Granger Causality Pattern Learning Equipped with Noise Invalidation Soft Thresholding

Khashayar Bayati, Karthikeyan Umapathy, Soosan Beheshti
Department of Electrical, Computer, and Biomedical Engineering,Toronto Metropolitan University, Toronto, Canada
Email: {khashayar.bayati, kumapath, soosan}@torontomu.ca

***Abstract*—This paper concentrates on pattern learning of Granger causality. In this context, the entities of the Granger causality matrix estimation derived from the state-space model indicate directional dependencies of the observations. Existing methods propose different forms of thresholding to eliminate insignificant entities from this matrix. These approaches do not exploit the difference in statistical characteristics of insignificant entities compared to the Granger causality itself. In this work, Noise Invalidation Soft Thresholding is chosen as the thresholding method to discard null relations in the Granger causality matrix pattern learning. Unlike existing approaches, the proposed method benefits from the statistical properties of insignificant entities. The simulation results demonstrate the advantages and superiority of the proposed method in the sense of accuracy and robustness for both randomly generated datasets with causal footprints as well as simulated electroencephalogram datasets.**

***Index Terms*—Granger Causality, State-Space Modelling, Noise Invalidation Soft Thresholding**

I. Introduction

Causality is a fundamental concept that lies at the heart of many fields, from philosophy to science to social sciences. The notion of causality is central to understanding the workings of the world around us, from the physical laws that govern the universe to the social and economic forces that shape our societies [1]. Causality has two distinct categories that continue to captivate researchers across various domains. The first category deals with the notion of time preceding, where events in the past influence those in the future [2]. The second category pertains to the influence of treatments or interventions on an outcome of interest [3], which is out of the scope of this paper. The majority of research examining causal footprints using the Granger technique in the time and frequency domain focuses on data modeling using vector auto-regressive (VAR) models [4]. It has been shown in [5] that, for a VAR model, in addition to regression coefficients, the auto-covariance sequence and the cross-power spectral density of the underlying process can also be used to measure Granger causality. Although the use of VAR modeling has advantages, such as the simplicity of the linear structure, VAR models are not accurate in the presence of moving average (MA) components [6]. There is a class of multivariate auto-regressive moving average (ARMA) models or equivalent finite-order state-space models [7] which performs accurately in the presence of MA components [8]. Taking into account the availability of efficient and effective state-space system procedures [9], a novel method for estimating Granger causality using state-space models was developed and has since become the primary method for Granger causal analysis [6]. However, one major issue here is that applying a significance test to the Granger causality matrix is still an issue since its theoretical distribution is unknown [10]. The majority of analyses often struggle with either discovering causal links that do not exist or not finding existing causal links [11]. Two approaches for learning Granger causality have been proposed in [10], but those are not considered data structures. Inspired by this existing problem, we propose a method that fully exploits the structure characteristics by transforming the problem of learning into determining a suitable threshold. This technique, known as noise invalidation soft thresholding (NIST), as presented in [12], is used to classify $causal$ and $null$ entities of the Granger causality matrix. The paper is organized as follows. Section II briefly reviews Granger causality using both VAR and state-space models. Section III deploys a noise signature from noisy Granger causality sample means and presents the proposed estimation of Granger causality using NIST. Simulation results are provided in Section IV, and conclusions are drawn in Section V.

II. Granger Causality

Defining "causality" is a challenging problem in complex systems without intuitive cause-and-effect relationships. Norbert Wiener introduced the concept of a causal relationship between two time series in 1956. Later in 1969, Clive Granger formalized this concept using linear VAR models of stochastic processes to implement this idea of causality [13]. Consider a discrete-time, stationary vector stochastic observable process consisting of r real-valued zero-mean scalar processes, $\mathbf{y}_n = [y_{1,n} \;\; y_{2,n} \;\; ... \;\; y_{r,n}]^T, -\infty < n < \infty$. Assume that the observable process is partitioned into sub-processes $\mathbf{y}_n = [\mathbf{y}_{i,n}^T \; \mathbf{y}_{j,n}^T \; \mathbf{y}_{k,n}^T]$, which are respectively the driver ($\mathbf{y}_{i,n}$), the target ($\mathbf{y}_{j,n}$), and remaining processes ($\mathbf{y}_{k,n}$). Denoting the infinite past of a process by $\mathbf{y}_N^- = [\mathbf{y}_{N-1}^T \; \mathbf{y}_{N-2}^T \; ...]$, conditional Granger causality in the context of vector auto-regressive models from driver to target ($\mathbf{y}_{i,n} \rightarrow \mathbf{y}_{j,n}|\mathbf{y}_{k,n}$) measures the degree to which the driver's past ($\mathbf{y}_{i,n}^-$) enhances the prediction of the target's future beyond the degree to which infinite past of target conditional on remaining processes ($\mathbf{y}_{j,n}^-|\mathbf{y}_{k,n}^-$) already anticipates its own future [14]. This definition is evaluated in the time domain by finding regression errors in (1) and (2), where $\mathbb{E}$ is the expectation operator.

$$\mathbf{e}_{j|ijk,n} = \mathbf{y}_{j,n} - \mathbb{E}[\mathbf{y}_{j,n}|\mathbf{y}_n^-] \tag{1}$$

$$\mathbf{e}_{j|jk,n} = \mathbf{y}_{j,n} - \mathbb{E}[\mathbf{y}_{j,n}|\mathbf{y}_{j,n}^-, \mathbf{y}_{k,n}^-] \tag{2}$$

In the context of maximum likelihood (ML), the relevant measure is the log-likelihood ratio [6]. This measure applies to nested autoregressive (AR) models that are associated with the projections $\mathbb{E}[\mathbf{y}_{j,n}|\mathbf{y}_n^-]$ and $\mathbb{E}[\mathbf{y}_{j,n}|\mathbf{y}_{j,n}^-, \mathbf{y}_{k,n}^-]$. Granger Causality is then the log-likelihood ratio of error variances in (1) and (2) as [15]:

$$F_{i\to j} = \ln \frac{\mathbb{E}[\mathbf{e}^2_{j|jk,n}]}{\mathbb{E}[\mathbf{e}^2_{j|ijk,n}]} \tag{3}$$

The above procedure can be repeated for $i, j = 1, 2, ..., r$. By calculating $F_{i\to j}$ for every possible (i, j) pair and constructing it as a matrix without taking the diagonals into account, the Granger causality matrix can be inferred from observations. A notable finding from [6] is that whenever (i, j) entry is zero, then the jth variable does not Granger cause the ith variable. Many types of time series data, especially those used in econometrics and neuroscience, have a significant MA component. Moreover, filtering, down-sampling, and recording noises all contribute to the presence of MA components, resulting in an ARMA process [16]. These MA components may not be well represented by a finite-order AR model. Thus, for a finite number of samples, a finite, possibly large model order must be chosen, which results in weak Granger causality with reduced statistical power and increased bias [6]. Contrary to linear AR models, the class of multivariate ARMA models or finite-order linear state-space (SS) models [7] are closed under the aforementioned procedures [16]. Given the general form of the linear state-space model, in terms of an innovation variable $\boldsymbol{\varepsilon}_n = \mathbf{y}_n - \mathbb{E}[\mathbf{y}_n|\mathbf{y}_n^-]$, a new state-space model can be achieved associated with Kalman filtering with Kalman gain of $\mathbf{K}$ in (4), which is called the "Innovation form of state-space (ISS)."

$$\begin{cases} \mathbf{z}_{n+1} = \mathbf{A}\mathbf{z}_n + \mathbf{K}\boldsymbol{\varepsilon}_n & \text{(4a)} \\ \mathbf{y}_n = \mathbf{C}\mathbf{z}_n + \boldsymbol{\varepsilon}_n & \text{(4b)} \end{cases}$$

The state equation (4a) in a dynamic system is represented by the transition matrix $\mathbf{A} \in \mathbb{R}^{l\times l}$, which describes how the system's current state $\mathbf{z}_n \in \mathbb{R}^l$ relates to its future state. On the other hand, the output equation (4b) shows the relationship between the system state $\mathbf{z}_n$ and the observable process $\mathbf{y}_n \in \mathbb{R}^r$, and it is represented by the observation matrix $\mathbf{C} \in \mathbb{R}^{r\times l}$. States and observations are assumed to be weakly stationary, which necessitates the stability of matrix $\mathbf{A}$. From [17], $\boldsymbol{\varepsilon}_n$ comprises a white noise process with a positive-definite covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^T]$. It is essential to solving a problem known as the discrete algebraic Ricatti equation (DARE), expressed in terms of the state error variance matrix, to get the ISS parameters [18]. The error variance in (1) is equivalent to the j-th diagonal element of the innovation covariance, $\mathbb{E}[\mathbf{e}^2_{j|ijk,n}] = \boldsymbol{\Sigma}(j, j)$. Forming a sub-model that eliminates the driving process yields (5), where the superscript (R) denotes a model that has been reduced by removing rows that correspond to the driver.

$$\mathbf{y}_n^R = \mathbf{C}^R\mathbf{z}_n + \boldsymbol{\varepsilon}_n^R \tag{5}$$

The sub-model (4a, 5) forms another innovation form of state-space model for which the error variance in (2) is equivalent to the j-th diagonal element of the innovation covariance, $\mathbb{E}[\mathbf{e}^2_{j|jk,n}] = \boldsymbol{\Sigma}^R(j, j)$. This proves that using the ISS parameters of an observed process $\mathbf{y}_n$; Granger Causality in (3) can be calculated numerically as [6]:

$$F_{i\to j} = \ln \frac{\boldsymbol{\Sigma}^R(j, j)}{\boldsymbol{\Sigma}(j, j)} \tag{6}$$

A. Estimation of Granger causality

In the case of using the VAR model to measure Granger causality by using regression parameters, the statistical test to learn a Granger Causality pattern becomes the log-likelihood ratio test [19]. When the VAR model is represented by an auto-covariance sequence of the underlying process, it is suggested that the statistics are well-approximated by a Γ distribution [5]. According to [6], for the case that there is no theoretical asymptotic distribution for Granger causality inferred from state-space models significant testing should be conducted using permutation or bootstrapping approaches. A method to learn Granger causality resulting from state space parameters has been put forth in [10] by creating a set of Granger causality matrix sample means ($\mathbf{F}_{mean}$) to remove insignificant entries in the Granger causality matrix. By constructing a vectorized version of $\mathbf{F}_{mean}$ ($\mathbf{f}_{vec}$), two methods for Granger Causality Pattern Learning (GCPL) are proposed in [10] to remove insignificant entities of $\mathbf{F}_{mean}$ by fitting a Gaussian Mixture Model (GMM) to $\mathbf{f}_{vec}$ as follows:

1) GCPL with Clustering (GCPL-C): The entries clustered by the posterior probabilities into the first Gaussian component with parameters (μ_1, σ_1^2) are considered $null$. Thus, $causal$ entities are those grouped into the other components.
2) GCPL with Thresholding (GCPL-T): The intersection of two first component distributions (f_1 and f_2) can be considered as a threshold (T_{PL-T}), defined to classify entities in $\mathbf{F}_{mean}$ as $causal$ or $null$. Where:

$$log f_1(T_{PL-T}; \mu_1, \sigma_1^2) = log f_2(T_{PL-T}; \mu_2, \sigma_2^2) \tag{7}$$

The block diagram of these methods is demonstrated in Figure 1.

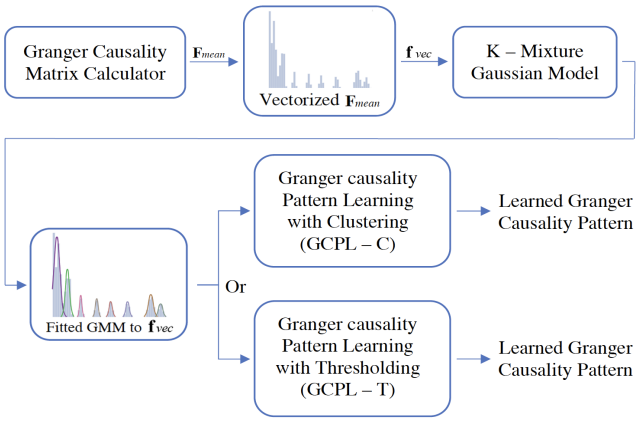

Fig. 1. The proposed scheme in [10] for learning Granger causality matrix

B. Motivation

It seems rational to consider what is denoted by insignificant entities as irrelevant nulls, which behave like unwanted noise. From [10], GCPL-C and GCPL-T, respectively, assume the values under the first component and the values less than the intersection of the two first components of fitted GMM to $\mathbf{f}_{vec}$ in Figure 1 as insignificant entities without considering the statistical characteristics of noise. Besides, both these methods require a sufficiently large number of trials to be performed, which may not be appropriate in some cases. Ignoring these assumptions can produce inaccurate results, especially in the presence of high-order and strong MA components. Therefore, a more robust approach toward learning Granger causality is to explore the statistical characteristics of Granger causality estimations. To this aim, in this research, NIST in [12] has been chosen to improve the Granger Causality learning performance.

III. Granger causality pattern learning with NIST (GCPL-NIST)

The objective of denoising is to eliminate noise in the observed data as much as possible. Among many approaches, thresholding seeks to remove the small values that are often associated with noise [12]. Consider noiseless $\overline{\mathbf{f}_{vec}}$ contaminated by an additive noise w to construct $\mathbf{f}_{vec}$, where $\boldsymbol{w} = [\boldsymbol{w}[1], \boldsymbol{w}[2], ..., \boldsymbol{w}[N]]^T$ is a white Gaussian random process with zero mean and variance σ_w^2. Thus, $\mathbf{f}_{vec}$ can be expressed as follows:

$$\mathbf{f}_{vec}[n] = \overline{\mathbf{f}_{vec}}[n] + \boldsymbol{w}[n] \tag{8}$$

A new variable γ_m is defined as follows once noisy $\mathbf{f}_{vec}$ has been sorted based on its absolute value ($\mathbf{f}_{sort}$):

$$\gamma_m = \sum_{n=m+1}^{N} |\mathbf{f}_{sort}[n]|^2 = \sum_{n=m+1}^{N} |\mathbf{f}^*_{sort}[n] + \boldsymbol{v}[n]|^2 \tag{9}$$

where m is the number of thresholding-induced non-zero coefficients, and $\mathbf{f}^*_{sort}[n]$ and $\boldsymbol{v}[n]$ are the corresponding noiseless and noise-only parts of $\mathbf{f}_{sort}[n]$. If the thresholding procedure with a soft threshold T_{NIST} discards values that are just corresponding to noise, then γ_m is represented by γ^v_m. Where γ^v_m is a sample of random variable Γ^v_m. From (9).
Whenever γ_m is a member of the set of γ^v_ms, it can be inferred that there are no noiseless values with the probability of $erf(\frac{\lambda}{\sqrt{2}}) = \frac{1}{\sqrt{\pi}} \int_0^{\frac{\lambda}{\sqrt{2}}} e^{-t^2} dt$. Otherwise, there is a similar high likelihood that γ_m includes nonzero rejected coefficients. λ must be chosen so that the probability is close to one. The following criteria can be used to evaluate this condition [12]:

$$\beta[m] = \frac{|\gamma_m - \mathbb{E}[\Gamma^v_m]|}{\lambda \sqrt{var[\Gamma^v_m]}} \tag{10}$$

If $\beta[m]$ is less than one, then $\gamma_m = \gamma^v_m$, is very likely to be true with the probability of $erf(\frac{\lambda}{\sqrt{2}})$, and there are no non-zero values involved. But once $\beta[m]$ gets bigger than one, it is likely with the same probability that γ_m cannot have the same structure as the additive noise. Consequently, the desired m^* can be determined by considering the criteria in (11). The threshold of this method (T_{NIST}), which is first introduced in [12], can be expressed as (12).

$$m^* = \underset{m}{\operatorname{argmin}}(\beta[m] \geq 1) \tag{11}$$

$$T_{NIST} = |\mathbf{f}_{sort}[m^*]| \tag{12}$$

Once T_{NIST} is determined using the NIST method, the elements of the $\mathbf{F}_{mean}$ that have a greater value of the threshold represent a valid Granger causality ($causal$), and values smaller than the threshold do not indicate any Granger causality relationship between corresponding source and target variables ($null$). The block diagram of proposed methods for Granger causality learning using NIST has been shown in Figure 2.

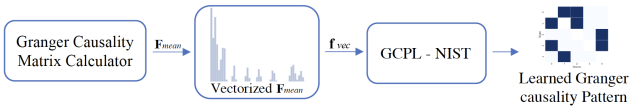

Fig. 2. Proposed scheme for learning Granger causality matrix using NIST

IV. Simulation Results

A. Simulated multivariate ARMA Model Dataset

Consider a vector auto-regressive moving average (VARMA) model of order (p, q):

$$\mathbf{y}_n = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{y}_{n-k} + \boldsymbol{\eta}_n + \sum_{j=1}^{q} \mathbf{C}_k \boldsymbol{\eta}_{n-j} \tag{13}$$

where $\mathbf{A}_k$'s represent autoregressive coefficients, and $\mathbf{C}_k$'s are moving average coefficients. VAR polynomial of order p has to be stable by checking if all eigenvalues of the system lie inside the unit circle. Moreover, the MA polynomial of order q is considered to be minimum-phase. To assess the proposed method, simulated time series were generated by forming sparse $\mathbf{A}_k$ as a random array with dimensions $p \times n \times n$ and scaling it to make a stable model. After the VARMA model has been generated, It has been shown in [20] that the Granger causality pattern of the VARMA model can be controlled by the sparsity pattern of $\mathbf{A}_k$ matrices [10].

Let N_0 represent the total number of trials in the simulated time series. After performing subspace identification on each trial of the process $\mathbf{y}_n$ to find parameters in (6), the Granger causality matrix ($\mathbf{F}$) for each trial can be computed by solving DARE. Then, by averaging over N_1 samples out of N_0 samples of $\mathbf{F}$, $N_2 = N_0/N_1$ samples of $\mathbf{F}_{mean}$ will be produced as it is shown in Figure 3 (a).

Reference [10] assumed that the Central Limit Theorem applies, which means that if N_1 is sufficiently large, the entries in $\mathbf{F}_{mean}$ will converge to a Gaussian distribution. By creating a vectorized version of $\mathbf{F}_{mean}$, denoted by $\mathbf{f}_{vec}$, and pooling all its entries, a histogram of $\mathbf{f}_{vec}$ will exhibit a mixture of Gaussian components, as shown in Figure 4. However, it is important to note that when using the Central Limit Theorem to analyze the distribution of entries in $\mathbf{F}_{mean}$, one must consider the assumptions and limitations of the analysis. If the

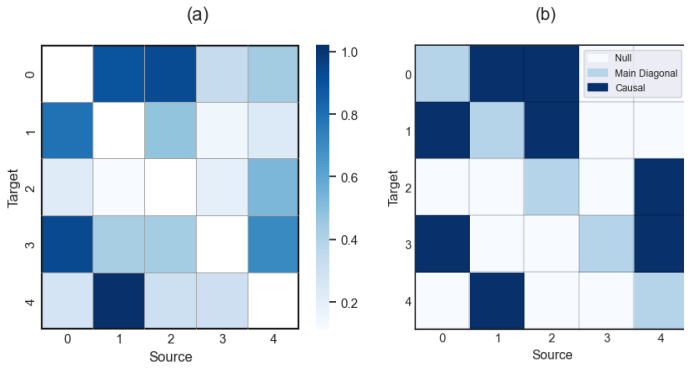

Fig. 3. a) An example of $\mathbf{F}_{mean}$, an averaged matrix over N_1 samples out of N_0 Samples of $\mathbf{F}$, b) Learned Granger causality pattern

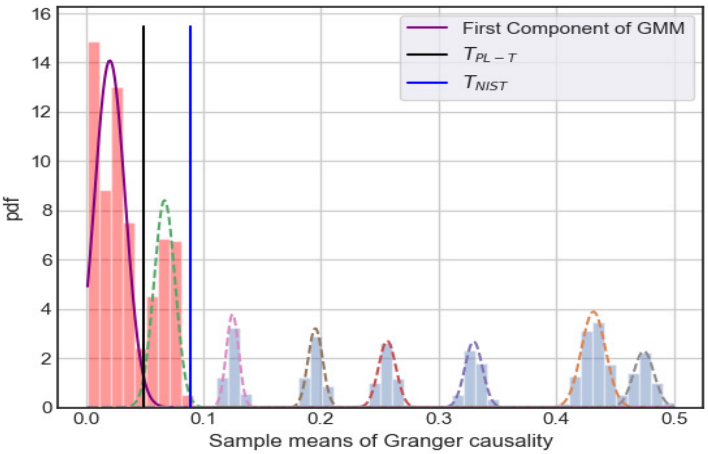

Fig. 4. Fitted Gaussian Mixture Model to a Vectorized $\bar{F}$

noise in the process is stationary and Gaussian, and the moving average components are either known or negligible, the Central Limit Theorem may be applicable, as illustrated in Figure 4. However, in many real-world applications, the noise is non-stationary or non-Gaussian, and the moving average components may be significant and unknown, making it difficult to apply the Central Limit Theorem with the limited number of observations. In such cases, alternative methods that consider the underlying structure of the noise may be required to model its behavior and distribution. For example, if the noise is colored, a more suitable approach may be to use techniques that explicitly account for the statistical characteristics of the noise. The GCPL-NIST approach proposes a new method that accounts for the underlying structure of the noise, resulting in a more precise and robust model that can better capture the complex dynamics of the underlying process.

For this section of the simulation results, $N_0 = 20000$ trials containing 5 observation ($n = 5$) with the length of $T = 1000$ were generated as the output of a ground-truth model. By finding the Granger causality matrix for each trial and averaging over $N_1 = 20$ of them, $N_2 = 1000$ samples of $\mathbf{F}_{mean}$ were produced. Then three methods, GCPL-T, GCPL-C, and GCPL-NIST, were applied to form the Granger causality pattern matrix. In Figure 4, the results of denoising of $\mathbf{f}_{vec}$ using the NIST method can be found. As demonstrated in Figure 4, the threshold obtained by the GCPL-NIST method (T_{NIST}) is quite different from the threshold obtained by GCPL-T in [10]. Also, as shown in red bars, not only are insignificant values of $\mathbf{F}_{mean}$ detected by GCPL-NIST not limited to the first component of the GMM model, as suggested by the GCPL-C method, but they also include most values of the second component. Thus, it seems that using denoising methods to remove insignificant values in $\mathbf{F}_{mean}$ is a more efficient and accurate choice for learning the Granger Causality pattern since GCPL-NIST takes advantage of statistical characteristics of insignificant entities. As explained in Section III The matrix of the Granger pattern will be formed by using T_{NIST} as it is illustrated in Figure 3 (b) for the $\mathbf{F}_{mean}$ matrix in Figure 3 (a).

In order to verify the accuracy of the proposed method, ten distinct random VAR models were generated. For each of them, N_2 samples of $\mathbf{F}_{mean}$ were split into train and test sets using 10-fold cross-validation to test the performance of methods for unseen data. In the training set, required parameters were found to classify entities in test data as $causal$ or $null$ and compared with the ground-truth model. The measure of accuracy is the ratio of the total number of true $causal$ and true $null$ entities to the number of non-diagonal elements of the $\mathbf{F}_{mean}$. In any analysis, when a method is repeated several times, a mean performance value is calculated alongside the standard deviation of the results to evaluate the robustness of the method. The accuracy for two methods in [10], and the proposed method in this paper (GCPL-NIST) are presented in Table I. As this table shows, NIST has higher accuracy, and the smaller standard deviation indicates that NIST offers more robustness.

TABLE I
ACCURACY AND STD PERFORMANCE COMPARISON OF GCPL-NIST, GCPL-C, AND GCPL-T

	Method		
	GCPL-NIST	**GCPL-C**	**GCPL-T**
Accuracy	90.5	83.5	78.5
Standard Deviation	1.49	2.29	3.21

B. Simulated Electroencephalogram Dataset

To demonstrate the effectiveness of the proposed Granger causality method on real-world electroencephalogram (EEG) data, simulations were conducted using an EEG dataset. Simulated EEG signals were generated from ten electrodes across a range of SNR levels and for a variety of different scenarios, including both unidirectional and bidirectional causal relationships. The SEED-G toolbox [21] was utilized to generate the simulated EEG signals, which included realistic noise and artifact components. These simulated datasets were then analyzed using the proposed Granger causality learning method. The resulting causal relationships were compared to known ground truth causal relationships obtained from the simulation process. All measures used in this comparison were

similar to those explained in Section IV-A. As shown in Table II, the proposed method (GCPL-NIST) performed better and achieved higher accuracy than the other two methods in [10], and its smaller standard deviation for different values of SNR, even when the SNR was relatively low, indicates that GCPL-NIST offers more robustness. These findings suggest that the proposed method can accurately identify causal relationships in EEG datasets, and its superior performance makes it a promising method for EEG signal analysis.

TABLE II
ACCURACY AND STD OF PERFORMANCE COMPARISON OF GCPL-NIST, GCPL-C, AND GCPL-T FOR EEG DATASET ACROSS DIFFERENT VALUES OF SNR

	Method		
SNR	**GCPL-NIST**	**GCPL-C**	**GCPL-T**
5 dB	90.55 $\pm$ 1.87	84.89 $\pm$ 2.86	78.67 $\pm$ 3.47
10 dB	91.89 $\pm$ 1.57	86.11 $\pm$ 2.54	80.99 $\pm$ 3.23
15 dB	93.1 $\pm$ 1.47	87.89 $\pm$ 2.35	83.67 $\pm$ 3.94
20 dB	94.88 $\pm$ 1.24	89.78 $\pm$ 2.09	86.22 $\pm$ 3.65

V. CONCLUSION

This paper addresses the Granger causality learning problem for a time series by employing state-space models. Due to the lack of knowing theoretical asymptotic distribution, applying a significance test to the Granger causality matrix remains problematic. To address this problem, NIST, as a method that can explore the statistical structure of insignificant entities in the Granger causality matrix, has provided a robust approach to improve the performance of Granger causality learning. The simulation results for both the randomly generated VARMA model and the simulated EEG dataset show the beneficial advantages of this method compared to previously introduced methods in the sense of accuracy and robustness. It is worthwhile to mention that this method is capable of being applied for learning causality in other biomedical time series, such as electrocardiogram or functional magnetic resonance imaging, to determine the connectivity between different sources' activities.

REFERENCES

[1] J. Pearl, *Causality*. Cambridge university press, 2009.

[2] K. Friston, "Dynamic causal modeling and granger causality comments on: The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution," *Neuroimage*, vol. 58, no. 2-2, p. 303, 2011.

[3] J. Pearl *et al.*, "Models, reasoning and inference," *Cambridge, UK: CambridgeUniversityPress*, vol. 19, no. 2, 2000.

[4] K. Friston, R. Moran, and A. K. Seth, "Analysing connectivity with granger causality and dynamic causal modelling," *Current opinion in neurobiology*, vol. 23, no. 2, pp. 172–178, 2013.

[5] L. Barnett and A. K. Seth, "The mvgc multivariate granger causality toolbox: a new approach to granger-causal inference," *Journal of neuroscience methods*, vol. 223, pp. 50–68, 2014.

[6] ——, "Granger causality for state-space models," *Physical Review E*, vol. 91, no. 4, p. 040101, 2015.

[7] E. J. Hannan and M. Deistler, *The statistical theory of linear systems*. SIAM, 2012.

[8] Y. Antonacci, L. Astolfi, and L. Faes, "Testing different methodologies for granger causality estimation: a simulation study," in *2020 28th European signal processing conference (EUSIPCO)*. IEEE, 2021, pp. 940–944.

[9] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.

[10] N. Plub-in and J. Songsiri, "Estimation of granger causality of state-space models using a clustering with gaussian mixture model," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 3853–3858.

[11] M. Wang and Z. Fu, "A new method of nonlinear causality detection: Reservoir computing granger causality," *Chaos, Solitons & Fractals*, vol. 154, p. 111675, 2022.

[12] S. Beheshti, M. Hashemi, X.-P. Zhang, and N. Nikvand, "Noise invalidation denoising," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6007–6016, 2010.

[13] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

[14] L. Faes, G. Nollo, S. Stramaglia, and D. Marinazzo, "Multiscale granger causality," *Physical Review E*, vol. 96, no. 4, p. 042150, 2017.

[15] J. F. Geweke, "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984.

[16] V. Solo, "On causality i: sampling and noise," in *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007, pp. 3634–3639.

[17] M. Aoki, "Two complementary representations of multiple time series in state-space innovation forms," *Journal of Forecasting*, vol. 13, no. 2, pp. 69–90, 1994.

[18] P. Lancaster and L. Rodman, *Algebraic riccati equations*. Clarendon press, 1995.

[19] J. Neyman and E. S. Pearson, "Ix. on the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, 1933.

[20] L. Barnett and A. K. Seth, "Behaviour of granger causality under filtering: theoretical invariance and practical application," *Journal of neuroscience methods*, vol. 201, no. 2, pp. 404–419, 2011.

[21] A. Anzolin, J. Toppi, M. Petti, F. Cincotti, and L. Astolfi, "Seed-g: simulated eeg data generator for testing connectivity algorithms," *Sensors*, vol. 21, no. 11, p. 3632, 2021.