T.R.

ANTALYA BILIM UNIVERSITY

INSTITUTE OF POSTGRADUATE EDUCATION

DISSERTATION MASTER'S PROGRAM OF ELECTRICAL AND COMPUTER ENGINEERING

INFORMATION THEORETIC APPROACHES FOR MULTIVARIATE ANALYSIS AND THEIR APPLICATIONS

DISSERTATION

Prepared By

Humair Ali

Antalya - 2020

T.R.

ANTALYA BILIM UNIVERSITY

INSTITUTE OF POSTGRADUATE EDUCATION

DISSERTATION MASTER'S PROGRAM OF ELECTRICAL AND COMPUTER ENGINEERING

INFORMATION THEORETIC APPROACHES FOR MULTIVARIATE ANALYSIS AND THEIR APPLICATIONS

DISSERTATION

Prepared By

Humair Ali

Thesis Advisor

Asst. Prof. Dr. Deniz Gençağa

Antalya - 2020

APPROVAL/NOTIFICATION FORM

ANTALYA BİLİM UNIVERSİTY

INSTITUTE OF POST-GRADUATE EDUCATION

HUMAIR ALI, a M.Sc. student of Antalya Bılım University, Institute of Post Graduate Education, Electrical and Computer Engineering owning student ID 181212006, successfully defended the thesis/dissertation entitled "INFORMATION THEORETIC APPROACHES FOR MULTIVARIATE ANALYSIS AND THEIR APPLICATIONS", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Academic Tittle, Name-Surname, Signature

Thesis Advisor **:** Assistant. Prof. Dr. Deniz Gençağa, ………..………………

Jury Member **:** Prof. Dr. Cafer ÇALIŞKAN, ………………..

Jury Member **:** Assistant. Prof. Dr. Murat AK, ………………………………..

Director of The Institute **:** Prof. Dr. İbrahim Sani MERT……..…….. ………………

Date of Submission:

Date of Defence: 30/09/2020

ABSTRACT

Causality is one of the most challenging topics in science and engineering. In many applications, the cause and effect relationships among complex systems are not clear. In the literature, many information theoretic approaches, such as the Granger causality and Transfer Entropy, have been successfully applied to estimate the direction of interactions among random variables. However, the majority of these analysis have focused on the relationships between pairs of variables. In complex systems, the number of variables can increase to large numbers and analysis of the interactions of each pair can be problematic.

In this thesis, we propose using conditional Transfer Entropy in order to seek out the hidden information among many interacting variables. We show that pairwise transfer entropy can be effective in identifying the directional interactions but some of these relationships can be due to the interaction with a third variable. The computer simulations verify these on synthetic coupled autoregressive model and also on Protein A4 (S100A4) data, where we show that by conditioning on certain variables, we can obtain more insight on the interactions.

DEDICATION AND ACKNOWLEDGMENT

I dedicate this thesis to my parents, friends, and teachers who always supported me to this day of my life.

I am very grateful to my advisor "Assistant. Prof. Dr. Deniz Gençağa" who guided and encouraged me throughout the thesis, who opened the doors for me in this field of research.

I would also like to thank the jury members Prof. Dr. Cafer Çalışkan and Assistant Professor Dr. Murat Ak.

I would like to thank Prof. Dr. Erol Eroğlu and Dr. Ahmet Yildirim for his help in protein data.

I would like to thank Assoc. Prof. Dr. Joseph Lizier for the JIDT package.

CONTENTS

INSTITUTE OF SCIENCES

ELECTRICAL AND COMPUTER ENGINEERING

MASTER OF SCIENCE PROGRAM WITH THESIS

ACADEMIC DECLARATION

I hereby declare that this master's thesis titled " Information theoretic approaches for multivariate analysis and their applications" has been written by myself under the academic rules and ethical conduct of the Antalya Bilim University.

I also declare that the work attached to this declaration complies with the university requirements and is my work.

I also declare that all materials used in this thesis consist of the mentioned resources in the reference list. I verify all these with my honor.

30 /09/ 2020

Humair Ali

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

TE	Transfer entropy
KDE	Kernel density estimation
MI	Mutual information
PID	Partial information decomposition
CC	Cross correlation
PDF	Probability density function
IID	Interaction information decomposition
AR	Autoregressive
ITE	Interaction transfer entropy
MA	Moving average
SS	State Space
KNN	K-nearest neighbor
LE	Lorenz equation
CMI	Conditional Mutual information

CHAPTER ONE

1. Introduction

Information flow with related measures like storage, transfer of information, and interaction has gained importance in many fields after Claude Shannon's information theory was presented in the late forties. Information theory, as a mathematical tool, sets rules to examine the relationships between the variables of complex systems and their sub-systems.

In recent years, identifying interdependency and causal relations among multivariate time series in data analysis of complicated systems has been an active research topic in many fields, including neuroscience, psychology, or computational bioinformatics.

Variables closely coupled show that information related to one variable can give information related to other variables at another moment. Detecting and characterizing their relations by observing the system's behavior can be difficult depending on the number of variables involved, magnitude, and nature of the coupling. Indeed, identifying different parts of a system could be an ill-posed problem because the definition of weak or strong connection is necessarily subjective. The coupling path among two variables is frequently analyzed in terms of one variable driving another in order to detect causal relationships. One variable at a given time influence the future values of another variable, causing an information flow.

A term called entropy (TE) is an asymmetric measure suggested for estimating such coupling directions. However, it is not easy to accurately evaluate the causation relationship. Each technique has some specific tuning parameters, and there is no agreement on an optimized method of approximating the entropy from a dataset [1].

Objective

In this thesis, the main goal is to better understand the statistical relationships among two or more random variables using information-theoretic approaches. For this purpose, we propose using pairwise and conditional dependencies estimated from data. These are entropy, transfer entropy, and mutual information. Among many methods, Kraskov technique [10] has been used for these approaches due to its advantages compared to other approaches, such as histogram and kernel estimation.

Comparisons between the pairwise TE estimations and conditional TE estimations are demonstrated on synthetic Autoregressive (AR) process data and later the molecules data about the Protein A4 (S100A4).

In conclusion, we have shown that causal analysis of multivariate systems can be explained in more detail by using conditional TE rather than the pairwise TE among each pair. When the conditional information flows between the alpha carbons of proteins are analyzed, we demonstrate that according to the conditioned residue, it can be possible to better understand the spatial interactions.

Cause and effect relationships:

In earlier literature, Schreiber [2] proposed the TE for the first time to analyze the asymmetric statistical dependencies between random variables. Lizier et al. (2010) presented how the general concept of information techniques could be applied to the system with multiple dynamic interactions. Multiple interacting systems could be represented by using multivariate stochastic processes, which could be divided into very fundamental parts of computation within the information dynamics setup.

Faes et al. (2017) showed a way of dissecting the information carried by the target of a multivariate dynamic system into a form of atoms of information, and these subdivisions are essential parts of the traditional information dynamics like information storage, information transfer, and predictive data. These basic atoms are beneficial for explaining the contributions of the systems individually in-network linked to the target system's dynamics and helpful in describing the equilibrium of synergy and redundancy among sources. Theoretical and real-time data were analyzed to show how predictive information was obtained from the transfer, storage, and modification mutually for multiple source systems. It was indicated that several types of informatics measured must be computed together in the form of subparts of the entire statistical process related to a multivariate system's target because any change in the network causes these measurements to undergo concurrent changes [3].

Bertschinger et al. (2014) reported work on utilizing partial information decomposition. They applied it to a 2D Icing model, where the temperature was variable, to decompose information in the spins' triplets. Further, peaks of synergetic and redundant information occurred in the disordered and critical phases. Then decomposition was performed on the cells of data in 1D elementary cellular automata taking from source present in those cells' neighbors. They also achieved a reduction in dimensions of PID atoms [4].

Wibral et al. (2017) described the application of partial information decomposition as a unified approach in the biological field of neural sciences. Neural systems like canonical microcircuits are made of neo-cortex involved in many different processes. A common rule means a goal function of information processing was required to implement in this structure. Thus, a framework independent of the domain was needed to establish this principle, which could be achieved through transfer entropy. Causality analysis between a pair cannot provide the required framework, as the neural systems involve a large number of inputs. They used partial information decomposition (PID) theory to achieve this. Partial information decomposition enables evaluating the information that multivariable give individually, shared, and jointly about the output. PID allowed comparison among goal functions in a common framework and provided a technique to design new functions by utilizing first principles. A goal function with synergy was designed by combining previous knowledge in a synergetic way and by analyzing it. The newly developed process was not investigated before and found useful for information processing in neural sciences [5].

Barrett (2015) summarized information redundancy and synergy shared in Gaussian systems that could be static or dynamic. The information about the target carried by two or more source variables could be characterized by using partial information decomposition, which decomposes the information into unique, redundant, or synergetic parts. Earlier research gave formulas for PID suggested that redundancy shrinks to the minimal information, so do not depend upon correlation among sources, and synergy provides extra information given by a weak source for a strong source known value. It can be increased or decreased depending on the correlation among sources. He found that the joint information in both sources was larger than the sum of individual information for each source. It was shown that how these findings could be used to measure information-dependent measures related to complexity and information transfer for general purpose and neural sciences. Furthermore, an approach was given based on independent formulas provided for redundancy and synergy that could be applied on continuous data of time series and could be used to quantify and characterize information sharing among complex variables systems [6].

Williams et al. (2010) demonstrated how to measure multivariate information so that the structure of interactions among variables become visible, and the issue of negative information related to the method of interaction information was eliminated. In the first step, this was achieved by describing a generalized measure related to redundant information that fulfilled intuitive characteristics for redundancy measurement. The parameter induced a lattice structure, known as redundancy lattice, over the possible information sources sets different ways of showing that the information could be distributed among a set of sources. Then the lattice could be used to quantify partial information that finds the unique information denoted by each combination of sources. Furthermore, they showed partial information terms obtained on the decomposition of mutual information so that complete information obtained from the source is divided into partial information contributions [7].

Structure of the thesis

In this thesis, we propose a new approach to better understand the directional statistical interactions between multivariate systems. We show that we can get more insight about these interactions by using conditional TE rather than pairwise TE which is applied in a pairwise manner among each component. We compare both methods first on a synthetic AR process and then on the detection of allosteric effects among the residues of a protein molecule. After this problem statement, the sequel of thesis is as follows:

In the second chapter, we briefly give background on entropy, including transfer entropy, mutual information, and conditional transfer entropy. Further, we present the extension of a multivariate case in entropy and methods. Moreover, we also give partial information decomposition and usage.

Chapter three of the thesis includes different estimation techniques known as Kraskov estimator, Kernel density estimation and histogram, for the purpose of estimating information theoretic quantities from data.

In the fourth chapter, we demonstrate the simulation results pertaining to the applications of pairwise TE, MI and Conditional TE both among multivariate members of a coupled AR system and among the alpha carbons of protein A4 (S100A4). Simulations of this section have been accomplished by using MATLAB. In the last chapter, we discuss the results found in our research, followed by suggestions for future work.

CHAPTER TWO

2. Background

2.1 Entropy

Entropy is a way of quantifying information contained in a variable or within a signal. It is used to quantify the amount of the uncertainty about a random variable. The concept of entropy inside the information theory was first introduced by Shannon in a paper named "A Mathematical Theory of Communication" [12]. In his paper, Shannon states that Entropy will be maximum in the case where all the outcomes are equiprobable. For this condition, entropy will increase with the increase in the outcomes. He adds that a zero entropy will be possible if the outcome is known. The mathematical expression of entropy for a discrete random variable (f) is given as follows:

$$H_p(f) = -\sum_{a \in A} P(f = a) lnP(f = a) \qquad (2.1)$$

Generally, logarithms of base 2 is used in place of natural log. For these two different logarithms entropy show different units that is for natural logarithm entropy is measured in "nats" while for base 2 logarithm "bits" is used for entropy. A continuous random variable f has a probability density function (pdf) p, so we can also write:

$$H(f) = -\int_{f \in A} p(f) logp(f) \qquad (2.2)$$

where p denotes the probability density function.

2.2 Mutual Information

Mutual information is another information theoretic quantity defined over pdf. It measures the amount of information that is shared by two random variables. It means mutual information gives the idea about the information that we know about one random variable given the observance of another variable [11].

Let's take an example of a die. Suppose that there are two variables X and Y. X denotes the roll of a die having six sides, and Y shows about even or odd condition of the die. These variables are related to the single die; therefore, the variable Y will surely give some information about the variable X. So, mutual information can be used for these variables. But if the two variables are related to two different dies, then it will not be possible for any variable to give information about the other variable. In this case, they cannot make use of mutual information[12].

The other definition of mutual information is given as the average decrease in uncertainty within one variable X resulting from the known value of other variable Y. The mutual information can be represented formally for X and Y, two random variables, having combined distribution P (X, Y)

$$I(X;Y) = \sum_{x\epsilon\chi, y\in\Upsilon} P(x,y) log \frac{P(x,y)}{P(x)P(y)} \quad (2.3)$$

Here in the equation, the terms P(X) and P(Y) presenting marginal distributions for X and Y, and these distributions are obtained by a process called marginalization. The Kullback-Leibler divergence can be seen from the above expression between product (P(x) P(y)) and joint ($P(x, y)$) distributions. The term MI can further be given in the form of the expected value above.

When the expected value is evaluated over divergence among two probabilities named marginal probability and conditional probability, the resultant expression is also termed as mutual information [16].

$$\mathrm{I(X;Y)} = \sum_{x\epsilon\chi} P(X=x) \sum_{y\epsilon\Upsilon} P(Y=y|X=x) \cdot \log\left(\frac{P(Y=y|X=x)}{P(Y=y)}\right) \quad (2.4)$$

The Shannon's entropy is related to it through some of the expressions given below:

$$I(X;Y) = H[X] - H[(X|Y)] \quad (2.5)$$

$$= H[X] + H[Y] - H[X,Y] \quad (2.6)$$

$$= H[X,Y] - H[(X|Y)] - H[(Y|X)] \quad (2.7)$$

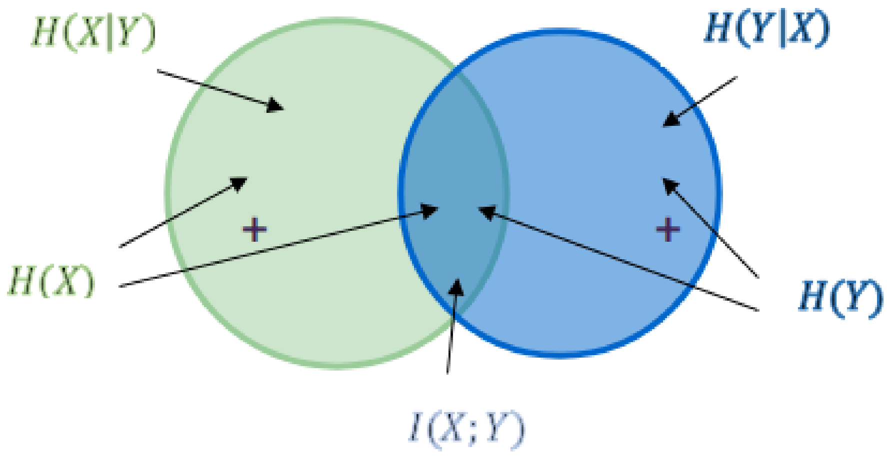

Figure 1: Venn diagram showing relation between entropy and mutual information.

If Y and X are independent, then the mutual information will always be zero. It will be lower than the minimum entropy of any of the variables:

$$0 \le I(X;Y) \le \min(H[X], H[Y]) \tag{2.8}$$

It can be further extended for continuous variables.

2.3 Transfer Entropy

Transfer entropy is an information theoretic quantity that can be utilized to quantify the amount of information flow between two random variables. Unlike the MI, TE is not asymmetric and it gives idea about the causal interactions. In a dynamic system, TE utilizes the time-bound information among two systems. It determines the reduction in the uncertainty about a random variable's dependency on its own past values by means of using its dependency on another random variable's past values, or it can be treated as the extra information we get by looking at the past values of another process accompanying the original one [12].

It can be defined explicitly for source node Y and X destination using the preceding state of the source named y_n and the destination's next state (x_{n+1}) conditioned on the destination' previous k states at the time n+1

$$TE_{YX}\, T\left(\mathrm{X}_{\mathrm{i}+1}|\mathrm{X}_{\mathrm{i}}^{(k)}, \mathrm{Y}_{\mathrm{i}}^{\mathrm{l}}\right) = \sum_{x_{\mathrm{i}+1},\; x_{\mathrm{i}}^{k}, \mathrm{y}_{\mathrm{i}}^{\mathrm{l}}} p\left(x_{\mathrm{i}+1}, x_{\mathrm{i}}^{k}, \mathrm{y}_{\mathrm{i}}^{\mathrm{l}}\right) \log_2 \frac{p\left(x_{\mathrm{i}+1}|x_{\mathrm{i}}^{k}, \mathrm{y}_{\mathrm{i}}^{\mathrm{l}}\right)}{p\left(x_{\mathrm{i}+1}\middle|x_{\mathrm{i}}^{(k)}\right)} \tag{2.9}$$

Where $x_i^{(k)} = \{x_i, x_{i-k+1}\}$ and $y_i^{(l)} = \{y_i, y_{i-l+1}\}$ are past states, and X and Y are k^{th} and I^{th} order markov processes, respectively, such that X depends on the *k* pervious values and *Y* depends on the *l* pervious values. In the literature, *k* and *l* are also know as the embedding dimensions.

It can also be expressed as

$$T_{Y\to X} = H(X_{t+1}|X_t) - H(X_{t+1}|X_t, Y_t) \quad (2.10)$$

Alternatively, transfer entropy can show in terms of the addition of two mutual information quantities or conditional information given below:

$$T_k(Y \to X) = I(Y; (X_0|X_k) \quad (2.11)$$

$$T_k(Y \to X) = I\big(Y; X_0, X_{(k)}\big) - I(Y; X_{(k)}) \quad (2.12)$$

Where

X'= destination's next state

$X^{(k)}$ = Previous k states

The term transfer entropy from X to Y has a value different from transfer entropy from Y to X; hence we can say that the TE is asymmetric. It measures a pair of variables. In the absence of proper conditioning for observations above two in number, it is impossible to differentiate between transfer entropy and indirect causality, resulting in spurious causality.

2.4 Extension to Multivariate Case: Multivariate entropy, MI and TE

We have discussed theoretic measures for just two variables. Still, there are various fields and researches involving a large number of variables, so we require measuring techniques for multiple variables. Here in this section, we will discuss multivariate analysis techniques and information measures for various variables [14].

2.4.1 Multivariate Entropy

Let us consider n number of variables continuous in nature, shown as follows:

$$v_n = \{X_1, X_2 \dots \dots \dots X_n\} \quad (2.13)$$

We will use notation v_{n-1} for set of variables in the absence of (X_n). Now the probabilities for the set of variables is denoted as

$\mathrm{p}(v_n)$ = Probability density function

$p(X_n|v_{n-1})$ = Conditioned probability density function

Joint entropy is used for multiple variables and it is the term to quantify uncertainty related to the multiple variables. It has a mathematical formula which is expressed as

$$H(v_n) = -\sum_s \mathrm{p}(v_n) \log(\mathrm{p}(v_n)) \tag{2.14}$$

Where's the set of values is (v_n).

2.4.2 Multivariate Mutual Information

Multivariate Mutual information is a way of defining how information is shared among variables. It is used for a set of variables. A general concept of mutual information for three or more than three variables is known as interaction information. Mutual information in the case of three variables is expressed as follows:

$$I(X_i, X_j, X_k) = I(X_i, X_j) - (X_i, X_j|X_k) \tag{2.15}$$

Now the mutual information for three variables in terms of entropies.

$$I(X_i, X_j, X_k) = H(X_i) + H(X_j) + H(X_k) - H(X_i, X_j)$$

$$-H(X_i, X_k) - H(X_j, X_k) + H(X_i, X_j, X_k) \tag{2.16}$$

The Multivariate MI may be negative, positive or zero. For three variables, X_i, X_j and X_k, the simplest case say, that X_i gives a certain amount of knowledge about X_k. This is seemed to be possible in the effect of belongs one of the variables may increase or decrease dependence between the others. As a trivial case, consider the situation in the multivariate product where variables X_i and X_j are independent when X_k is not known, but become dependent given X_k . For this case, $I(X_i, X_j, X_k)$ is may be negative. This shows that mutual information for multivariate variables need not be always positive, by expanding it in terms of parameters of probability up to the second order.

2.4.3 Multivariate Transfer Entropy

The information flow from a collection of variables to a target variable can be quantified by multivariate transfer entropy. In particular, $T_k(Y \rightarrow X)$ will capture the information transfer that occurs due to the interaction between a set of source variables (e.g., an exclusive-OR type interaction); uni-variate analysis cannot grasp this. As such, in the multivariate TE, we now have a measure for information transfer that is non-linear, directional, and captures collective interactions. Figure 2 shows the entropy among source Y and destination X for transfer of information [3].

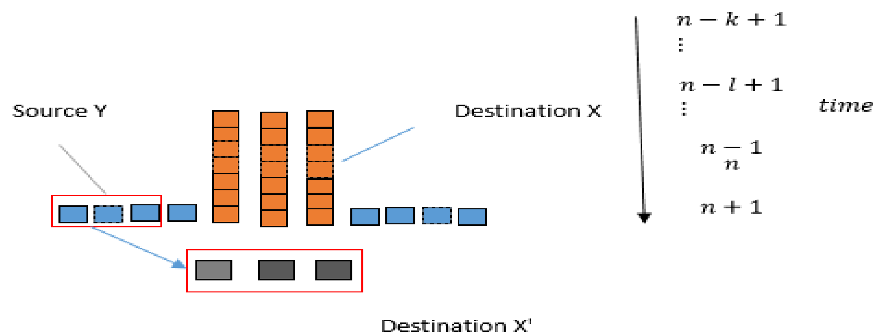

Figure 2: Transfer entropy for multiple variables.

Multivariate TE has proven to be better for measuring non-linear and linear casual relationships effectively in neuroscience. MTE is an extension of pairwise entropy transfer.

In an M variable system, if we show the source by X, target by Y and the remaining system $Z = \{Z^k\}_{k=1,\ldots,M-2}$, we can evaluate the information flow from a source system X to the target system Y, by considering the rest of the complex system. This can be represented as follows:

$$TE(X \rightarrow Y|Z) = \sum p(y_{1:n,}x_{1:n-1,}z_{1:n-1}) log \frac{p(y_n|x_{1:n-1,}y_{1:n-1,}z_{1:n-1})}{p(y_n|y_{1:n-1,}z_{1:n-1})} \qquad (2.17)$$

Where $x_{1:n-1}$ represents the entire previous history of the process x [43], $z_{1:n-1}$ shows the pas of Z, and $y_{1:n,}$ represents all past and the current states.

2.5 Partial Information Decomposition

Williams and beer presented the concept of PID that is partial information decomposition. It provides a basis for decomposing the information shared among the collection of variables known as inputs and the other variable output. In other words, we can say that the purpose of the partial information decomposition is decomposing the information known as joint mutual information among sources related to target in the form of non-negative values. It can apply to any desired number of sources [6]. It decomposes the information that a collection of the variable has about another variable into the following forms:

- Unique
- Redundant
- Synergistic
- Mixed

Hence, it decomposes information among set of variables and single variable.

2.5.1 Explanation: Redundancy, Synergy and Uniqueness

Now consider a simple case of two variables where

$$S = \{X_1, X_2\} \qquad (2.18)$$

The sum of the partial information values gives relevant mutual information. Now for a two-

variable case, we can say that: Unique information is given by each individual variable X about the variable Y, and Redundant type of information is achieved by both variables X. The combined variables X give synergetic effect about Y.

The mutual information is expressed in terms of partial information decomposition which can be written as follows:

$$I(X1, X2; Y) \equiv Synergy(Y; X1, X2) + Unique(Y; X1) \qquad (2.19)$$

$$+ Unique(Y; X2) + Redundancy(Y; X1, X2)$$

$$I(X1; Y) \equiv Unique(Y; X1) + Redundancy(Y; X1, X2) \qquad (2.20)$$

$$I(X2; Y) \equiv Unique(Y; X2) + Redundancy(Y; X1, X2) \qquad (2.21)$$

William and Beer introduced that the redundancy of information is equal to the minimum information term. The new function that is the minimum information function, shows the intuitive behavior of redundancy. It means that for variable Y, redundancy gives by both the variables X regarding Y [7]. The minimum function is further related to specific information and the expression of the particular information given as:

$$I_{spec}(y;X) = \sum_{x\in\chi} p(x|y)\left[\, log\left(\frac{1}{p(y)}\right) - log\left(\frac{1}{p(y|x)}\right)\right] \quad (2.22)$$

Here the equation shows how much of the information regarding the specific state of variable Y given by the variable X is provided by specific information [13].

The minimum function evaluates by comparing the amounts of data given by individual variables X about each set of variable Y. The expression for the minimum information given as:

$$I_{min}(Y;X_1,X_2) = \sum_{y\in Y} p(y) min_{X_i}\, I_{spec}(y;X_i) \quad (2.23)$$

Where each variable X is considered individually.

Further, redundancy is equivalent to the minimum information function given by William and Beer.

$$Redundancy(Y;X_1,X_2) = I_{min}(Y;X_1,X_2) \quad (2.24)$$

However, minimum information and redundant information has distinct but critical difference among them. The difference lies in the way the variables X contributed to providing information. Redundancy is the information obtained by both the variables X, while the information named minimum information shows the minimum average amount of the information provided by any variable X [9].

2.6 Conditional entropy, MI, and Transfer entropy

2.6.1 Conditional entropy

For a known value of one random variable, how much information is required to define the random variable’s outcome is shown by a term called conditional entropy. It can also be specified in terms of expected values. When the expected value for the entropies related to conditional distributions averages above the conditioned variable, it is also called conditional entropy.

Mathematical expression for conditional entropy of X, a random variable, specified another Y variable shows the uncertainty for X for the known value of Y then for (X, Y) ~ p(x, y) conditional entropy is given as follows:

$$H(Y|X) = \sum_i p(x = x_i) H(Y|X = x_i) \quad (2.25)$$

$$= -\sum_i p(x = x_i) \sum_i p(y = y_i|X = x_i) logp(y = y_i|X = x_i) \quad (2.26)$$

$$= -\sum_i \sum_j p((y = y_i|X = x_i)) log(y = y_i|X = x_i) \quad (2.27)$$

$$= -E\ logp(Y|X) \quad (2.28)$$

This expression shows that the entropy of two random variables is equal to the sum of the entropy of one and the other variable's conditional entropy. If and only if one variable is a deterministic function of the other variable, then for this condition, the conditional entropy will be zero or greater than zero [1].

2.6.2 Conditional mutual information

An essential parameter of information-theoretic that relates the notion of mutual information with conditions is conditional mutual information. It is a quantity that quantifies the dependence of two variables X and Y, given in the presence of a third variable Z. Let's consider complementary as well as mutually orthogonal subspaces X, Y, Z. Then the expression of mutual information between X and Y when conditioned on Z is given as

$$I_p(X, Y| Z) = D(p(x , y|z) \parallel p(x|z)\, p(y|z) \quad (2.29)$$

$$= E_p \left[log \frac{(x, y|z)}{p(x|z)\, p(y|z)} \right] \quad (2.30)$$

$$= H_p(X, Z) + H_p(Y, Z) - H_p(Z) - H_p(X, Y, Z) \quad (2.31)$$

Different informatics measures for x, y, z variables show three bottoms left, bottom right, and upper circles—different colors like yellow, brown, orange, and green representing condition-based mutual information [12].

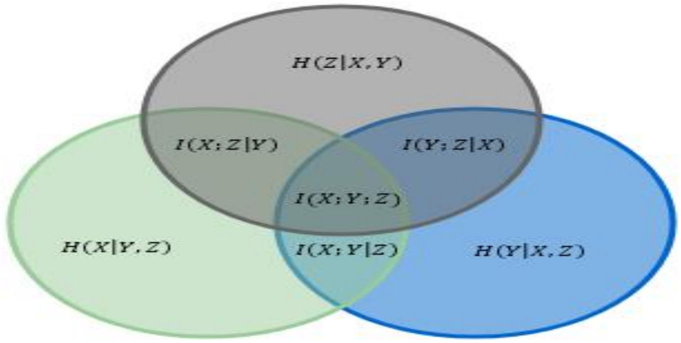

Figure 3: Venn diagram showing Conditional Mutual informatics measures.

2.6.3 Conditional Transfer entropy

The transfer entropy for two series X and Y is given as

$$T_{X \to Y} = I\left[(Y_t : X_{0:t} | Y_{0:t})\right] \qquad (2.32)$$

This expression shows that X influences Y. But a question arises is that whether the influence is direct? The answer lies in a term named conditional entropy transfer. This term determines whether the dependence is direct or not by the use of a third series. The term conditional entropy defines the sharing of information among two series based on the condition of the third series past value.

Mathematically it is given as

$$C_{X \to Y\,|(Z)} = I\left[Y_t : X_{o:t} \mid Z_{0:t}\right] \qquad (2.33)$$

Conditional entropy also terms as causation entropy. For example, considering a system representing $X \to Z \to Y$. Z influenced by X and Y is influenced by Z. Any effect that X imposed on Y also goes to Z. For this scenario, the entropy will be greater than zero, regardless of whether X does not directly affect Y. In comparison, the conditional entropy will be zero because of the conditioning imposed on Z [1].

Chapter Three

3. Estimation Techniques

It is becoming popular and useful to analyze the complicated system as distributed information treating systems in computational bioinformatics, artificial life, and neuroscience. It leads to the intense use of information-theoretic quantities for analyzing the dynamic process of complicated systems related to these fields. This research is aimed to develop a deep insight for information-theoretic measures and its application in the field of biochemistry. Many dynamic processes are related to proteins and we propose using conditional TE to better understand these processes. In the literature, entropy, MI, TE and their variants are estimated using data in many ways. This section briefly covers the most frequently used approaches.

There are lot of methodologies presented for estimating information-theoretic quantities. Modeling the probability density function (PDF) from data is the most challenging part of the majority of the methods. Estimating probability density functions can be performed using the following techniques:

- Histograms
- Kernels density estimation
- Kraskov estimation

3.1 The Histogram

The simplest and oldest technique for estimating PDF is histogram. This technique consists of simple process in which we divide observations into non-overlapping bins and by counting the number of observations in each bin. Once these counts in each bin are normalized with the total number of data points, a crude estimate of any PDF can be obtained. However, this approach has severe problems in three or more dimensional estimations and therefore it is not implemented in this thesis where multivariate PDF's are considered.

3.2 The Kernel Estimation

There are many shortcomings related to the histogram, and further, the sensitivity related to the size of bins and origin's choice is also a problem. An alternative technique to bypass the problem of discontinuities is the kernel estimation. Mathematically it is expressed as

$$\hat{f}X(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right) \tag{3.1}$$

Here kernel function is denoted by K, the number of observations by N and window width by h. A possible choice is a Gaussian kernel for K, leading towards an estimator called density estimator given as follows:

$$\hat{f}(x) = \frac{1}{Nh\sqrt{2\pi}}\sum_{i=1}^{N} exp\left(\frac{-(x-x_i)^2}{2h^2}\right) \tag{3.2}$$

In practice, the fundamental element is the choice of bandwidth h. Its optimum Value is that which decreases the asymptotic mean integrated squared error among estimation and actual density. Thus, the bandwidth is expressed as follows:

$$\hat{h}_{rot} \approx \sigma\left(\frac{4}{f+2}\right)^{1/(f+4)} N^{-1/(f+4)} \tag{3.3}$$

Where dimensionality is denoted by f.

There are various kinds of kernels available, but the most common kernels are Rectangular and Gaussian. Each kernel has a specific bandwidth or radius. This estimation technique can further utilize to quantify entropy by using generalized correlation integral that is estimated by correlation sum [Prichard and Theiler,]. For estimating entropy if rectangular kernel use, it proved to be more accurate than the histogram method. For this technique, correlation sums utilize to estimate entropies and correlation sum evaluated by using Grassberger-Procaccia (GP) algorithm or some other efficient method [10].

3.3 The Kraskov Estimation

The most challenging part of the estimation of mutual information is the probability densities. So, it is preferred to use a technique that does not involve probability density estimations. One such technique is the Kraskov estimation (KGS) that directly estimates mutual information. It is a common variant of the Naïve kNN estimation. The basic concept behind this estimation is that there must be strong relation among spaces *X* and *Y* if the neighbor for an observation in space *X* corresponds to the same neighbor in the space *Y*. The general rule of this estimation is that for all density estimation in different spaces, the same length scales will be used for *k* closest neighbor distance as in joint space, making the bias smaller. This estimation is

$$\hat{H}\ (X) = -\Psi(K) + \Psi(n) + \log(c_d) + \frac{d}{N}\sum_{n=1}^{N} \log\left(\varepsilon_X(N, K)\right) \qquad (3.4)$$

Where

Ψ=digamma function

N=number of samples in space X

d=dimensionality of samples

c_d=volume of a unitary ball having d-dimension

Then two little bit different estimators derived, the most common of which is given as

$$\hat{I}(X;Y) = \Psi(N) + \Psi(K) - \frac{1}{K} - \frac{1}{N}\sum_{i=1}^{N}\left(\Psi\left(\tau_{x_i}\right)\right) + (\Psi\left(\tau_{y_i}\right)) \qquad (3.5)$$

The distance of τ_{x_i} from $x_i$is not more than $0.5 \times \varepsilon(n.K) = 0.5 \times \max\left(\varepsilon_X(n,K), \varepsilon_Y(n,K)\right)$.

The factor k is a free parameter that shows the neighborhood's size for local density estimation. A larger *k* will reduce the variance related to estimation, so the smaller k proves to be more accurate. Whereas it requires that the value of *k* increases with *N* for estimation of consistent density and some weak conditional entropy approximates converge for any fixed value of *k*. To evaluate mutual information in the joint space *x* we set value of *k* that finds the value of ϵ for each point *x*. KSG is one of the most accurate techniques to evaluate mutual information. *K* a free parameter, is first selected, and then the estimation is applied to the data. It can vary the value of *k*, which allows exploring characteristics within the probability distribution across spatial scales, and it will result in a bias-variance tradeoff. The optimum value of *k* based on the structure of spatial characteristics within the data can be non-trivial and might occur on several spatial scales. The optimum *k* value also dependent upon the *N* because the fine characteristic features exist only at a higher level of sampling density [10]

CHAPTER FOUR

4. Application and Simulations

In this section, the performance of information-theoretic quantities, such as MI, TE, are demonstrated by computer simulations on two cases, namely the Coupled Autoregressive model and coupled molecular model. In this regard, three estimation techniques are discussed and evaluated in detail. Moreover, as the main contribution of the thesis, the comparison between pairwise TE, conditional TE have been analyzed on these two situations. Furthermore, the proposed approach is assessed by applying different reliable techniques discussed in the literature. Kraskov and Kernel models are implemented on MATLAB to provide insights into a non-linear autoregressive model. To analyze the applicability of models, the proposed approach was used on various alpha carbons of proteins.

4.1 Application to the AR Model:

In this section, it is shown how information-theoretic approaches could be used for the analysis of complex dynamic systems. We assume three time series that are non-linearly dependent and coupled via autoregressive model (AR) as shown below:

$$\begin{aligned} y_1(t) &= a_1 y_1(t-1) + n_1(t) \\ y_2(t) &= b_1 y_2(t-1) + b_2 y_2(t-2) + b_3 y_1(t-1) + n_2(t) \\ y_3(t) &= c_1 y_3(t-1) + c_2 y_2(t-1) + c_3 y_1(t-2) + n_3(t) \end{aligned} \quad (4.1)$$

Where n_1, n_2, n_3 are drawn from Gaussian noise. The following diagram depicts the dependency of three state variables on their own past samples and on the other variables' past samples. The blue arrows indicate a one-time lag whereas the orange represents a two-time lag.

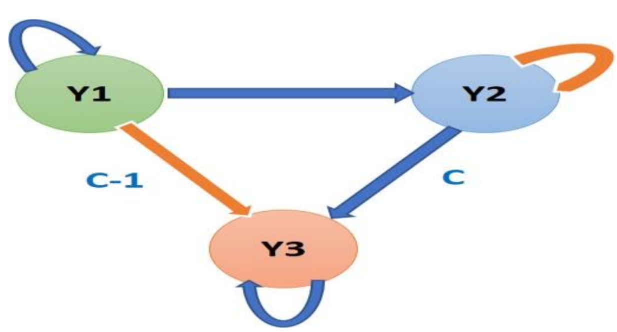

Figure 4: Interaction between variables in the AR System.

This model shown by Fig. 4 depicts multivariate variables and the relations and interactions among them in a complicated dynamic system. The sign connected two variables shows their interaction, and the other sign just related to the single variable indicates that the variable also depends on itself.

4.1.1 Results of Conditional and pairwise transfer entropies on the AR model

Here, a conditional and pairwise approach on the couples is applied for this method and a model-free assessment in TE, using the time series simulation with equation (4.1) is performed. The AR model involves three interacting variables. These simulations were repeated 100 times for ensemble averaging purposes. We analyze the TE between each pair of variables with one conditional. This analysis is setup with Kraskov estimator with the *k=4* for the nearest point, and maximum lags 2 in our model.

Table 1: AR matrix of Conditional and Pairwise TE.

AR Conditional TE		**AR Pairwise TE**	
$Y_1 \rightarrow Y_3 \mid Y_2$ 0.02	$Y_2 \rightarrow Y_3 \mid Y_1$ 0.17	$Y_1 \rightarrow Y_3$ 0.11	$Y_2 \rightarrow Y_3$ 0.27
$Y_1 \rightarrow Y_2 \mid Y_3$ 0.04		$Y_1 \rightarrow Y_2$ 0.04	

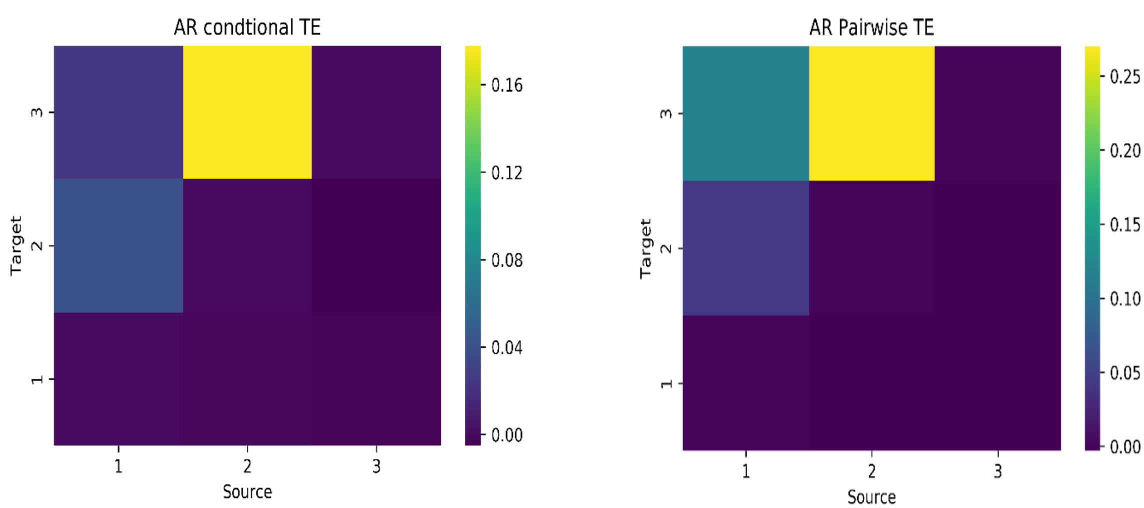

Figure 5:The matrix represents the Conditional and pairwise TE estimation in the AR system. The color (dark to light) indicates the TE average over 100 realizations of the simulations.

In Fig. 5, the conditional and pairwise TE values for the 3 variable AR system are illustrated. In Fig. 5a, the number on the x axis denotes the relevant source, the number on the y axis represents the relevant target, and the remaining third variable is the conditional. That is, the value of the conditional TE (Y1→Y2|Y3) is given at the second from the top of the first column of the matrix.

As we can see in Fig. 5, our approach detects the accurate information transfer from Y1 to Y2 given Y3 (Y1→Y2|Y3), which is higher than the TE(Y2→Y1|Y3), which are obtained as a result of ensemble averaging of 100 realizations. The rest of the matrices denote the conditional TE results pertaining to (Y3, Y2), (Y3, Y1), (Y1, Y3), (Y2, Y3) by hidden conditional variables in the non-linear AR system. When the results are compared with the equations, it is observed that the transfer directions of the actual information can be found successfully.

In Fig5b, the actual number values of the matrices are shown as well, due to the fact that the small differences are not remarkable in the heatmaps. According to the comparison between the pairwise and the conditional results, we notice that the pairwise TE between 2 and 3 is 0.27, whereas this value goes down to 0.17 when the same pair is taken by conditioning on the first variable. This means that pairwise TE includes some redundant amount which is actually coming from 1 and affecting 3. Similarly, pairwise TE between 1 and 3 also includes redundant information coming from the indirect effect of 1 via 2 to 3. These results conclude that we can have more in-depth analysis with conditional results compared to pairwise results.

4.2 Applications of Transfer Entropy in Allosteric effects of residues in proteins

In biochemistry, a protein's activity can be regulated by an effector at a site which is outside of the protein's own region. This effect is known as allostery and the dynamics of the biomolecules can be changed by means of such effects. Based on recent [24], it has been shown that the allosteric effects originate because of the atomic fluctuations and these are also related to the changes in entropy interactions. Here, "Data used in this study which were derived from 1 micro-second long molecular dynamics (md) simulations of calcium-binding protein A4 (S100A4), were provided by Dr. Erol Eroglu and Dr. Ahmet Yildirim. This protein comprises 186 amino acids long residues, is considered as a drug target for the treatment of some cancer tumors. For data production, the experimental x-ray crystallographic structure of the protein (PDB access code; 3ZWH) was used for md simulations. The simulations were carried out using Gromacs software, the atomic coordinate frames of the protein were written every pico-second and total of 5000 frames were recorded for TE analysis. Only atomic fluctuation data of alpha carbons (CA) of residues were used for TE calculations". Here, the conditional interactions between different residues compared with pairwise relationships for 50 of these alpha carbons.

Table 2:Tested on the below methods.

APPLICATION	METHOD	SOURCES	DESTINATION
TRANSFER ENTROPY	Pairwise	R1 ……. R50	R2 ……. T50
	Mutual Information	R1 …. R50	R2 …. R50
	Conditional	R1 ……. R50	R2 ……. R50

4.3 Pairwise TE on Allosteric effect analysis

In a complex network like those formed by residues of a protein molecule, the pairwise interaction is affected on a broader network framework. Estimation of the TE between residues helps us to define and identify the functional and structural links. In the following figure, a cartoon depicts such interactions among three residues:

Here, using the time-series data, TE values are estimated for each pair of 50 alpha carbons, i.e. ($TE_i \rightarrow TE_j$ and $TE_j \rightarrow TE_i$), where $i, j = 1,2, \ldots ,50$.

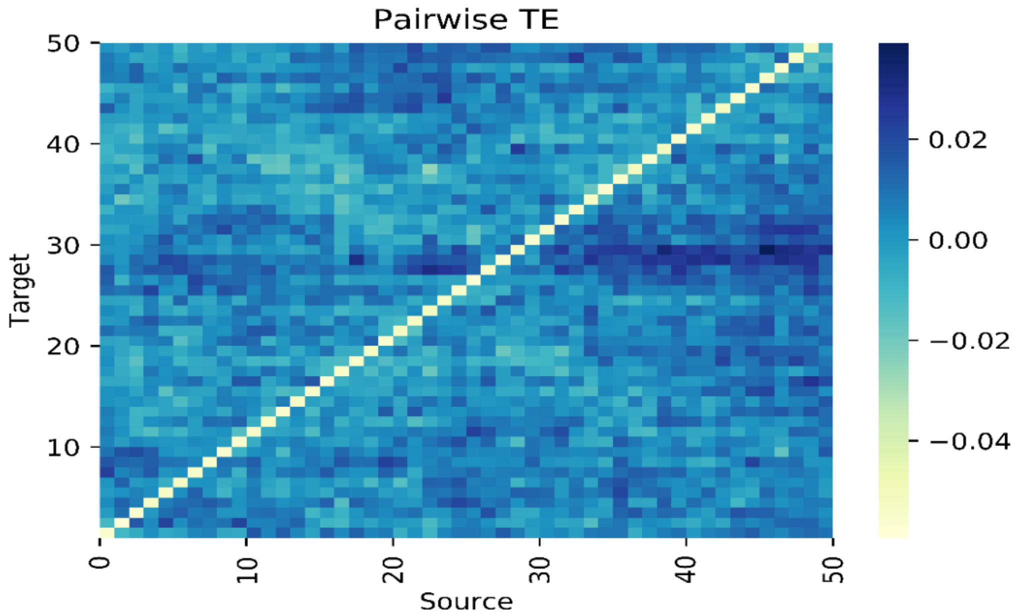

Figure 6:Pairwise TE as a function of the source to target. The results were obtained over 50 realizations in proteins residue.

Fig. 6 shows the pairwise TE values between each residue. We observe that pairwise TE results are close to each other for the chosen 50 residues. However, it should be noted that the information flow from the residues between 30 and 50 to residue number 30 (and vicinity) is much more remarkable compared to the rest. In addition to this, the values around the diagonals seem to be lower than the rest.

4.4 Estimation of the Mutual information among the residues

MI is an information-theoretic quantity explaining the reduction in uncertainty of one random variable given the knowledge of the other. Since MI is a symmetrical measure, it cannot differentiate between possible directions or causalities. However, using MI allows us to clarify the full level of dependence above and beyond that shown by correlation. Second, it can help provide an indirect evaluation of models assuming a non-linear causal dependence as the primary interaction between the two variables. Here, the estimation of mutual information between every pair of 50 alpha carbons is done using the Kraskov method. In the following figure, these are shown:

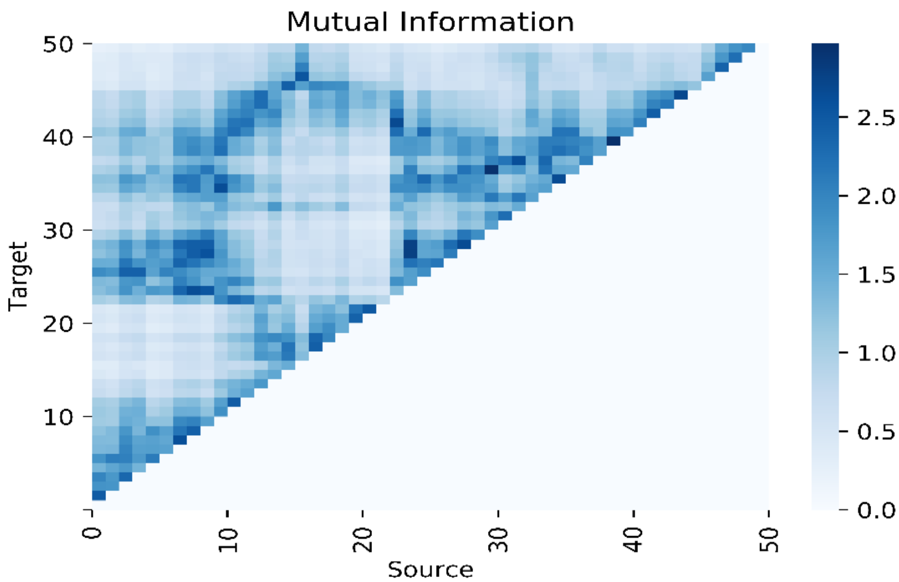

Figure 7:Matrix represents the residues samples of mutual information Between (1-50).

In Fig. 7, we observe certain dependency patterns showing the interactions between 50 residues in protein. For example, residues between 45 and 50 seem to be independent from residues 1-44. In this band, there is only one exception at residue 16 where the above independence is not valid. Moreover, residues between 12 and 22 are observed to be almost independent from residues 1 to 9, whereas those between 22 and 30 are highly dependent on those in the range of 1-10.

4.5 Analysis of Conditional TE among residues

In this section, we propose to analyze the interactions between the residues of the same Protein A4 (S100A4) by using conditional TE. As we see in the AR section, here we also observe additional information brought by conditioning other residues. This lets us identify which pairs have common effectors at the background.

In the following section, by considering each residue pair, R_i and R_j representing the source and target, the conditional TE values are illustrated for 50 variables. In each figure, conditional TE from a single source to all targets are estimated separately and for each of them a 50x50 matrix is obtained.

For example, in Fig. 8a, each column of the matrix shows the values of TE from source 1 to a specific target, given the residue number 2. Similarly, in Fig. 8b, each column is composed of TE values from source 2 to every target, conditioned on residue number 3. In the rest of the section, this procedure is repeated for each source with different conditions.

When we analyze the results, we note that conditioning on a third residue considerably changes the pairwise TE between many residue pairs. Most of the time, the conditional TE values are found to be smaller than their pairwise counterparts, meaning that we measure “redundant” information between the pairs if we do not take the conditions into account.

Physically, this means that even if certain residues seem to be closely related, when the allosteric effect of a third one is analyzed by conditioning, this close relationship can be found to be smaller than its pairwise value. This means, a third residue in space can provide more information about the interaction between pairs fluctuations.

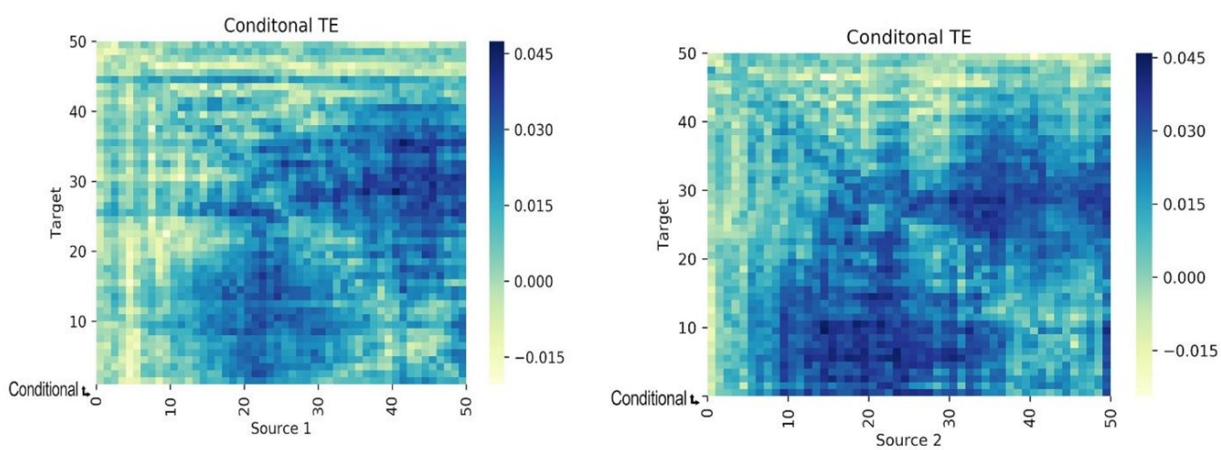

Figure 8: Conditional TE between one source to all destination.

The fig. 8 shows the conditional TE between source 1 and all targets conditioned on 2 to 50, in each column, respectively. We observe that the information flow is small between source 1 and all targets when conditioned on separate observations of residues 1 to 10 and it does not change much as a function of targets. On the other hand, information flow from source 1 to targets between 45 and 50 seem to be almost independent from the change in the conditioning residue. For the rest of the interactions, higher information flows are observed from source 1 to sources between 25 to 40 conditioned on 35 to 50 and from source 1 to targets 1 to 40 conditioned on residues from 20 to 30. Furthermore, when different conditionings to third variables are examined, we notice that mostly, under the 45 degree diagonal, where the source number is much smaller than the conditioned one, the TE is higher. This means that the locations of the residues with smaller number are affected more from the conditionings.

On the other hand, the residue sources with higher numbers become more connected to higher numbered targets for low numbered conditions. That is, the upper left of matrices tend to be higher in value, also meaning that we get more information when we condition on high numbered residues.

For example, for residue number 34, we see a huge change going down to minimum values when conditioned on the residues in the vicinity of 34, namely 32,35,etc.

When we analyze source number 6 conditioned on 45, TE to targets between 1 to 15 are affected considerably. This time, we can conclude that actually residue 6 and residue 45 are not that far away from each other in affecting targets between 1 to 15.

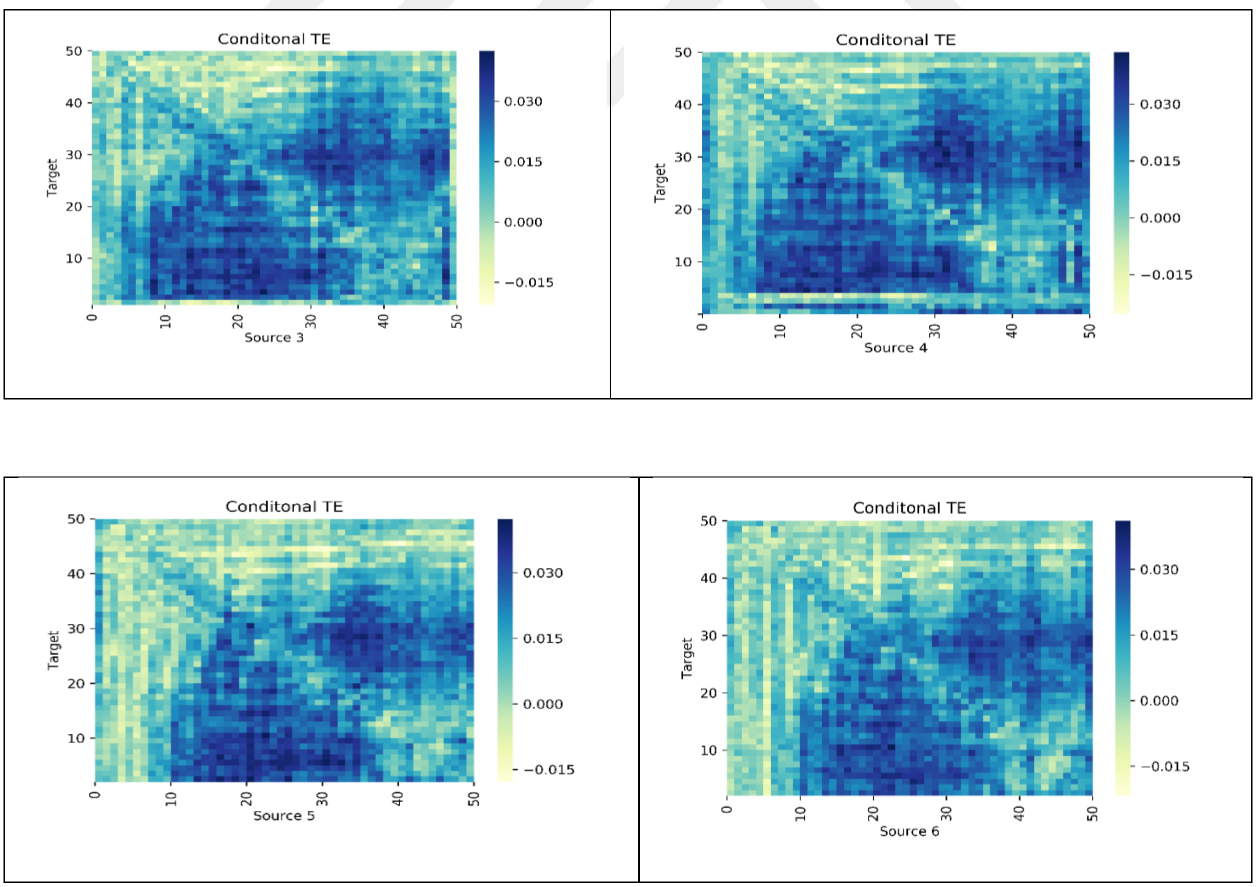

Figure 9:Sources between 3 to 6 and all target conditioned 2-50 in above simulation.

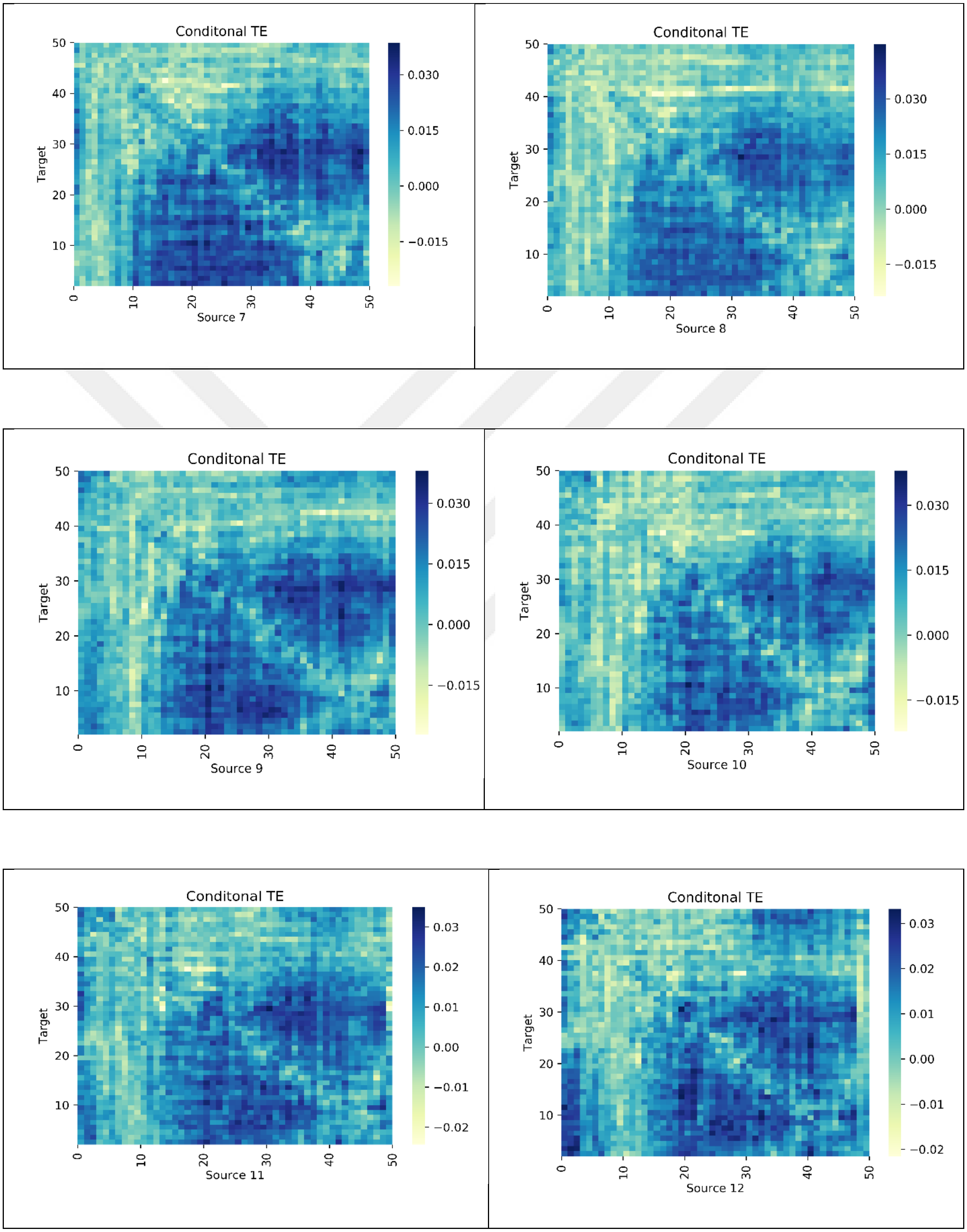

Figure 10: Sources between 7 to 12 and all target conditioned 2-50 in above simulation.

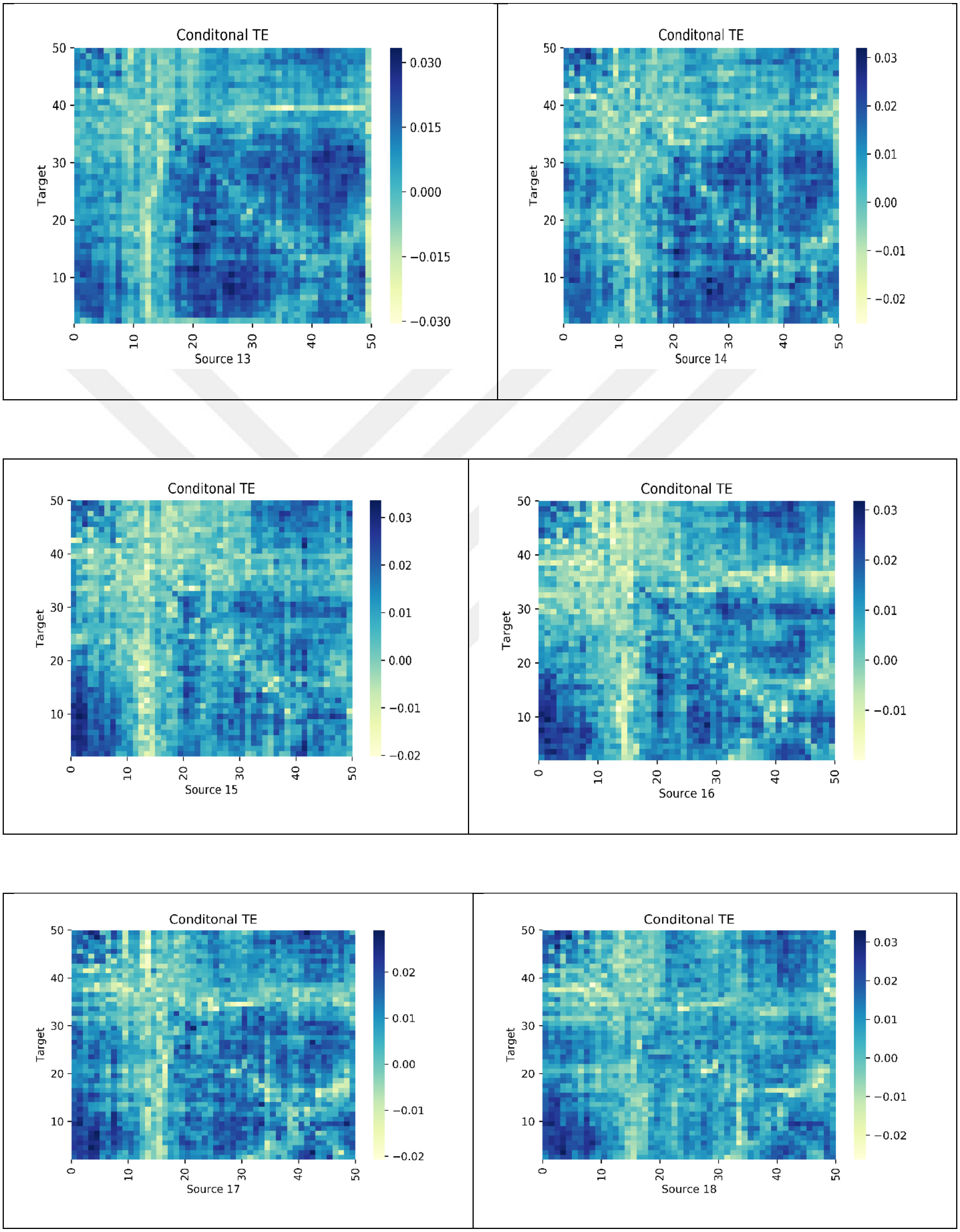

Figure 11:Sources between 12 to 18 and all target conditioned 2-50 in above simulation.

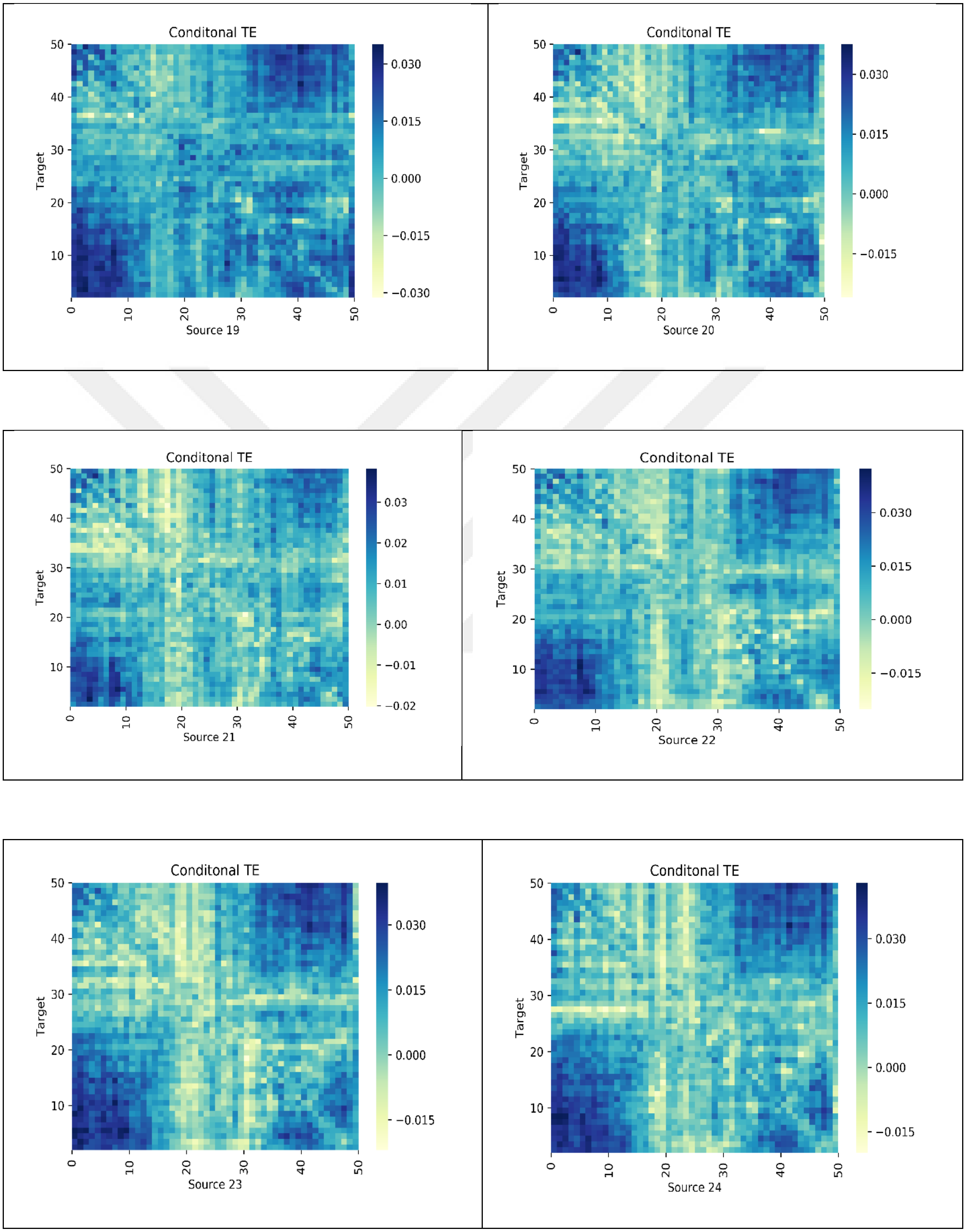

Figure 12: Sources between 18 to 24 and all target conditioned 2-50 in above simulation.

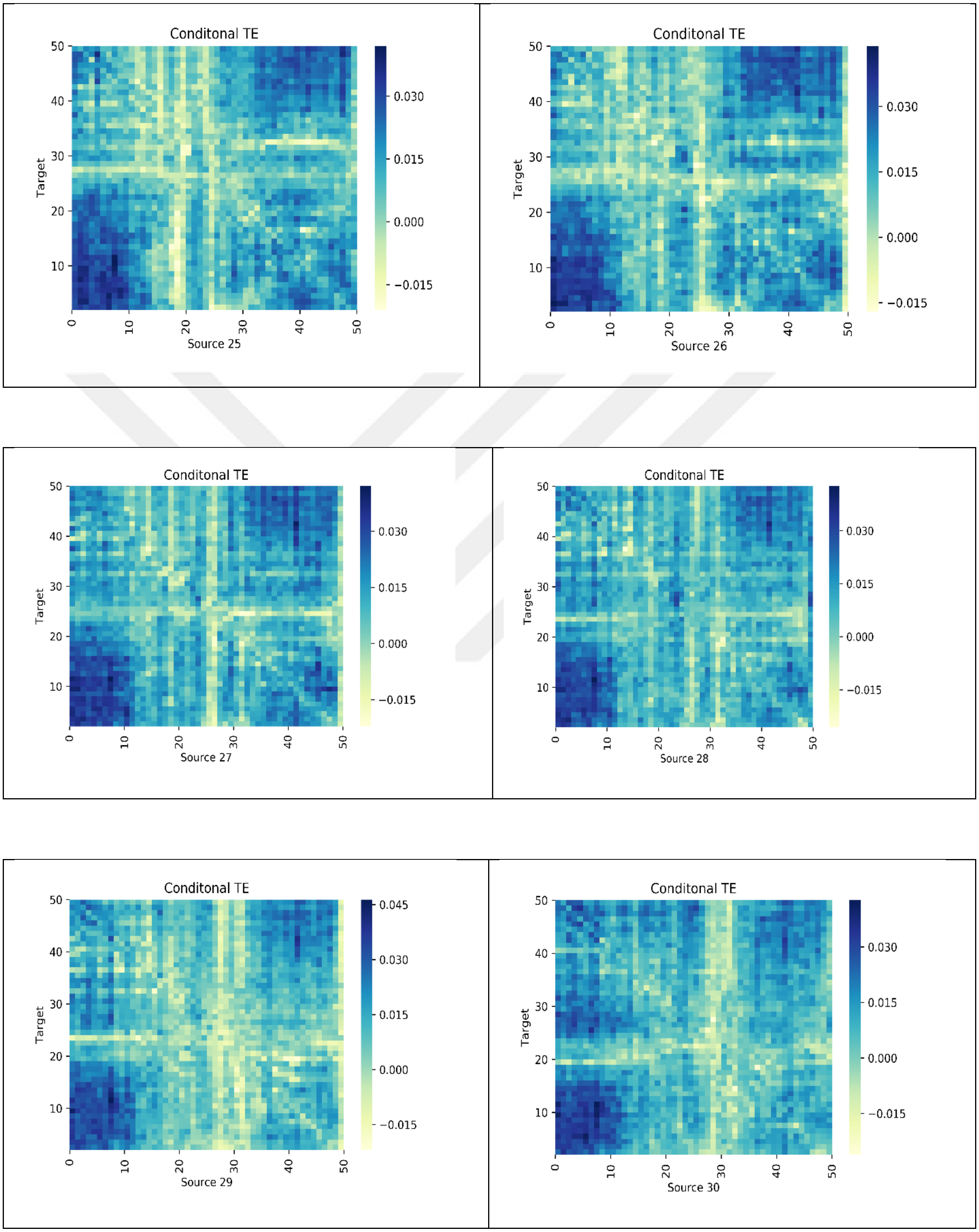

Figure 13: Sources between 25 to 30 and all target conditioned 2-50 in above simulation.

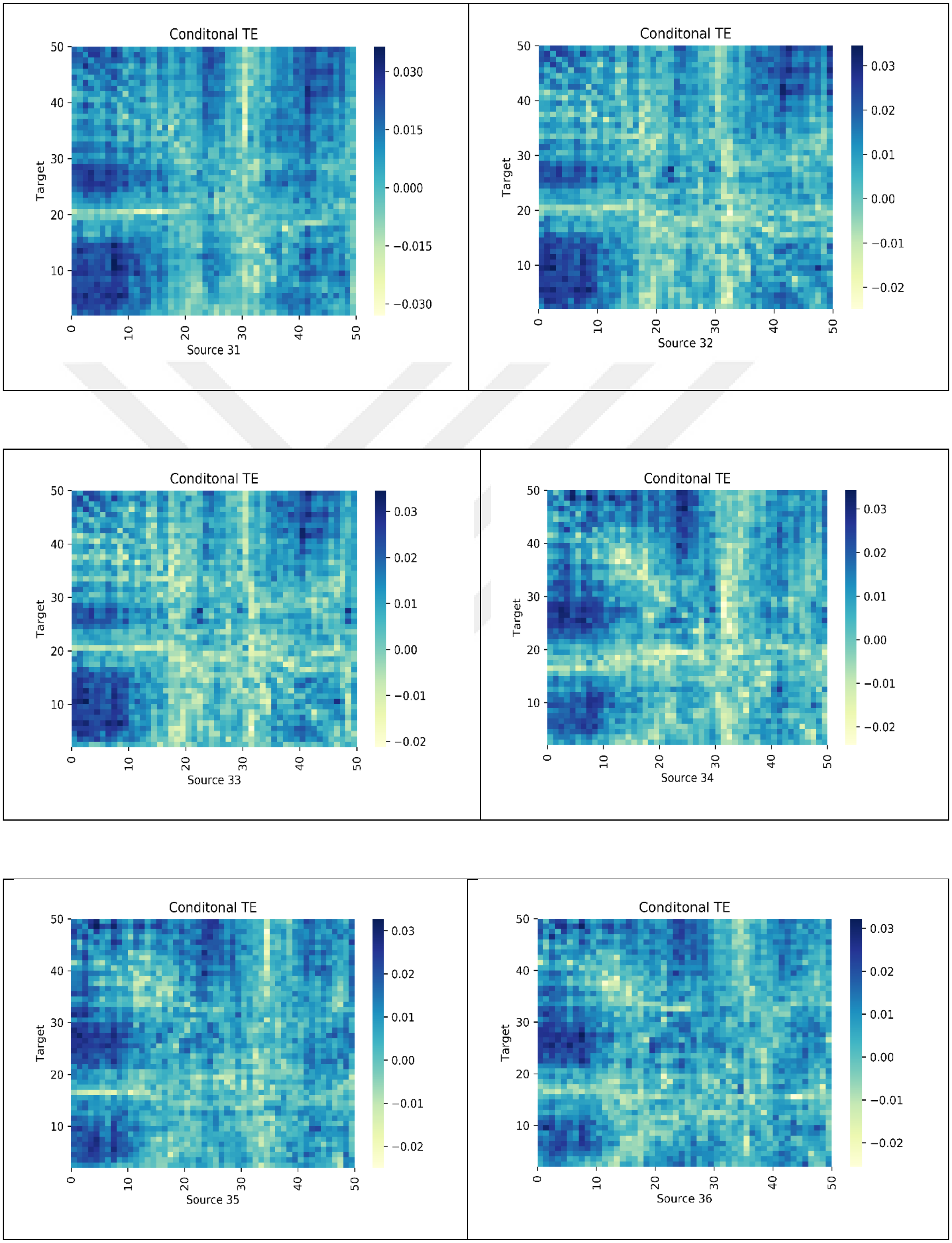

Figure 14:Sources between 31 to 36 and all target conditioned 2-50 in above simulation.

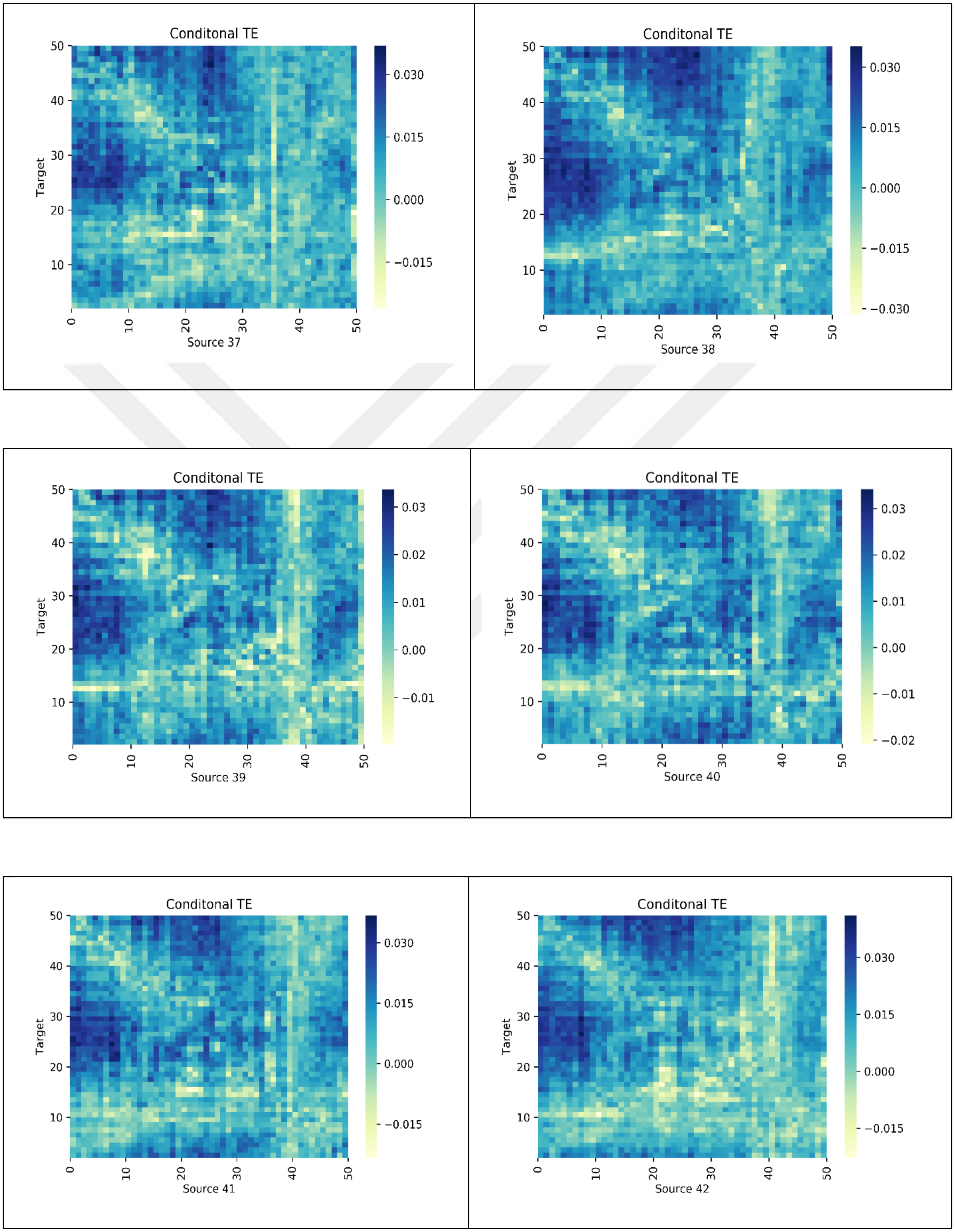

Figure 15:Sources between 37 to 42 and all target conditioned 2-50 in above simulation.

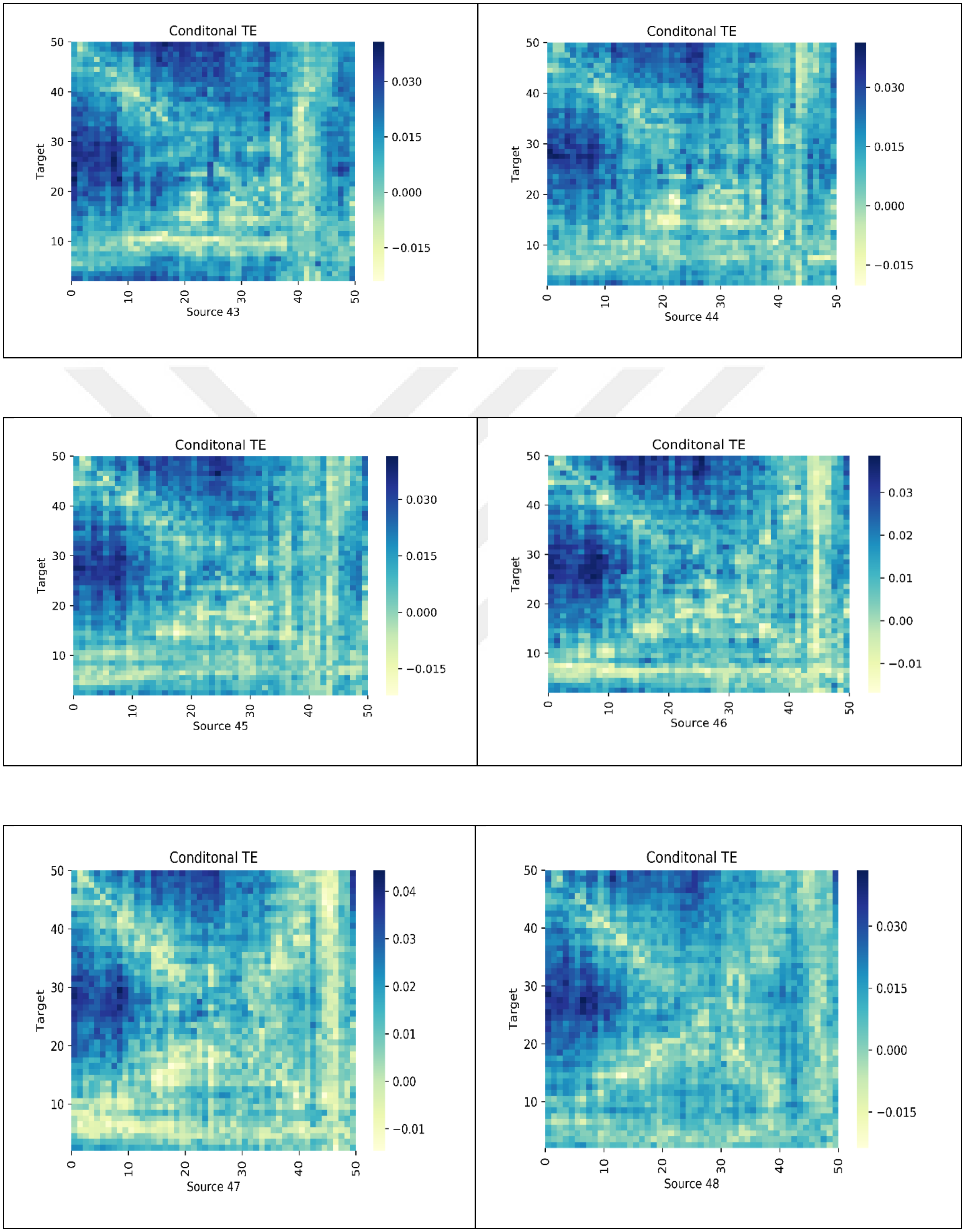

Figure 16: Sources between 43 to 48 and all target conditioned 2-50 in above simulation.

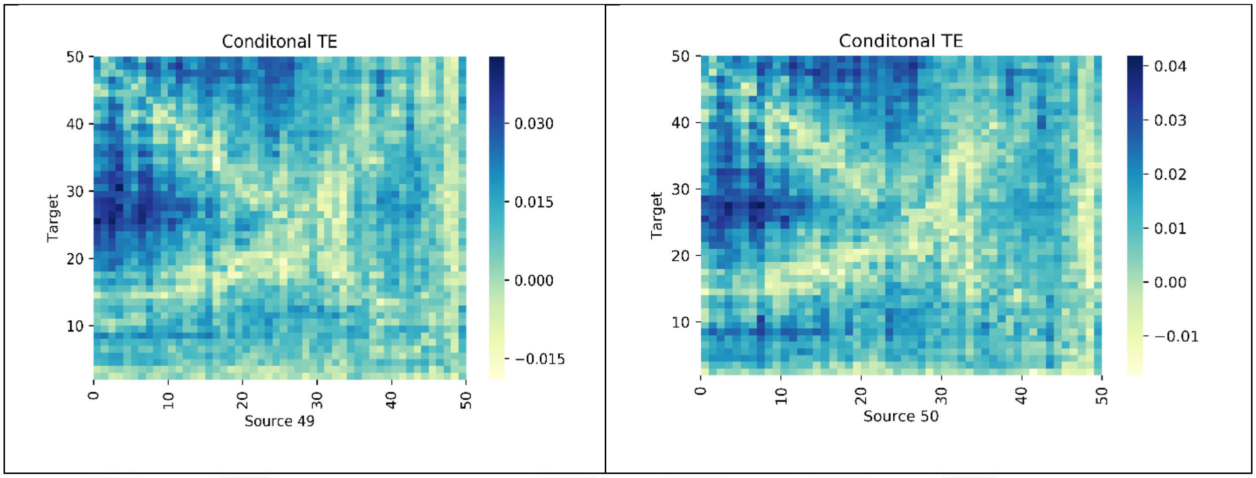

Figure 17: Sources between 49 to 50 and all target conditioned 2-50 in above simulation.

Based on these observations, where we see a significant pattern change between pairwise TE and conditional TE values, we can conclude that we can get much more insight on the hidden interactions among multiple random variables by taking conditional estimations into account. Above, we present that the information flow between any two residues can change significantly based on the conditions.

CHAPTER FIVE

5. Discussion and Future Work

In this thesis, comparison between different information-theoretic approaches are analyzed to better understand the allosteric interactions in proteins. In this perspective, we examined the behavior of MI, TE and conditional TE. In order to demonstrate the extra information that we can gain by conditioning on a third variable, we present two-way directional information flows between the elements of protein. For the proof of concept, we show the importance of conditioning in TE on an application where three autoregressive processes are coupled to each other. This simulation verifies that pairwise TE is not always ideal to understand the directional information flows and conditioning on other variables can significantly affect this interaction. In the simulations, we demonstrate this effect on Protein A4 (S100A4) data and show that when we use conditional TE, we can get extra insight of the spatial interactions in protein.

In conclusion, the use of conditional TE is highly encouraged for the directional analysis between random variables in many applications in science and engineering.

References

[1] N. M. Timme and C. C. Lapish, *A tutorial for information theory in neuroscience,* eNeuro, 5 (3): 2018. doi: 10.1523/ENEURO.0052-18.2018.

[2] C. Wang, H. Yu, R. W. Grout, K. Ma and J. H. Chen, "Analyzing information transfer in time-varying multivariate data," 2011 IEEE Pacific Visualization Symposium Hong Kong, 2011, pp. 99-106, doi: 10.1109/PACIFICVIS.2011.5742378.

[3] Lizier, J.T., Prokopenko, M., Zomaya, A.Y. 2010.Information modification and particle collisions in distributedcomputation. Chaos: An Interdisciplinary Journal of Nonlinear Science, 20 (03).

[4] Faes, L., Porta, A., Nollo, G. and Javorka, M. 2016. Information Decomposition in Multivariate Systems: Definitions, Implementation and Application to Cardiovascular Networks. Entropy, 19(01):5.

[5] Bertschinger, N., Rauh, J., Olbrich, E., Jost, J. and Ay, N. 2014. Quantifying unique information. Entropy, 16(04): 2161-2183.

[6] Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T. and Phillips, W. A. 2017. Partial information decomposition as a unified approach to the specification of neural goal functions. Brain and Cognition, 112:25-38.

[7] Barrett, A., B. 2015. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. Phys. Rev. E, 91.

[8] Williams, P.L. and Beer, R.D. 2010. Nonnegative Decomposition of Multivariate Information. Cognitive Science Program and School of Informatics and Computing Indiana University, Bloomington, Indiana 47406 USA.

[9] Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Causal information approach to partial conditioning in multivariate data sets. Comput. Math. Methods Med. **2012**, 2012, 303601.

[10] Joseph T. Lizier[1, 2], December 4, 2014. "JIDT: An information-theoretic toolkit for studying the dynamics of complex systems" [1]Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany 2CSIRO Digital Productivity Flagship, Marsfield, NSW 2122, Australia.

[11] Deniz Gencaga 1,*, Kevin H. Knuth 2 and William B. Rossow 1 "A Recipe for the Estimation of Information Flow in a Dynamical System", *Entropy* **2015**, *17*, 438-470; doi:10.3390/e17010438.

[12] http://staff.ustc.edu.cn/~cgong821/Wiley.Interscience.Elements.of.Information.Theory.Jul.2006.eBook-DDU.pdf.

[13] M.E. Pflieger, R.E. Greenblatt, "Using Conditional Mutual Information to Approximate Causality for Multivariate Physiological Time Series", Vol. 7, No. 2, 2005, Source Signal Imaging, Inc., San Diego, CA, USA.
[14] Ryan G. James,1, 2, ∗ Nix Barnett,1, 3, and James P. Crutchfield1, 2, 3, "Information Flows? A Critique of Transfer Entropies", Complexity Sciences Center 2Physics Department 3Mathematics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, June 20, 2016.
[15] Joseph T. Lizier? · Jakob Heinzle? · Annette Horstmann · John-Dylan Haynes · Mikhail Prokopenko, "Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity", August 12, 2010.
[16] Chih-Sheng Huang,[1,2] Nikhil R. Pal,[3] Chun-Hsiang Chuang,[1,4] and Chin-Teng Lin[1,2,4] "Identifying changes in EEG information transfer during drowsy driving by transfer entropy", Published online 2015 Oct 23. doi: 10.3389/fnhum.2015.00570.
[17] B. Gierlichs, L. Batina, P. Tuyls, B. Preneel, Mutual Information Analysis - A Generic Side-Channel Distinguisher, in the proceedings of CHES 2008, Lecture Notes in Computer Science, vol. 5154, pp. 426-442, Washington DC, USA, August 2008.
[18] http://folk.uio.no/feder/Fys3130/Entropy/Shannon_entropy.pdf.
[19] Robert M. Gray Information Systems Laboratory Electrical Engineering Department Stanford University, "Entropy and Information Theory First Edition, Corrected", c 1990 by Springer Verlag. Revised 2000, 2007, 2008, 2009, 2013 by Robert M. Gray.
[20] ANNICK LESNE (2014). Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics .Mathematical Structures in Computer Science, 24, e240311 doi:10.1017/S0960129512000783.
[21] Paul L. Williams1 and Randall D. Beer1, 2," Generalized Measures of Information Transfer",[1]Cognitive Science Program and [2]School of Informatics and Computing Indiana University, Bloomington, Indiana 47406 USA, October 22, 2018.
[22] Wollstadt P, Martı´nez-Zarzuela M, Vicente R, Dı´az-Pernas FJ, Wibral M (2014) Efficient Transfer Entropy Analysis of Non-Stationary Neural Time Series. PLoS ONE 9(7): e102833. doi:10.1371/journal.pone.0102833.
[23] Daniel Chicharro1, ∗ and Anders Ledberg2, 3 "A framework to study dynamic dependencies in networks of interacting processes", 1Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Via Bettini 31, 38068 Rovereto (TN) 2Center of Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain 3SoRAD, Stockholms Universitet, Stockholm, Sweden September 14, 2012.
[24] Hacisuleyman A, Erman B (2017) Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. PLoS Comput Biol 13(1): e1005319. doi:10.1371/journal.pcbi.1005319.
[25] joseph.lizier@gmail.com; This manuscript is a modified pre-print of: Joseph T. Lizier, "JIDT: An information-theoretic toolkit for studying the dynamics of complex systems", Frontiers in Robotics and AI 1:11, 2014; doi:10.3389/frobt.2014.00011.
[26] L. Barnett, A. B. Barrett, and A. K. Seth. \Granger causality and transfer entropy are equivalent for Gaussian variables". In: Physical review letters 103.23 (2009), p. 238701.
[27] Vicente, R., Wibral, M., Lindner, M. *et al.* Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* **30,** 45–67 (2011). https://doi.org/10.1007/s10827-010-0262-3.

[28] Lizier JT, Prokopenko M, Zomaya AY (2010) Information modification and particle collisions in distributed computation. Chaos 20: 037109–037109.
[29] Besserve M, Scholkopf B, Logothetis NK, Panzeri S (2010) Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. J Comput Neurosci 29: 547–566.
[30] J. Sukharev, C. Wang, K.-L. Ma, and A. T. Wittenberg. Correlation study of time-varying multivariate climate data sets. In Proc. IEEE Pacific Visualization, pages 161–168, 2009.
[31] Stramaglia, S.; Angelini, L.; Wu, G.; Cortes, J.; Faes, L.; Marinazzo, D. Synergetic and Redundant Information Flow Detected by Unnormalized Granger Causality: Application to Resting State fMRI. IEEE Trans. Biomed. Eng. **2016**, 63, 2518–2524.
[32] Quax, R.; Har-Shemesh, O.; Sloot, P.M.A. Quantifying Synergistic Information Using Intermediate Stochastic Variables. Entropy **2017**, 19, 85.
[33] Florin, E.; Gross, J.; Pfeifer, J.; Fink, G.R.; Timmermann, L. The effect of filtering on Granger causality based multivariate causality measures. Neuroimage **2010**, 50, 577–588.
[34] Lizier, J.T.: The local information dynamics of distributed computation in complex systems. Springer, Heidelberg (2013)
[35] Faes, L., Nollo, G., Porta, A.: Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. Phys. Rev. E 83, 051112 (2011).
[36] Porta, A., Castiglioni, P., Bari, V., Bassani, T., Marchi, A., Cividjian, A., Quintin, L., Di Rienzo, M.: K-nearest-neighbor conditional entropy approach for the assessment of the short-term complexity of cardiovascular control. Phys. Meas. 34, 17–33 (2013).
[37] Bashan, A., Bartsch, R.P., Kantelhardt, J.W., Havlin, S., Ivanov, P.C.: Network physiology reveals relations between network topology and physiological function. Nat. Communicat.3 (2012).
[38] Kraskov,A.,Stögbauer,H.,andGrassberger,P.(2004).Estimatingmutualinforma- tion. *Phys.Rev.E* 69, 066138.doi:10.1103/PhysRevE.69.066138.
[39] Bauer,T.L.,Colbaugh,R.,Glass,K.,andSchnizlein,D.(2013).“Useoftransfer entropy to infer relationships from behavior,”in Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop,CSIIRW’13 (New York,NY:ACM).doi:10.1145/2459976.2460016.
[40] Porta, A., Gnecchi-Ruscone, T., Tobaldini, E., Guzzetti, S., Furlan, R., Montano, N.: Progressive decrease of heart period variability entropy-based complexity during graded head-up tilt. J. Appl. Physiol. 103, 1143–1149 (2007).
[41] Faes, L., Nollo, G., Porta, A.: Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. Phys. Rev. E 83, 051112 (2011).
[42] Kugiumtzis, D.: Direct-coupling information measure from nonuniform embedding. Phys. Rev. E 87, 062918 (2013).
[43] Harmah DJ, Li C, Li F, Liao Y, Wang J, Ayedh WMA, Bore JC, Yao D, Dong W and Xu P (2020) Measuring the Non-linear Directed Information Flow in Schizophrenia by Multivariate Transfer Entropy. Front. Comput. Neurosci. 13:85. doi: 10.3389/fncom.2019.00085.
[44] T. S. Han. Multiple Mutual Informations and Multiple Interactions in Frequency Data. Information and Control, vol. 46(1), pp. 26-45 (1980).

APPENDIX

Appendix 1

The code of AR model is attached in the following figure.

```
        n_matrix = 0.2*xlsread('Protein_radiii_data2');
%         n_matrix= 0.2*randn(3,5000);
       n_matrix = n_matrix(1:3,:)

matrix1 = [a1, 0, 0; ...
          b3, b1, 0; ...
          0, c2, c1];

matrix2 = [0, 0, 0; ...
           0, b2, 0; ...
           c3, 0, 0];

y_t_1 = randn(3,1);  % t-1 values [1.5; 1.5; 1.5];
y_t_2 = randn(3,1); % t-2 values[1; 1; 1];

for ii = 1: size(n_matrix,2)

    Y(:,ii) = matrix1 * y_t_1 + matrix2 * y_t_2 + + n_matrix(:,ii);
    y_t_2 = y_t_1;
    y_t_1 = Y(:,ii);


end
```

Appendix 2

Simulations Parameters for AR Model.

Parameters	**Values**
a1	0.99
b1	0.8

b2	0.1
b3	0.1
c1	0.1
c2	0.3
c3	0.08

Appendix 3

All the codes used in this research can be obtained in the following link.

https://github.com/humairpalh/humairali