based on the MDL principle, causal connectivity of brain regions is investigated by initiating a unified description length guided framework, which ensures closure and continuity of the description process to obtain a more convincing result. Furthermore, in the current work for distributed collaborative brain, the isomorphism between tasks may be generalized existed, which requires that the method used to describe brain behavior should be consistent and follows the same mathematical theory.

V. CONCLUSION

From the parsimony principle in modeling, we propose that the causal connectivity can be obtained by initiating a unified framework guided by description length based on the MDL principle, which guarantees continuity and closure of the original correlation of data in processing, so as to obtain an optimal approximate description for data. In this article, we developed several uGCA forms to approach the original descriptive space from different sides. We verified the superiority of several uGCA forms in simulated data and real fMRI experiments. Meanwhile, further derivation of uGCA needs to reveal and investigate nonlinear phenomena in the brain, this is one of our future works.

REFERENCES

[1] C. L. Martin and M. Chun, "The BRAIN initiative: Building, strengthening, and sustaining," *Neuron*, vol. 92, no. 3, pp. 570–573, Nov. 2016.
[2] K. Amunts, C. Ebell, J. Müller, M. Telefont, A. Knoll, and T. Lippert, "The human brain project: Creating a European research infrastructure to decode the human brain," *Neuron*, vol. 92, no. 3, pp. 574–581, Nov. 2016.
[3] H. Okano, E. Sasaki, T. Yamamori, A. Iriki, T. Shimogori, Y. Yamaguchi, K. Kasai, and A. Miyawaki, "Brain/MINDS: A Japanese national brain project for marmoset neuroscience," *Neuron*, vol. 92, no. 3, pp. 582–590, Nov. 2016.
[4] Australian Brain Alliance Steering Committee, "Australian brain alliance," *Neuron*, vol. 92, no. 3, pp. 597–600, 2016.
[5] S.-J. Jeong, H. Lee, E.-M. Hur, Y. Choe, J. W. Koo, J.-C. Rah, K. J. Lee, H.-H. Lim, W. Sun, C. Moon, and K. Kim, "Korea brain initiative: Integration and control of brain functions," *Neuron*, vol. 92, no. 3, pp. 607–611, Nov. 2016.
[6] M.-M. Poo, J.-L. Du, N. Y. Ip, Z.-Q. Xiong, B. Xu, and T. Tan, "China brain project: Basic neuroscience, brain diseases, and brain-inspired computing," *Neuron*, vol. 92, no. 3, pp. 591–596, Nov. 2016.
[7] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: Emerging methods and principles," *Trends Cognit. Sci.*, vol. 14, no. 6, pp. 277–290, Jun. 2010.
[8] K. J. Friston, "Functional and effective connectivity: A review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, Jan. 2011.
[9] M. Siegel, T. H. Donner, and A. K. Engel, "Spectral fingerprints of large-scale neuronal interactions," *Nature Rev. Neurosci.*, vol. 13, no. 2, pp. 121–134, Feb. 2012.
[10] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, "Dynamic functional connectivity: Promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, Oct. 2013.
[11] S. L. Valk, B. C. Bernhardt, A. Böckler, F.-M. Trautwein, P. Kanske, and T. Singer, "Socio-cognitive phenotypes differentially modulate large-scale structural covariance networks," *Cerebral Cortex*, vol. 27, no. 2, Jan. 2016, Art. no. bhv319.
[12] Y. Li, M. Lei, W. Cui, Y. Guo, and H.-L. Wei, "A parametric time-frequency conditional Granger causality method using ultra-regularized orthogonal least squares and multiwavelets for dynamic connectivity analysis in EEGs," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3509–3525, Dec. 2019.
[13] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003.
[14] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.
[15] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, Feb. 2010.
[16] K. E. Stephan and A. Roebroeck, "A short history of causal modeling of fMRI data," *NeuroImage*, vol. 62, no. 2, pp. 856–863, Aug. 2012.
[17] J. Tsunada, A. S. K. Liu, J. I. Gold, and Y. E. Cohen, "Causal contribution of primate auditory cortex to auditory perceptual decision-making," *Nature Neurosci.*, vol. 19, no. 1, pp. 135–142, Jan. 2016.
[18] T. S. Zarghami and K. J. Friston, "Dynamic effective connectivity," *NeuroImage*, vol. 207, Feb. 2020, Art. no. 116453.
[19] T. T. Fernandes, B. Direito, A. Sayal, J. Pereira, A. Andrade, and M. Castelo-Branco, "The boundaries of state-space Granger causality analysis applied to BOLD simulated data: A comparative modelling and simulation approach," *J. Neurosci. Methods*, vol. 341, Jul. 2020, Art. no. 108758.
[20] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.
[21] C. W. J. Granger and P. Newbold, "Spurious regressions in econometrics," *J. Econometrics*, vol. 2, no. 2, pp. 111–120, Jul. 1974.
[22] A. Tank, I. Covert, N. Foti, A. Shojaie, and E. Fox, "Neural Granger causality for nonlinear time series," 2018, *arXiv:1802.05842*. [Online]. Available: http://arxiv.org/abs/1802.05842
[23] M. Farokhzadi, G.-A. Hossein-Zadeh, and H. Soltanian-Zadeh, "Nonlinear effective connectivity measure based on adaptive neuro fuzzy inference system and Granger causality," *NeuroImage*, vol. 181, pp. 382–394, Nov. 2018.
[24] N. Talebi, A. M. Nasrabadi, I. Mohammad-Rezazadeh, and R. Coben, "NCREANN: Nonlinear causal relationship estimation by artificial neural network; applied for autism connectivity study," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2883–2890, Dec. 2019.
[25] A. Hatemi-J, "Asymmetric causality tests with an application," *Empirical Econ.*, vol. 43, no. 1, pp. 447–456, Aug. 2012.
[26] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *J. Amer. Stat. Assoc.*, vol. 77, no. 378, pp. 304–313, Jun. 1982.
[27] P. A. Stokes and P. L. Purdon, "A study of problems encountered in Granger causality analysis from a neuroscience perspective," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 34, pp. E7063–E7072, Aug. 2017.
[28] L. Ning and Y. Rathi, "A dynamic regression approach for frequency-domain partial coherence and causality analysis of functional brain networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 1957–1969, Sep. 2018.

[29] W. Liao, D. Marinazzo, Z. Pan, Q. Gong, and H. Chen, "Kernel Granger causality mapping effective connectivity on fMRI data," *IEEE Trans. Med. Imag.*, vol. 28, no. 11, pp. 1825–1835, Nov. 2009.

[30] Z. Chen, K. Zhang, L. Chan, and B. Schölkopf, "Causal discovery via reproducing kernel Hilbert space embeddings," *Neural Comput.*, vol. 26, no. 7, pp. 1484–1517, Jul. 2014.

[31] H. Guo, W. Zeng, Y. Shi, J. Deng, and L. Zhao, "Kernel Granger causality based on back propagation neural network fuzzy inference system on fMRI data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1049–1058, May 2020.

[32] R. S. Hacker and A. Hatemi-J, "Tests for causality between integrated variables using asymptotic and bootstrap distributions: Theory and application," *Appl. Econ.*, vol. 38, no. 13, pp. 1489–1500, Jul. 2006.

[33] M. Kaminski, *Multichannel Data Analysis in Biomedical Research*, 2nd ed. Pacific Grove, CA, USA: Duxbury Resource Center, 2007.

[34] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, vol. 100, no. 14, Apr. 2008, Art. no. 144103.

[35] C. L. Zou, C. Ladroue, S. X. Guo, and J. F. Feng, "Identifying interactions in the time and frequency domains in local and global networks—A Granger causality approach," *BMC Bioinf.*, vol. 337, no. 11, pp. 1–17, 2010.

[36] A. M. DSouza, A. Z. Abidin, L. Leistritz, and A. Wismüller, "Exploring connectivity with large-scale Granger causality on resting-state functional MRI," *J. Neurosci. Methods*, vol. 287, pp. 68–79, Aug. 2017.

[37] X. Wang, R. Wang, F. Li, Q. Lin, X. Zhao, and Z. Hu, "Large-scale Granger causal brain network based on resting-state fMRI data," *Neuroscience*, vol. 425, pp. 169–180, Jan. 2020.

[38] X. Li, G. Marrelec, R. F. Hess, and H. Benali, "A nonlinear identification method to study effective connectivity in functional MRI," *Med. Image Anal.*, vol. 14, no. 1, pp. 30–38, 2010.

[39] L. Barnett and A. K. Seth, "Detectability of Granger causality for subsampled continuous-time neurophysiological processes," *J. Neurosci. Methods*, vol. 275, pp. 93–121, Jan. 2017.

[40] F. Li, X. Wang, Q. Lin, and Z. Hu, "Unified model selection approach based on minimum description length principle in Granger causality analysis," *IEEE Access*, vol. 8, pp. 68400–68416, 2020.

[41] R. L. Wasserstein and N. A. Lazar, "The ASA's statement on *p*-values: Context, process, and purpose," *Amer. Statistician*, vol. 70, no. 2, pp. 129–133, 2016.

[42] D. J. Benjamin, J. O. Berger, M. Johannesson, B. A. Nosek, E.-J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer, and D. Cesarini, "Redefine statistical significance," *Nature Hum. Behav.*, vol. 2, no. 1, pp. 6–10, 2018.

[43] V. Amrhein, S. Greenland, and B. McShane, "Scientists rise up against statistical significance," *Nature*, 2019.

[44] R. L. Wasserstein and N. A. Lazar, "ASA statement on statistical significance and *p*-values," in *The Theory of Statistics in Psychology*. Cham, Switzerland: Springer, 2020, pp. 1–10.

[45] A. Kolmogorov, "Three approaches to the quantitative definition of information," *Problems Inf. Transmiss.*, vol. 1, no. 1, pp. 1–7, 1965.

[46] A. Kolmogorov, "Logical basis for information theory and probability theory," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 5, pp. 662–664, Sep. 1968.

[47] C. S. Wallace and D. M. Boulton, "An information measure for classification," *Comput. J.*, vol. 11, no. 2, pp. 185–194, 1968.

[48] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[49] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *J. Amer. Statist. Assoc.*, vol. 96, no. 454, pp. 746–774, 1998.

[50] P. D. Grünwald, *The Minimum Description Length Principle*. Cambridge, MA, USA: MIT Press, 2007.

[51] P. G. Bryant and O. I. Cordero-Brana, "Model selection using the minimum description length principle," *Amer. Statistician*, vol. 54, no. 4, pp. 257–268, 2000.

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

[53] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[54] P. D. Grünwald, I. J. Myung, and M. A. Pitt, *Advances in Minimum Description Length: Theory and Applications* (Neural Information Processing series). Cambridge, MA, USA: MIT Press, 2005.

[55] M. Hansen and B. Yu, "Bridging AIC and BIC: An MDL model selection criterion," in *Proc. Worskhop Detection, Estimation, Classification, Imag.*, Santa Fe, NM, USA, 1999, pp. 24–26.

[56] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inf. Theory*, vol. 36, no. 3, pp. 453–471, May 1990.

[57] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.

[58] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.

[59] J. Rissanen, "A predictive least-squares principle," *IMA J. Math. Control Inf.*, vol. 3, nos. 2–3, pp. 211–222, 1986.

[60] Y. M. Shtar'kov, "Universal sequential coding of single messages," *Problemy Peredachi Informatsii*, vol. 23, no. 3, pp. 3–17, 1987.

[61] J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 40–47, Jan. 1996.

[62] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 11, no. 2, pp. 416–431, Jun. 1983.

[63] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*. Amsterdam, The Netherlands: Elsevier, 1986.

[64] J. Rissanen, "MDL denoising," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2537–2543, Nov. 2000.

[65] F. Li, X. Wang, P. Shi, Q. Lin, and Z. Hu, "Neural network can be revealed with functional MRI: Evidence from self-consistent experiment," *Nature Commun.*, 2020.

[66] D. L. Medin, R. L. Goldstone, and D. Gentner, "Respects for similarity," *Psychol. Rev.*, vol. 100, no. 2, p. 254, 1993.

[67] D. Gentner and A. B. Markman, "Structure mapping in analogy and similarity," *Amer. Psychologist*, vol. 52, no. 1, p. 45, 1997.

[68] D. R. Hofstadter, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York, NY, USA: Basic Books, 1995.

[69] D. R. Hofstadter, "Analogy as the core of cognition," in *The Analogical Mind: Perspectives From Cognitive Science*. Cambridge, MA, USA: 2001, pp. 499–538.

[70] D. Hofstadter, *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York, NY, USA: Basic books, 2008.

[71] D. R. Smith and M. T. Keane, "Analogical problem solving," *Amer. J. Psychol.*, vol. 103, no. 4, p. 581, 1990.

[72] J. E. Hummel and K. J. Holyoak, "Distributed representations of structure: A theory of analogical access and mapping," *Psychol. Rev.*, vol. 104, no. 3, p. 427, 1997.

[73] S. Phillips and W. H. Wilson, "Categorial compositionality: A category theory explanation for the systematicity of human cognition," *PLoS Comput. Biol.*, vol. 6, no. 7, Jul. 2010, Art. no. e1000858.

[74] S. Phillips and W. H. Wilson, "Categorial compositionality II: Universal constructions and a general theory of (quasi-) systematicity in human cognition," *PLoS Comput. Biol.*, vol. 7, no. 8, Aug. 2011, Art. no. e1002102.

[75] S. Phillips and W. H. Wilson, "Categorial compositionality III: F-(co) algebras and the systematicity of recursive capacities in human cognition," *PLoS ONE*, vol. 7, no. 4, Apr. 2012, Art. no. e35028.

ZHENGHUI HU received the Ph.D. degree in physics from Zhejiang University, Hangzhou, China, in 2005. From 2005 to 2007, he was a Postdoctoral Fellow in electrical and computer engineering with The Hong Kong University of Science and Technology. From 2008 to 2010, he was a Postdoctoral Fellow in computing and information science with the Rochester Institute of Technology, Rochester, NY, USA. Since 2015, he has been a Professor with the College of Science, Zhejiang University of Technology, Hangzhou, China. His research interests include functional brain imaging, magnetocardiography, and signal reconstruction.

FEI LI was born in Hubei, China, in 1995. He received the M.S. degree from the College of Science, Zhejiang University of Technology, Hangzhou, Zhejiang, China, in 2017, where he is currently pursuing the Ph.D. degree.

XUEWEI WANG was born in Shanxi, China, in 1997. He is currently pursuing the M.S. degree with the College of Science, Zhejiang University of Technology, Hangzhou, Zhejiang, China.

QIANG LIN was born in Zhejiang, China, in 1964. He received the M.S. degree in optics from Hangzhou University in 1988 and the Ph.D. degree in optics from Zhejiang University in 2003. In 1994, he became a Professor and the Head of the Optics Division, Department of Physics, Hangzhou University. From 1996 to 1997, he was with the Technical University of Vienna, Austria. In 1998, he joined the Department of Physics, Zhejiang University, and became the Director of the Optics Institute. From 2005 to 2006, he was an Alexander von Humboldt Fellow with the Max-Born-Institute of Berlin, Germany. Since 2014, he has been the Dean of the College of Sciences, Zhejiang University of Technology. He has authored or coauthored more than 200 scientific articles and three books.