Nonlinear Conditional Time-varying Granger Causality of Task fMRI Data via Deep Stacking Networks

Kai-Cheng Chuang[1,2] KAI.CHUANG@PBRC.EDU
[1] *Medical Physics Graduate Program, Louisiana State University, Baton Rouge, LA, USA*
[2] *Biomedical Imaging Center, Pennington Biomedical Research Center, Baton Rouge, LA, USA*

Sreekrishna Ramakrishnapillai[2] SREEKRISHNA.RAMAKRISHNAPILLAI@PBRC.EDU
Lydia Bazzano[3] LBAZZANO@TULANE.EDU
[3] *Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, LA, USA*

Owen Carmichael[2] OWEN.CARMICHAEL@PBRC.EDU

Editors: Under Review for MIDL 2022

Abstract

Time-varying Granger causality refers to patterns of causal relationships that vary over time between brain functional time series at distinct source and target regions. It provides rich information about the spatiotemporal structure of brain activity that underlies behavior. Current methods for this problem fail to quantify nonlinear relationships in source-target relationships, and require ad hoc setting of relationship time lags. This paper proposes deep stacking networks (DSN), with adaptive convolutional kernels (ACK) as component parts, to address these challenges. The DSN use convolutional neural networks to estimate nonlinear source-target relationships, ACK allow these relationships to vary over time, and time lags are estimated by analysis of ACK coefficients. When applied to synthetic data and data simulated by the STANCE fMRI simulator, the method identified ground-truth time-varying causal relationships and time lags more robustly than competing methods. The method also identified more biologically-plausible causal relationships in a real-world task fMRI dataset than a competing method. Our method is promising for modeling complex functional relationships within brain networks.

Keywords: Granger causality, Adaptive convolutional kernels, fMRI

1. Introduction

Effective connectivity refers to the influence that functional activity in one brain system exerts over functional activity in another (Friston, 2011). It has become an important tool to understand the organization of human neural circuitry underlying perception and cognition in health and disease based on time series data from functional neuroimaging methods such as functional magnetic resonance imaging (fMRI) (Seth et al., 2015; Deshpande et al., 2009). Granger causality, the predominant method for quantifying effective connectivity, assesses the degree to which time series data at a current time point in a target region is predicted by time series data at the current or earlier time points from a different (source) region, after accounting for the influence of other sources and target regions data from previous time points (Granger, 1969; Friston et al., 2013; Goebel et al., 2003). This method has proven useful for clarifying various aspects of brain dynamics (Liao et al., 2009; Zhou et al., 2011).

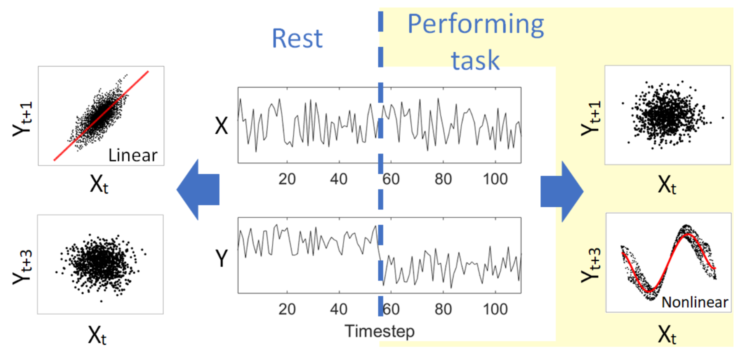

Figure 1: A hypothetical time-varying causal relationship, where source X_t has a strong linear relationship with target Y_{t+1} during rest (lefthand side), and X_t has a nonlinear relationship with Y_{t+3} during task performance (righthand side). This paper presents a method for automatically discovering such time-varying causal relationships.

Because connectivity relationships between brain regions are believed to change dynamically over the course of task performance (Ambrosi et al., 2021; Marcinkevičs and Vogt, 2021; Sato et al., 2006), and even during periods of rest (Cekic et al., 2018), extensions of Granger causality that quantify **time-varying causal relationships** (Figure 1) have the potential for high impact. To date, three solutions to this problem have been presented, all based on the vector autoregressive (VAR) model, which allows modeling of linear relationships between source and target (Ambrosi et al., 2021; Marcinkevičs and Vogt, 2021; Sato et al., 2006). Time-varying VAR parameters were estimated using wavelet functions (Sato et al., 2006), generalized VAR (GVAR) (Marcinkevičs and Vogt, 2021), and particle filtering (PF) (Ambrosi et al., 2021).

This paper seeks to overcome two key limitations of prior time-varying Granger causality methods. First, the prior methods were only able to model linear relationships between source and target signals, thus precluding modeling of nonlinear causal relationships that are expected to arise from complex neural dynamics (Liao et al., 2009; Marinazzo et al., 2011; Schoukens and Ljung, 2019; Príncipe et al., 2011). Second, prior methods for time-varying Granger causality were limited in their ability to handle time lags: the number of timesteps that elapse between causal brain activity at the source and resulting brain activity at the target. One prior method limited the time lag to exactly one time point to reduce computational complexity (Ambrosi et al., 2021), while the other methods required the user to specify the time lag *a priori* (Marcinkevičs and Vogt, 2021; Sato et al., 2006). This is important because time lags are not expected to be known *a priori*.

We propose to use deep stacking networks (DSN) to overcome these limitations. DSN allow estimation of nonlinear Granger causality between source (X_t) and target (Y_t), after accounting for the influence of activity in other source regions (Z_t), using convolutional neural network (CNN) modules; stacking multiple such modules allows modeling among multiple sources and targets and each CNN module efficiently capture temporally-localized features between a pair of source and target (Figure 2 (a)). Within each CNN is an adaptive convolutional kernel (ACK), whose estimated kernel coefficients reveal time-varying causal relationships and time lags at each time point in the time series (Figure 2 (b)). This approach extends our previous study, which used DSN to estimate nonlinear Granger causality but was unable to handle time-varying causal relationships (Chuang et al., 2021). We show

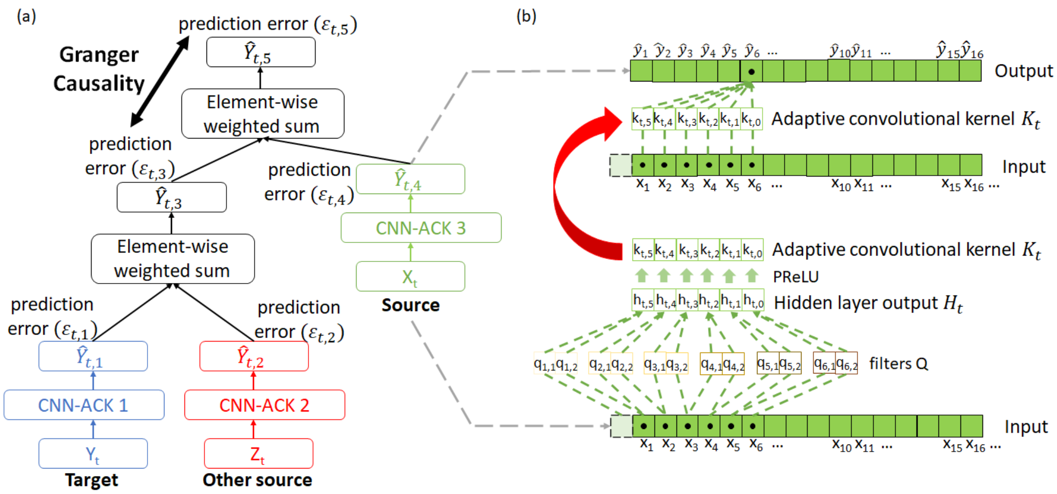

Figure 2: **(a)** Proposed DSN with CNN-ACKs to estimate nonlinear time-varying Granger causality between source X_t and target Y_t, conditioned on source Z_t. **(b) Top:** Target time series Y_t is modeled in terms of a source time series X_t that is convolved by an ACK K_t, whose values change over the course of time (t). **Bottom:** K_t is estimated at training time through estimation of six (1 x 2) convolutional filters Q which are applied to the input source time series X_t, followed by application of a PReLU activation function.

that the method identifies time-varying causal relationships, including time-varying time lags, when applied to synthetic datasets and simulated data from a public-domain fMRI simulator. We also show that it provides richer information about causal structures in a real-world task fMRI dataset than traditional non-time-varying causal modeling does.

2. Methods

2.1. Multivariate Granger Causality via Deep Stacking Networks

Given sources X_t and Z_t, and target Y_t, we train a DSN whose CNNs with ACKs (CNN-ACKs) use X_t, Y_t, and Z_t to reconstruct Y_t. Conditional Granger causality is assessed in terms of how much better X_t reconstructs Y_t, after accounting for how well previous time points of Y_t, along with Z_t, jointly reconstruct Y_t (Figure 2 (a)). First, CNN-ACK 1 and 2 are trained to transform previous time points of Y_t into Y_t, and Z_t into Y_t, resulting in estimates $\hat{Y}_{t,1}$ and $\hat{Y}_{t,2}$ and prediction errors $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$. Then, to represent the best reconstruction of Y_t based on both of Y_t and Z_t, $\hat{Y}_{t,1}$ and $\hat{Y}_{t,2}$ provide inputs to the third module, which estimates an element-wise weighted sum of the inputs to predict Y_t, resulting in estimate $\hat{Y}_{t,3}$ and prediction error $\varepsilon_{t,3}$. To reconstruct Y_t based on X_t, time series X_t is provided as input to CNN-ACK 3, again with Y_t as the target, resulting in predicted time series $\hat{Y}_{t,4}$ and prediction error $\varepsilon_{t,4}$. To reconstruct Y_t in terms of all of X_t, Y_t, and Z_t jointly, $\hat{Y}_{t,3}$ and $\hat{Y}_{t,4}$ are provided as inputs to an element-wise weighted sum to produce the final estimate of Y_t, $\hat{Y}_{t,5}$, and prediction error $\varepsilon_{t,5}$. The Granger causality of source X_t to target Y_t, conditioned on other source Z_t ($X \rightarrow Y|Z$), is defined in terms of the reduction in modeling error when X_t, Y_t, and Z_t are used to reconstruct Y_t, compared to when only Y_t and Z_t are used to reconstruct Y_t:

$$GCindex_{X \rightarrow Y|Z} = \ln(\frac{|\varepsilon_{t,3}|}{|\varepsilon_{t,5}|}) \tag{1}$$

If incorporating X_t improves the reconstruction of Y_t after accounting for effects of Y_t and Z_t, $GCindex_{X \to Z|Y}$ will be a large positive number, providing evidence for conditional Granger causality. Complex causal relationships among several time series can be disentangled by calculating conditional Granger causality with differing assignments of time series to the roles of X_t, Y_t, and Z_t.

2.2. CNN-ACKs for Time-varying Granger Causality

Inspired by Jia et al. (2016) and Zamora Esquivel et al. (2019), we used CNN-ACKs in our DSN architecture to estimate time-varying causal relationships. An ACK is defined by a dynamic filter that changes its weights automatically depending on the data in the input time series. The ACK K_t is generated by convolving filters Q with input time series X_t and using an activation function to transform the result into K_t. The first step is that at each timestep (t), the (1 x 6) hidden layer output $H_t = [h_{t,5}, \ldots, h_{t,0}]$ is calculated as the dot product of six 1 x 2 filters $Q = [q_{1,1}, q_{1,2}; \ldots; q_{6,1}, q_{6,2}]$ with the input time series $X_t = [x_{t-6}, \ldots, x_t]$ (Figure 2 (b) Bottom).

$$h_{t,5} = [q_{1,1}, q_{1,2}] \cdot [x_{t-6}, x_{t-5}]; ...; h_{t,0} = [q_{6,1}, q_{6,2}] \cdot [x_{t-1}, x_t] \tag{2}$$

Then, the Parametric Rectified Linear Unit (PReLU) activation function is applied to each element of hidden layer output H_t to generate the ACK K_t,

$$K_t = [k_{t,5}, ..., k_{t,0}] = PReLU_t(H_t) = [PReLU_{t,5}(h_{t,5}), ..., PReLU_{t,0}(h_{t,0})] \tag{3}$$

The coefficients of K_t can be interpreted as evidence of Granger causality at specific time lags at time t of the time series. For example, if $k_{t,5}$ (lag 5 causality) has a large magnitude, it suggests that at time t, there is a causal relationship between the source at time t-5 and the target at time t. The estimate of the target time series, $\hat{Y}_t$, is the dot product of K_t with the input time series $X_t = [x_{t-5}, \ldots, x_t]$ (Figure 2 (b) Top).

$$\hat{Y}_t = K_t^T X_t \tag{4}$$

In each CNN-ACK, the six filters Q (2 weights and 1 bias terms for each filter) and the parameters of PReLU (6 weights for each timestep) are the learnable parameters. We used the TensorFlow and Keras software packages to build our network architecture and optimize it with Adam optimizer (Chollet et al., 2015; Abadi et al., 2016).

2.3. Design of Experiments

We applied the proposed method to synthetic time series data, simulated task fMRI data from the public-domain STANCE simulator (Hill et al., 2017), and a real-world task fMRI dataset. For each synthetic and simulated dataset, 100 X_t, Y_t, Z_t time series triples were generated as described in subsequent sections, and the real-world task fMRI dataset included 100 fMRI scans. For each dataset, ten-fold cross validation was used to repeatedly train s and quantify causal relationships within the testing set of the fold. Conditional Granger causality between source and target, independent of source was considered evident when the mean GC index over the ten-folds of cross validation was significantly greater than 0 via a one-sample t-test (p-value <0.05). Identified causal relationships were compared to those programmed into the synthetic and simulated data sets, and causal relationships identified in the real-world data were compared to published data about the brain functional underpinnings of the task.

2.4. Synthetic Datasets

We designed two synthetic datasets that focused on testing the method's ability to model nonlinear causal relationships whose functional form differed between well-defined epochs of the time series, but whose time lag was constant; and testing the ability to model relationships whose functional form is constant, but whose time lags differ between epochs.

Synthetic dataset 1: identical time lags in all epochs. Each time series had 110 timesteps generated according to the equations in Table 1. N(0,0.1) represents Gaussian noise with zero mean and 0.1 standard deviation.

Synthetic dataset 2: different time lags in different epochs. Each time series had 110 timesteps generated according to the equations in Table 2.

Table 1: Synthetic dataset 1: identical time lags in all epochs.

	X-Y relationship	**Equation**	**Timestep (t)**
Epoch 1	Linear	$\mathrm{Y}_t = 5X_{t-1} + 0.5N(0,0.1)$	1-22
Epoch 2	Quadratic	$\mathrm{Y}_t = -0.5(X_{t-1})^2 + 0.5N(0,0.1)$	23-44
Epoch 3	Exponential	$\mathrm{Y}_t = 0.5e^{X_{t-1}} + 0.5N(0,0.1)$	45-66
Epoch 4	Cubic	$\mathrm{Y}_t = -5(X_{t-1})^3 + 0.5N(0,0.1)$	67-88
Epoch 5	No	$\mathrm{Y}_t = 0.5N(0,0.1)$	89-110

$\mathrm{X}_t \sim N(0,0.1)$; $\mathrm{Z}_t \sim N(0,0.1)$

Table 2: Synthetic dataset 2: different time lags in different epochs.

	X-Y relationship	**Equation**	**Timestep (t)**
Epoch 1	Time lag 2	$\mathrm{Y}_t = 0.5X_{t-2} + 0.5N(0,0.1)$	1-55
Epoch 2	Time lag 5	$\mathrm{Y}_t = -0.5X_{t-5} + 0.5N(0,0.1)$	56-110

$\mathrm{X}_t \sim N(0,0.1)$; $\mathrm{Z}_t \sim N(0,0.1)$

2.5. Simulated Task fMRI Datasets

For each simulated dataset, triples of 130-timestep time series, each of which contained single-timestep-duration events, were produced, with causal relationships existing between events in one time series, and events in another. Each time series of events was convolved with a canonical hemodynamic response function (HRF), followed by addition of simulated system and physiological noise at a magnitude of 1% of the event-related fMRI signal.

Simulated dataset 1: identical time lags in all epochs. Each X_t, Y_t, Z_t time series triple was initially generated with 52 randomly placed events. Then, time series Y_t, between time points 1 and 65, was edited so that an event at Y_t was added if there was also an event at X_{t-1}; i.e., X had an *excitatory* effect on Y during this epoch. Similarly, time series Y_t, between time points 66 and 130, was edited so that an event at Y_t was deleted if there were events at X_{t-1} and Y_t (Table 3); i.e., X had an *inhibitory* effect on Y.

Simulated dataset 2: different time lags in different epochs. All X_t, Y_t, Z_t time series triples were generated initially with 52 randomly placed events. Then, as for simulated dataset 1, events in the Y_t were edited to reflect the differing excitatory and inhibitory effects of X_t at differing time lags within each epoch (Table 4).

2.6. Real-World Task fMRI Dataset

We applied the proposed method to task fMRI data collected from the Bogalusa Heart Study (Berenson et al., 2001). One hundred participants performed a Stroop task during fMRI on a GE Discovery 3T scanner at Pennington Biomedical Research Center. Acquisition of T1-weighted structural MPRAGE and axial 2D gradient echo EPI BOLD fMRI data was described previously (Carmichael et al., 2019). Preprocessing of fMRI included slice timing correction, head motion correction, smoothing, co-registration to the T1-weighted image, and warping of T1-weighted data to a Montreal Neurological Institute (MNI) coordinate. The regions of interest (ROI) in MNI coordinate previously identified as activated by the Stroop task (fusiform gyrus, occipital gyrus, precuneus, and thalamus) were extracted (Sheu et al., 2012). The proposed method was applied to all possible assignments of ROIs to the roles of source, target, and other source (i.e., to X_t, Y_t, and Z_t) to explore time-varying Granger causalities. A specific time lag is selected when its time lag causality coefficients were significantly different from 0 via a one-sample t-test. Differences in causal relationships between rest and task conditions were assessed by testing whether the difference in Granger causality coefficients between conditions was significantly different from 0 by permutation testing with 10,000 permutations.

Table 3: Simulated dataset 1: identical time lags in all epochs.

	X-Y relationship	**Time lag**	**Timestep (t)**
Epoch 1	Excitation	1	1-65
Epoch 2	Inhibition	1	66-130

Table 4: Simulated dataset 2: different time lags in different epochs.

	X-Y relationship	**Time lag**	**Timestep (t)**
Epoch 1	Excitation	1	1-26
Epoch 2	Inhibition	1	27-52
Epoch 3	Excitation	3	53-78
Epoch 4	Inhibition	4	79-104
Epoch 5	No relationship	-	105-130

3. Results

3.1. Synthetic Datasets

Synthetic dataset 1: identical time lags in all epochs. The proposed method correctly identified the true Granger causality $X \to Y|Z$ in synthetic dataset 1 (p-value <0.0001). The other possible assignments of X_t, Y_t, and Z_t to sources and target had no evidence of Granger causality (minimum p-value = 0.7437). In addition, the time dependence of the $X \to Y|Z$ causal relationships was correctly tracked by the Granger causality coefficients (the coefficients of ACK K_t) that quantified each time lag (Figure 3 & Figure 4 (a)). Specifically, Granger causality coefficients corresponding to ground-truth causal relationships (e.g., the lag 1 causality coefficient during epoch 1) were nonzero while other Granger causality coefficients were correctly estimated to be close to the nominal null value. Both GVAR and PF successfully identified the linear causal relationship in the linear epoch as expected; but they failed to identify other causal relationships in other epochs, likely due to the linear assumptions they make. (Figure 4 (a)).

Synthetic dataset 2: different time lags in different epochs. The proposed method correctly identified the true Granger causality $X \to Y|Z$ (p-value <0.0001). As above, Granger causality coefficients correctly tracked changes in time lags across epochs (see, e.g., the low errors in the lag 2 and lag 5 causality coefficients during the epochs they exerted causal effects, Figure 4(b)), and other coefficients were correctly estimated to be null. GVAR and PF failed to correctly track time-varying causalitie.

3.2. Simulated Task fMRI Datasets

The proposed method correctly identified the true Granger causality $X \to Y|Z$ in simulated dataset 1 and 2 (p-value = 0.0011 and 0.0015, respectively). Also, the time-

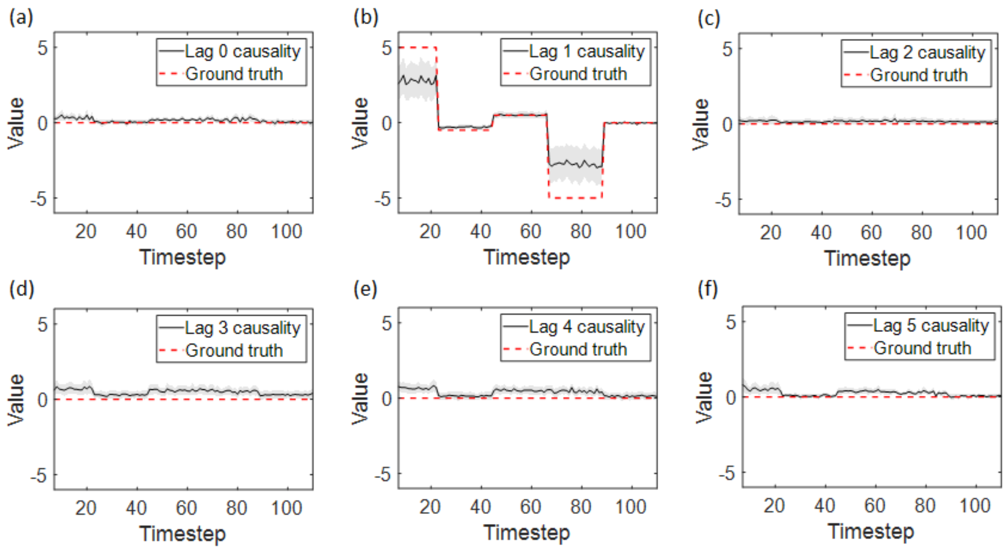

Figure 3: Time courses of ground-truth Granger causality coefficients (red) and estimates (K_t) from DSN-ACKs (mean in black, values within 1 standard deviation of the mean in gray) for differing time lags in synthetic dataset 1.

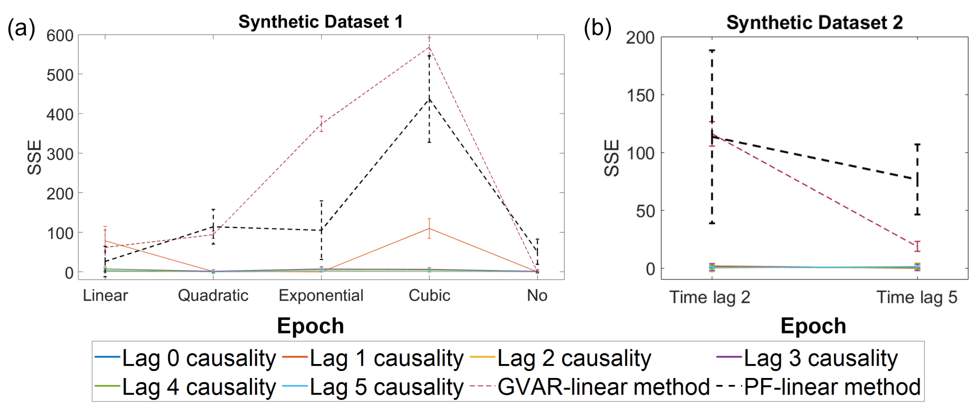

Figure 4: Sum of squared errors (SSE) differences between ground truth Granger causality coefficients and estimates of time lag 0-5 Granger causality coefficients provided by DSN-ACKs, GVAR, and PF for **(a)** synthetic dataset 1 and **(b)** synthetic dataset 2.

varying Granger causality coefficients for $X \rightarrow Y|Z$ were correctly estimated across all epochs by DSN-ACKs (Figure 5). All the other causality relationships among X_t, Y_t, and Z_t were correctly determined to be null (minimum p-value $= 0.1382$ and 0.2229 for simulated dataset 1 and 2, respectively). Both GVAR and PF failed to estimate Granger causality coefficients accurately in both simulated datasets.

3.3. Real-World Task fMRI Dataset

Figure 6 shows which lag 0 Granger causality coefficients were estimated to be non-zero among the four Stroop task ROIs during rest and task execution. Reciprocal causal relationships between the occipital gyrus and thalamus were identified independent of the fusiform gyrus, in agreement with previous findings (Guido, 2018; Usrey and Alitto, 2015). In addition, a causal relationship from the fusiform gyrus to occipital gyrus, independent of precuneus, has been identified. Reciprocal Granger causalities between the occipital

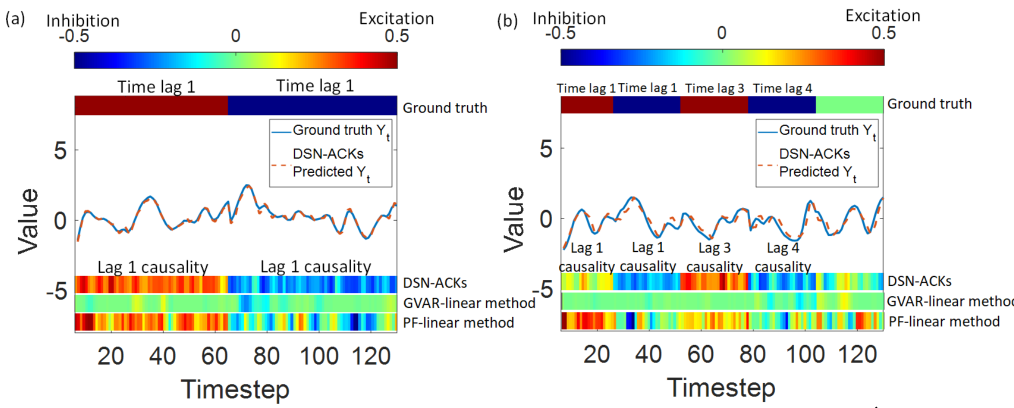

Figure 5: Time courses of ground-truth time series Y_t and the predicted $\hat{Y}_t$ by DSN-ACKs for simulated dataset 1 **(a)** and 2 **(b)**. Ground truth Granger causality coefficients at time lags relevant to each time series epoch are shown above the time series; estimates of those Granger causality coefficients generated by DSN-ACKs, GVAR, and PF are depicted below.

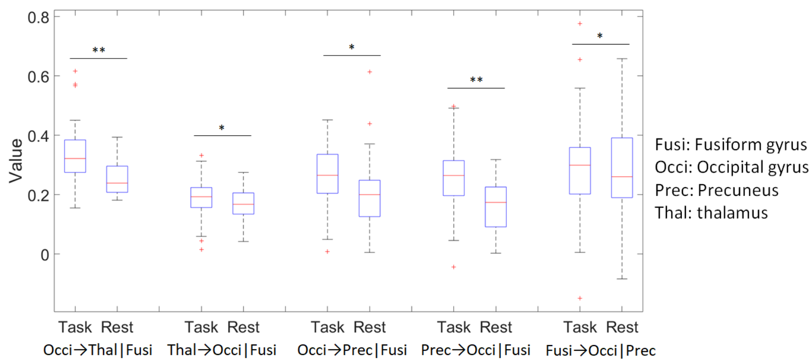

Figure 6: The estimated time lag 0 Granger causality coefficients ($k_{t,0}$) between various triples of brain regions from the real-world fMRI dataset during Stroop task performance and rest. Permutation test *p <0.05, **p <0.001.

gyrus and precuneus independent of fusiform gyrus have also been identified. No causal relationship has been identified by GVAR and PF. These lag 0 Granger causality coefficients were statistically different between task and rest conditions. This suggests differences in functional causal dynamics corresponding to differences in behaviors being performed by those brain regions. The latter three causal relationships were not successfully identified by our previous method that failed to account for time-varying causality (Chuang et al., 2021), suggesting that the current method captures richer information about complex brain network functioning than prior methods.

4. Conclusion

Our DSN-ACKs architecture that characterizes time-varying nonlinear conditional Granger causality identifies time-varying causal relationships programmed into synthetic and simulated fMRI data. When applied to real task fMRI data, the method identifies plausible causal brain functional relationships among brain regions that prior methods were unable to identify. Future work should extend this approach to account for spatially- and temporally variable hemodynamic response functions that could impact discovery of causal relationships (Ambrosi et al., 2021; Duggento et al., 2021).

Acknowledgments

Funding for this work was provided by NIH grants R01AG041200 and R01AG062309 as well as the Pennington Biomedical Research Foundation.

References

Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

Pierfrancesco Ambrosi, Mauro Costagli, Ercan E Kuruoglu, Laura Biagi, Guido Buonincontri, and Michela Tosetti. Modeling brain connectivity dynamics in functional magnetic resonance imaging via particle filtering. *bioRxiv*, 2021.

Gerald S Berenson et al. Bogalusa Heart Study: a long-term community study of a rural biracial (black/white) population. *The American journal of the medical sciences*, 322(5): 267–274, 2001.

Owen Carmichael, Patrick Stuchlik, Sreekrishna Pillai, Geert-Jan Biessels, Ram Dhullipudi, Anna Madden-Rusnak, Shane Martin, Daniel S Hsia, Vivian Fonseca, and Lydia Bazzano. High-normal adolescent fasting plasma glucose is associated with poorer midlife brain health: Bogalusa Heart Study. *The Journal of Clinical Endocrinology & Metabolism*, 104 (10):4492–4500, 2019.

Sezen Cekic, Didier Grandjean, and Olivier Renaud. Time, frequency, and time-varying Granger-causality measures in neuroscience. *Statistics in medicine*, 37(11):1910–1931, 2018.

François Chollet et al. Keras. https://keras.io, 2015.

Kai-Cheng Chuang, Sreekrishna Ramakrishnapillai, Lydia Bazzano, and Owen T Carmichael. Deep stacking networks for conditional nonlinear Granger causal modeling of fMRI data. In *International Workshop on Machine Learning in Clinical Neuroimaging*, pages 113–124. Springer, 2021.

Gopikrishna Deshpande, Stephan LaConte, George Andrew James, Scott Peltier, and Xiaoping Hu. Multivariate Granger causality analysis of fMRI data. *Human brain mapping*, 30(4):1361–1373, 2009.

Andrea Duggento, Maria Guerrisi, and Nicola Toschi. Echo state network models for nonlinear Granger causality. *Philosophical Transactions of the Royal Society A*, 379(2212): 20200256, 2021.

Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1): 13–36, 2011.

Karl J Friston, Rosalyn Moran, and Anil K Seth. Analysing connectivity with Granger causality and dynamic causal modelling. *Current opinion in neurobiology*, 23(2):172–178, 2013.

Rainer Goebel, Alard Roebroeck, Dae-Shik Kim, and Elia Formisano. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic resonance imaging*, 21(10):1251–1261, 2003.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

William Guido. Development, form, and function of the mouse visual thalamus. *Journal of neurophysiology*, 120(1):211–225, 2018.

Jason E Hill, Xiangyu Liu, Brian Nutter, and Sunanda Mitra. A task-related and resting state realistic fMRI simulator for fMRI data validation. In *Medical Imaging 2017: Image Processing*, volume 10133, page 101332N. International Society for Optics and Photonics, 2017.

Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. Dynamic filter networks. *Advances in neural information processing systems*, 29:667–675, 2016.

Wei Liao, Daniele Marinazzo, Zhengyong Pan, Qiyong Gong, and Huafu Chen. Kernel Granger causality mapping effective connectivity on fMRI data. *IEEE transactions on medical imaging*, 28(11):1825–1835, 2009.

Ričards Marcinkevičs and Julia E Vogt. Interpretable models for Granger causality using self-explaining neural networks. *arXiv preprint arXiv:2101.07600*, 2021.

Daniele Marinazzo, Wei Liao, Huafu Chen, and Sebastiano Stramaglia. Nonlinear connectivity by Granger causality. *Neuroimage*, 58(2):330–338, 2011.

José C Príncipe, Weifeng Liu, and Simon Haykin. *Kernel adaptive filtering: a comprehensive introduction*. John Wiley & Sons, 2011.

João Ricardo Sato, Edson Amaro Junior, Daniel Yasumasa Takahashi, Marcelo de Maria Felix, Michael John Brammer, and Pedro Alberto Morettin. A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. *Neuroimage*, 31(1):187–196, 2006.

Johan Schoukens and Lennart Ljung. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, 39(6):28–99, 2019.

Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.

Lei K Sheu, J Richard Jennings, and Peter J Gianaros. Test–retest reliability of an fMRI paradigm for studies of cardiovascular reactivity. *Psychophysiology*, 49(7):873–884, 2012.

W Martin Usrey and Henry J Alitto. Visual functions of the thalamus. *Annual review of vision science*, 1:351–371, 2015.

Julio Zamora Esquivel, Adan Cruz Vargas, Paulo Lopez Meyer, and Omesh Tickoo. Adaptive convolutional kernels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

Zhenyu Zhou, Xunheng Wang, Nelson J Klahr, Wei Liu, Diana Arias, Hongzhi Liu, Karen M von Deneen, Ying Wen, Zuhong Lu, Dongrong Xu, et al. A conditional Granger causality model approach for group analysis in functional magnetic resonance imaging. *Magnetic resonance imaging*, 29(3):418–433, 2011.