September 2015

Spatiotemporal Sensing and Informatics for Complex Systems Monitoring, Fault Identification and Root Cause Diagnostics

Gang Liu
University of South Florida, gliu@mail.usf.edu

Spatiotemporal Sensing and Informatics for Complex Systems Monitoring, Fault Identification and Root Cause Diagnostics

by

Gang Liu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management System Engineering
College of Engineering
University of South Florida

Major Professor: Hui Yang, Ph.D.
José Zayas-Castro, Ph.D.
Alex Savachkin, Ph.D.
Fabio M. Leonelli, M.D.
Ravi Sankar, Ph.D.

Date of Approval:
April 21, 2015

Keywords: High-dimensional Modeling, Model-driven Parametric Monitoring, Self-organized Clustering, High-redundant Variables

DEDICATION

To my loving parents, Qiwen Liu and Yuxia Xiang, who have been so supportive and considerate over the years.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to Dr. Hui Yang for his instantly inspiring guidance, hearty encouragement, valuable advices and constant support through the entire course of this study. Completing this dissertation would not have been possible without him. He is the best mentor and adviser that any graduate student could hope for. I appreciate him so much for providing me this great opportunity to work in the Complex Systems Monitoring, Modeling and Analysis Lab and introducing me to this exciting research field.

Secondly, I am extremely thankful to the members of my committee, Dr. José Zayas-Castro, Dr. Alex Savachkin, Dr. Ravi Sankar and Fabio M. Leonelli, M.D. Deep thanks go to them for the constructive suggestions they proposed for my dissertation and the valuable advices that will help my career development in the future.

Thirdly, my deep gratitude also goes to Gloria Hanshaw, Elizabeth Conrad, Catherine Burton and Norma Paz for their consistent help and support in my graduate study. Thanks also go to colleagues in our laboratory and my friends who give me sincere suggestions on my research work and help me with my study and life.

I gratefully acknowledge the financial support from NSF (CMMI-1454012, CMMI-1266331, IIP-1447289 and IOS-1146882).

Finally, I am heavily indebted to my parents who have been firmly supporting me and encouraging me from the behind. I would like to express my greatest appreciation and deepest love to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In order to cope with system complexity and dynamic environments, modern industries are investing in a variety of sensor networks and data acquisition systems to increase information visibility. Multi-sensor systems bring the proliferation of high-dimensional functional Big Data that capture rich information on the evolving dynamics of natural and engineered processes. With spatially and temporally dense data readily available, there is an urgent need to develop advanced methodologies and associated tools that will enable and assist (i) the handling of the big data communicated by the contemporary complex systems, (ii) the extraction and identification of pertinent knowledge about the environmental and operational dynamics driving these systems, and (iii) the exploitation of the acquired knowledge for more enhanced design, analysis, monitoring, diagnostics and control.

My methodological and theoretical research as well as a considerable portion of my applied and collaborative work in this dissertation aims at addressing high-dimensional functional big data communicated by the systems. An innovative contribution of my work is the establishment of a series of systematic methodologies to investigate the complex system informatics including multi-dimensional modeling, feature extraction and selection, model-based monitoring and root cause diagnostics.

This study presents systematic methodologies to investigate spatiotemporal informatics of complex systems from multi-dimensional modeling and feature extraction to model-driven monitoring, fault identification and root cause diagnostics. In particular, we developed a

multiscale adaptive basis function model to represent and characterize the high-dimensional nonlinear functional profiles, thereby reducing the large amount of data to a parsimonious set of variables (i.e., model parameters) while preserving the information. Furthermore, the complex interdependence structure among variables is identified by a novel self-organizing network algorithm, in which the homogeneous variables are clustered into sub-network communities. Then we minimize the redundancy of variables in each cluster and integrate the new set of clustered variables with predictive models to identify a sparse set of sensitive variables for process monitoring and fault diagnostics. We evaluated and validated our methodologies using real-world case studies that extract parameters from representation models of vectorcardiogram (VCG) signals for the diagnosis of myocardial infarctions. The proposed systematic methodologies are generally applicable for modeling, monitoring and diagnosis in many disciplines that involve a large number of highly-redundant variables extracted from the big data.

The self-organizing approach was also innovatively developed to derive the steady geometric structure of a network from the recurrence-based adjacency matrix. As such, novel network-theoretic measures can be achieved based on actual node-to-node distances in the self-organized network topology.

CHAPTER 1: INTRODUCTION AND BACKGROUND

Rapid advancement of sensing and information technology brings the spatially and temporally big data. Facing the massive data and spatiotemporal signals (usually nonlinear and nonstationary), there is a pressing need to develop advanced methodologies that will enable and assist (i) the modeling and characterization of the big data communicated by the complex systems, (ii) the extraction and identification of pertinent knowledge about the environmental and operational dynamics driving these systems, and (iii) the exploitation of the acquired knowledge for more enhanced monitoring, diagnosis, and control of them. However, addressing this need is very challenging because of a collection of factors, which include the inherent complexity of the system and the uncertainty associated with the system's operation and its environment, the high dimensionality of the data communicated by the system, and the increasing expectations and requirements posed by real-time decision-making.

The objective of this dissertation is to develop a series of systematic methods to investigate the complex system informatics with respect to high-dimensional modeling, pertinent knowledge discovery, system diagnostics, performance monitoring and control.

1.1 Motivation and Needs

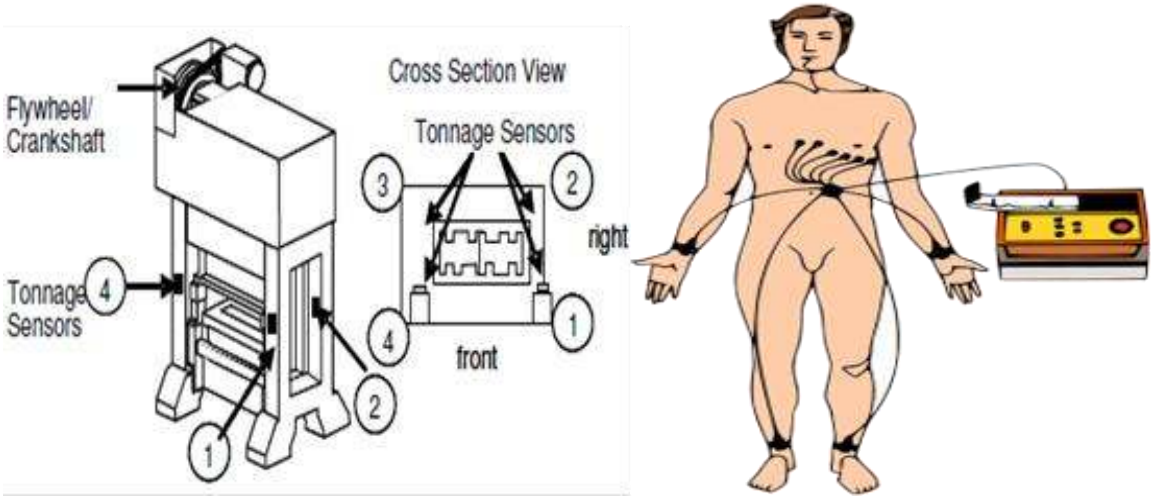
With the rapid advances in sensor technology and data acquisition systems, multichannel sensing is capable of providing us with comprehensive information about complex systems. The all-round sensing and high-powered computing lead to a spatially and temporally data-rich environment. Facing the massive data and spatiotemporal signals (usually nonlinear and

nonstationary), there are urgent needs to efficiently handle big data and systematically investigate the complex systems, including signal representation and characterization, informative feature selection, performance monitoring and root cause analysis. An effective data dimensionality reduction is critical to handle massive data. A sparse representation model with significant accuracy characterizes both spatial and temporal behaviors of spatiotemporal systems. The characteristics of multichannel sensory signals are depictured by the features extracted from the sparse representation model. An effective feature selection method optimally selects informative features for optimal decision-making. An integrative monitoring for spatiotemporal system identifies the change points of the complex system and the out-of-control states. The root-cause analysis detects the physical or environmental variation that leads to the out-of-control states. A series of systematic methods are in need to address these issues.

1.2 Multichannel Sensing of Spatiotemporal Dynamics

A multichannel sensing, with sensors distributed at different positions on the complex system, can facilitate the all-round data collection and comprehensive investigation of complex system. Multichannel sensing is widely used in meteorology, oceanography, geology, physiology and manufacturing systems. For example, the sensing system of forging process in Fig. 1a contains four strain gage sensors located spatially to measure the forces at different parts on the machine. The sensing signals provide rich information about the product quality and process condition. Fig. 1b shows a patient connected to the 10 electrodes necessary for a 12-lead Electrocardiography (ECG) system, which observes the heart from 12 different angles. It is widely used for the diagnosis of cardiovascular diseases in the clinical practice. The differences among the characteristics of the signals from a variety of sensors show the spatial diversity of the system. The changes along the time in each signal from one sensor explain the temporal

variation. Only the integrative study of both spatial and temporal behaviors will discover the underlying system mechanisms.



(a) Forging Process (b) 12-lead ECG

Figure 1 Examples of spatiotemporal sensing.

1.3 Research Background and Challenges

The research on spatiotemporal sensing and complex system informatics are challenging due to a variety of environmental and operational dynamics including the nonlinear and nonstationary systems, high-dimensional representation, complex interdependence structure among the extracted features.

1.3.1 Nonlinear and Nonstationary Systems

The previous methods that were used in various fields of data analysis simply assumed the linearity and stationarity of the underlying processes. It is now extensively accepted that the complex physical systems often have nonlinear and nonstationary properties as follows:

1) Nonlinearity: In mathematics, a nonlinear system is one that does not satisfy the superposition principle, or one whose output is not directly proportional to its input. Most

physical systems are inherently nonlinear in nature, which makes the systems different to represent and analyze. This is because the parameter estimation and model inference are somewhat more involved for nonlinear systems than they are in the linear cases. For example, when the model of least squares is applied to a nonlinear system, the resulting normal equations are nonlinear and often difficult to solve [1]. The usual approach is to directly minimize the residual sum of squares by an iterative procedure. Furthermore, the normal-theory inference used in the linear system does not apply exactly to nonlinear systems. Instead, inference based on asymptotic or large-sample theory must be employed to address this issue.

2) Nonstationarity: Besides the nonlinearity, the nonstationarity is another challenge in data analysis because the nonstationarity of the complex systems means that the system varies from time periods to time periods. The typical statistical methods require stationarity. For example, in the theory of statistics expectation values are defined through ensemble averages and for a stationary process the ergodic theorem enables one to replace an ensemble average by a time average [2]. However, in a nonstationary process, the average of two time scales fails to find the mean value and bring the loss of precise meanings. To explicitly represent and characterize the complex systems with both nonlinear and nonstationary properties is a challenging and necessary.

1.3.2 Sparse Representation of Spatiotemporal Profiles

In various scientific fields, the analysis of spatiotemporal patterns emerging from complex systems plays an important role. An investigation of multidimensional data allows us to learn more about the internal dynamics of the system. An effective representation model of the spatiotemporal signals is a general and crucial issue in the investigation. Depending on the

intended use, in the previous researches the spatiotemporal signals were transformed to other domains for analysis, i.e., time domain, frequency domain, time-frequency domain and state space domain. This transformation often brings the information loss, i.e., the projection of spatiotemporal signals onto 1-dimensional time domain will diminish the spatial information. The representations using these methods often suffered from the incompactness because they are not adaptive. The representation methods are not designed smartly with respect to the characteristics of the spatiotemporal signals. A systematic method for sparse and effective representation of spatiotemporal signals is not presented in the literature.

The previous methods have the drawback in the following aspects: (i) The representations are not compact; (ii) The representations are not adaptive and (iii) The methods transform the spatiotemporal signals into a specific domain and may cause important information loss.

1.3.3 Feature Redundancy and Irrelevancy

The principle of parsimony requires the simplest models with a satisfied accuracy. In the analysis of a complex system, a variety of features can be extracted from the system itself or descriptive models or analytical models. The extracted feature sets are often suffered from the redundant features and some of the features are irrelevancy with respect to system performance.

1) Feature Redundancy: Redundant features are those which provide no more information than the currently selected features. Fig. 2a shows the redundancy between two features. It is noted that the two features are highly correlated. Therefore, one of them should be excluded from the mathematical model that characterizes the system.

2) Feature Irrelevancy: Irrelevant features provide no useful information in any context, i.e., describing the system performance, classifying two categories or predicting the

output. Fig. 2b shows the irrelevant feature with respect to the model performance. It is seen from the figures that the change of input2 does not affect the target value.

Thus in the descriptive models or the analytical models, input2 should be excluded to satisfy the principle of parsimony. In the literature, the maximum relevancy and minimum redundancy method is proposed. However, a systematic method to efficiently eliminate the feature redundancy and irrelevancy for multichannel signals at both sensor level and individual feature level is not found.

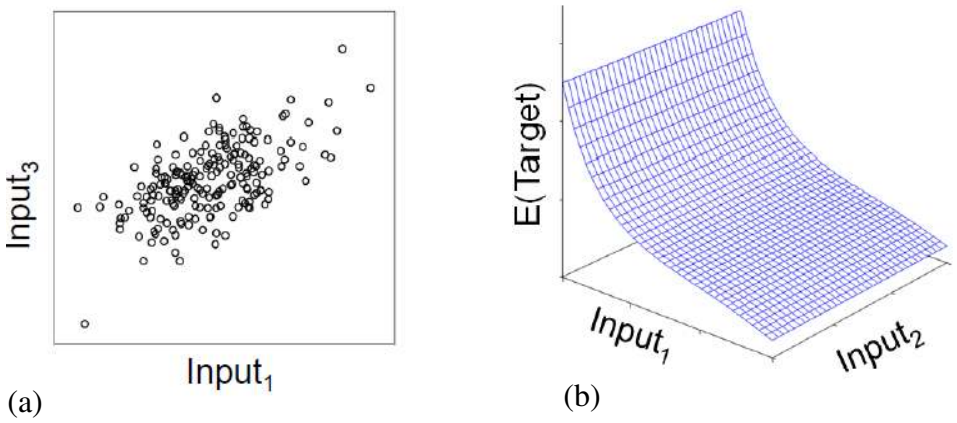


Figure 2 Examples of (a) feature redundancy and (b) feature irrelevancy.

1.3.4 Feature Selection

The motivation of informative feature selection is the feature redundancy and irrelevancy as described in section 1.3.3. The feature selection in the literature was often investigated using regression and classification models, because the incorrect inclusion of unimportant variables in the model may seriously affect the prediction accuracy of the models. Moreover, the importance of a feature can be reflected by the magnitude of the coefficient for it. From the viewpoint of regression and classification models, a parsimonious one is preferred and thus the number of

importance features is small. A general regression model with the group variable structure is presented as follows:

$$y_i = \beta_0 + \sum_{k=1}^{K} \sum_{j=1}^{p_k} \beta_{kj} x_{i,kj} + \varepsilon_i \quad (1)$$

where y_i denotes the response variable, $x_{i,kj}$ is predictor j ($j = 1, \dots, p_k$), belonging to group k ($k = 1, \dots, K$), β_0 and β_{kj} are the corresponding model parameters, ε_i is the random error and index i ($i = 1, \dots, n$) indicates the sample number. The general regularization method for estimation the parameters of model (1) can be written as:

$$\min_{\beta_0, \beta_{kj}} a \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{k=1}^{K} \sum_{j=1}^{p_k} \beta_{kj} x_{i,kj})^2 + bJ(\beta_{kj}) \quad (2)$$

where $J(\beta_{kj})$ is penalty function, a and b are the tuning parameters. a is the significance assigned to the model performance on representing the response variable and b is the significance assigned to the penalty function for adjustment.

In the literature, different types of penalty function and values of a and b were used to generate different models. For example, Yuan and Lin [3] extended the regular lasso and developed the group lasso by using an L2-norm penalty in the form of $J(\beta_{kj}) = \sum_{k=1}^{K} \|\beta_k\|$ with the $\|\beta_k\| = \sqrt{\sum_{j=1}^{p_k} \beta_{kj}^2}$. Zhao et al. [4] proposed an alternative method using an L∞-norm penalty, i.e., $J(\beta_{kj}) = \sum_{k=1}^{K} \|\beta_k\|_\infty$ with $\|\beta_k\|_\infty = \max_j\{|\beta_{kj}|\}$. The unimportant feature group can be removed from the model by choosing appropriate tuning parameters. Yuan and Lin [5] developed group non-negative garrote (GNNG) method for group variable selection. In their model, the predictors are transformed such that $\sum_{i=1}^{n} x_{i,kj} = 0$ and $\sum_{i=1}^{n} x_{i,kj}^2 = 1$ and the response variable is centered such that $\sum_{i=1}^{n} y_i = 0$. When the non-negative shrinking factor is d_k introduced for group k, the coefficients $\beta_{kj} = d_k \widehat{\beta_{kj}^{ols}}$, where $\widehat{\beta_{kj}^{ols}}$ is the ordinary least square estimate of β_{kj}. In the GNNG model, the following penalized least square criterion is used:

$$\min_{d_k} a \left\| \boldsymbol{y} - \sum_{k=1}^{K} \boldsymbol{X}_k \widehat{\boldsymbol{\beta}_k^{ods}} d_k \right\|^2 + b \sum_{k=1}^{K} d_k, \quad s.t. \quad d_k \geq 0, k = 1,2,\dots,K \tag{3}$$

where y is the vector of the observed response variable, X_k is the matrix of the predictor variables for group k, $\widehat{\beta_k^{ods}}$ is the vector of estimated coefficients using the ordinary least square method corresponding to group k, a and b are tuning parameters.

The above models were developed to select the important groups. Another important task is to select the important features at individual level in the selected groups, because the unimportant features in the selected group may affect both prediction performances and model parsimony if all the features in the selected groups are used. To address this issue, Huang et al. [6] and Zhou and Zhu [7] proposed the group bridge and hierarchical lasso by using the following penalty: $\sum_{k=1}^{K} \|\beta_k\| 1$ with $\|\beta_k\|_1 = \sqrt{\sum_{j=1}^{p_k} |\beta_{kj}|}$. Paynabar et al. [8] proposed the hierarchical non-negative garrote (HNNG) based on the GNNG to first select the important groups and then select the significant individual features. In the group level, the method is the same as GNNG in equation (3). In the individual level, the following criterion is defined:

$$\min_{\boldsymbol{d}} a \left\| \boldsymbol{y} - \widetilde{\boldsymbol{X}} (\widehat{\boldsymbol{\beta}^r} \cdot \boldsymbol{d}) \right\|^2 + b_1 \|\boldsymbol{d}\|_1 + b_2 \|\boldsymbol{d}\|^2, \quad s.t. \quad \boldsymbol{d} \geq \boldsymbol{0} \tag{4}$$

where $\widehat{\beta^r}$ is the vector of ridge estimates calculated by $(\widetilde{X}^T \widetilde{X} + I b_2)^{-1} \widetilde{X}^T y$, $\widetilde{X}$ is the matrix of all predictors whose groups are identified as important in the group level, d is the vector of shrinking factors of individual features and the operator $\cdot$ represents the element-wise vector product. The vector of coefficients of all individual features β can be calculated by $\beta = (\widehat{\beta^r} \cdot d)$. Paynabar compared his methods with the previous methods by extensive study and show the priority of his method.

In the regression models, it often happens that individual features are highly correlated both within and between groups. If two features are highly correlated, their corresponding coefficients in the regression model should tend to be the same (up to a sign change if negatively correlated). This is called similar effects in [8]. Some methods have been proposed to work on this issue. For example, Zou and Hastie [9] proposed the elastic net with a combined L1- and L2-norm penalty function $J(\beta_{kj}) = \sum_{k=1}^{K}\sum_{j=1}^{p_k}|\beta_{kj}| + \sum_{k=1}^{K}\sum_{j=1}^{p_k}\beta_{kj}^2$ and showed that the elastic net benefits from the similar effects property. However, their method does not consider the group feature selection. HNNG considers both group level and individual level selection and gains the similar effect property. However, according to the discussion in feature redundancy in section 1.3.5, an effective and parsimony model should not have two high correlated features. The high correlated features should have one representative to achieve model sparsity. It is worth mentioning that most of existing variable selection methods focus on the relevancy between predictors and response variables. Interdependence structures among predictors are often overlooked, or not explicitly investigated. The interdependence structures among variables are investigated by variable clustering.

1.3.5 Variable Clustering

The motivation of feature clustering is to estimate the feature redundancy as discussed in the above section. The feature clustering is different from the traditional data clustering. The typical data clustering methods and their probabilistic mixture models consider how close the entities are (e.g. in terms of Euclidean distance). Fig. 3a is an example of data clustering, in which the working systems of air-conditioner are clustered according to the scaled temperature and scaled pressure. As shown in Fig. 3a, one point represents a working system and the coordinates of one point are the real values of scale temperature and scaled pressure for a that

working system. The similar working systems are clustered into one group considering the Euclidean distances among this data points. However, the feature clustering algorithm, on the other hand, clusters the correlated features into one group. Fig. 3b is an example of feature clustering, in which the features extracted from a multiscale adaptive basis function model of spatiotemporal vector cardiogram signals are clustered. As shown in Fig. 3b, one point represents a feature which changes with time. Or rather each point represents a time series of the feature. The feature clustering algorithm clusters the feature considers how dependent they are, either linearly or nonlinearly.

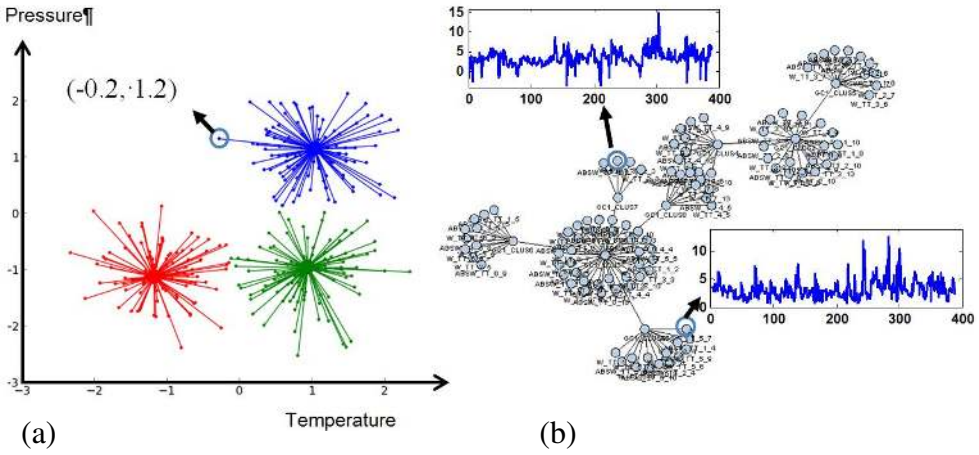


Figure 3 Examples of (a) data clustering and (b) feature clustering

Variable clustering depends on similarity measurements between variables such as linear correlation or mutual information. Nonetheless, linear correlation cannot capture nonlinear interdependences among variables. Mutual information characterizes linear and nonlinear correlation, but requires the stationarity assumption. Note that latent-variable methods are also commonly used for variable clustering, e.g., oblique principal component clustering. However, both oblique rotation and principal component analysis are based on linear projections of variables. As such, nonlinear interdependences among variables are not fully considered. Also,

variable clustering mainly focuses on the redundancy among variables while neglecting the relevancy between predictors and the response. New methodologies that integrate variable clustering with variable selection to improve effectiveness and efficiency of predictive analytics are urgently needed. In this dissertation, we developed a new methodology of self-organizing network that leverages merits of variable clustering and variable selection to investigate both relevancy and redundancy structures among variables for improving the performance of predictive modeling.

1.4 Research Objectives

My research is motivated by the urgent needs to efficiently handle big data and extract useful information for system informatics investigation including complex system modeling, diagnostics, monitoring and root cause analysis. The detailed objectives are described as following four:

1) Representation and Characterization of Multichannel Spatiotemporal Signals: This work aims: i) to develop a systematic method for multi-channel spatiotemporal signal representation using multiscale adaptive basis function modeling, ii) to characterize the temporal behaviors of spatiotemporal signal, iii) to investigate the practical issues in multiscale adaptive basis function modeling, i. e., optimal basis selection, model complexity vs. model performance, model robustness.

2) Model-driven Parametric Monitoring of High-dimensional Nonlinear Functional Profiles: This study aims: i) to develop a sparse basis function model of high-dimensional profiles, thereby reducing the large amount of data to a parsimonious set of model parameters while preserving the information; ii) to select a low-dimensional set of sensitive predictors for fault diagnostics.

3) Self-organizing Network for Variable Clustering and Predictive Analytics: This study aims: i) to develop a new approach of nonlinear coupling analysis to measure nonlinear interdependence structures among variables; ii) to develop a self-organizing approach to identify the interdependence structures among a large number of variables; iii) to leverage the advantages of variable clustering and variable selection to investigate both redundancy and relevancy among variables for improving the performance of predictive analytics and diagnosis.

4) Self-organized Recurrence Networks: The self-organizing network approach is extended to the recurrence analysis. In this study, I will derive the steady geometric structure of a network from the recurrence-based adjacency matrix.

1.5 Research Contributions

The project is creative and original, because: (1) Spatiotemporal signal representation: The previous representation methods for spatiotemporal signal representation have the drawbacks of information loss (because the spatiotemporal signal is transform to other domains) and incompactness (because the representation is not adaptive). The proposed research will develop an innovative multiscale adaptive basis function model of spatiotemporal signals, which is new and solves the drawbacks of the previous methods. (2) Pertinent knowledge/feature discovery: The previous informative feature selection physically choose the feature groups (i.e., from one sensor), ignoring the feature redundancy. The proposed research introduces a smart feature clustering algorithm to informative feature selection for estimating the feature redundancy. And the theoretically-justified tuning parameter selection in the mathematical model shows the priority over the previous empirical selection. (3) Model-based real-time monitoring: The traditional monitoring focused on the mean shift or significant variation changes in the

process. However, the monitoring on a complex system should be comprehensive to detect the detailed changes in either spatial or temporal behaviors. The proposed model-based monitoring bridges this gap by examining the changes not limited to mean shift and variation change. (4) Root cause analysis: The previous studies on root cause analysis are all static models. For a complex system with varying environmental or system's operational factors, a dynamical causal inference combined with system performance monitoring and optimal decision-making is innovative and challenging.

The project has the potential to broadly impact the society because it will yield a fundamental understanding of complex system informatics with respect to system modeling, data compression, informative feature selection, performance monitoring, diagnostics and control. When the representation model is used to medical records, the highly reduced data dimensionality will facilitate the store of huge medical data. When the model-based real-time monitoring system assists the smart healthy system, a lower death rate is expected. The patient specific monitoring will improve the personal medicine and in-time medical care. Therefore the expense on the healthcare would be reduced for in-time treatment. When the methods are introduced to the manufacturing systems, the informative selection of sensors will reduce the expense because the unimportant sensory signals are not necessary. Meanwhile, the monitoring system and the root cause analysis will high improve the system reliability and the quality of products. A variety of areas with the complex system will receive the benefits of the proposed systematic methodologies and its associated tools.

1.6 Dissertation Organization

The dissertation is organized is organized as following: Chapter 2 presents multiscale adaptive basis function modeling of spatiotemporal vectorcardiogram signals, Chapter 3

introduces a model-driven parametric monitoring of high-dimensional nonlinear functional profiles, Chapter 4 presents a self-organizing network for variable clustering and predictive analytics, Chapter 5 shows a self-organized recurrence network and Chapter 6 derives the conclusions arising out of this dissertation.

1.7 References

[1] G. K. Smyth, "Nonlinear Regression," Encyclopedia of Environmetrics, vol. 3, pp. 1405-1411, 2002.

[2] H. Kantz and T. Schreiber, Eds., Nonlinear Time Series Analysis. New York: Cambridge University Press, 2004.

[3] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variable," J. R. Statist. Soc. B, vol. 68, pp. 49-67, 2006.

[4] P. Zhao, G. Rocha and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," The Ann. of Stat., vol. 37, pp. 3468-3497, 2009.

[5] M. Yuan and Y. Lin, "On the nonnegative garrote estimate," J. R. Statist. Soc. B, vol. 69, pp. 143-161, 2007.

[6] J. Huang, S. Ma, H. Xie and C. Zhang, "A group bridge approach for variable selection," Biometrika, vol. 96, pp. 339-355, 2009.

[7] N. Zhou and J. Zhu, "Group variable selection via a hierarchical lasso and its oracle property," Tech. Rep. University of Michigan, Dept. of Statistics, 357-358.

[8] K. Paynabar, "Nonlinear Profile data Analysis for System Performance Improvement," 2012.

[9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Statist. Soc. B, vol. 67, pp. 301-320, 2005.

CHAPTER 2: MULTISCALE ADAPTIVE BASIS FUNCTION MODELING OF SPATIOTEMPORAL SIGNALS[1]

Mathematical modeling of cardiac electrical signals facilitates the simulation of realistic cardiac electrical behaviors, the evaluation of algorithms, and the characterization of underlying space-time patterns. However, there are practical issues pertinent to model efficacy, robustness, and generality. This paper presents a multiscale adaptive basis function modeling approach to characterize not only temporal but also spatial behaviors of vectorcardiogram (VCG) signals. Model parameters are adaptively estimated by the "best matching" projections of VCG characteristic waves onto a dictionary of nonlinear basis functions. The model performance is experimentally evaluated with respect to the number of basis functions, different types of basis function (i.e., Gaussian, Mexican hat, customized wavelet and Hermitian wavelets), and various cardiac conditions, including 80 healthy controls and different myocardial infarctions (i.e., 89 inferior, 77 anterior-septal, 56 inferior-lateral, 47 anterior, 43 anterior-lateral). Multi-way Analysis of Variance shows that the basis function and the model complexity have significant effects on model performances while cardiac conditions are not significant. The customized wavelet is found to be an optimal basis function for the modeling of space-time VCG signals. The comparison of QT intervals shows small relative errors (<5%) between model representations and real-world VCG signals when the model complexity is greater than 10. The proposed model shows great potentials to model space-time cardiac pathological behaviors, and

[1] This chapter was published in IEEE Journal of Biomedical and Health Informatics [32]. Permission is included in Appendix A.

can lead to potential benefits in feature extraction, data compression, algorithm evaluation and disease prognostics.

2.1 Introduction and Motivation

Cardiac electrical signals, e.g., electrocardiogram (ECG), are widely used for the diagnosis of cardiovascular diseases in the clinical practice. The surface ECG is usually obtained by recording the potential difference between electrodes placed on the skin surface. The most commonly used clinical ECG system, i.e., 12-lead ECG, monitors the underlying cardiac electrical activities from 12 different measurement angles. In 1956, Ernst Frank designed a new vectorcardiogram (VCG) lead system to capture cardiac electrical activities in the form of spatial vectors. The 3-lead VCG is observed along three orthogonal X, Y, Z planes of the body (i.e., frontal, transverse, and sagittal) and shows cardiac electrical activities in space and time [1].

VCG signals are not as broadly used in the medical practice as 12-lead ECG because, among other reasons, the interpretation of 3-dimensional VCG is not commonly taught and requires specialized knowledge of space-time decomposition of cardiac vectors. Dower et al. [2] and our previous investigation [3] showed that 12-lead ECG is derivable from 3-lead VCG with a linear transform matrix. Although this transform matrix may need to be personalized, such a linear transform shows the preservation of clinically useful information between 12-lead ECG and 3-lead VCG. However, when it comes to the development of computer algorithms, 12-lead ECG has higher dimensionality than 3-lead VCG and potentially introduces the “curse of dimensionality” problem. Hence, the 3-lead VCG surmounts not only the information loss from only one or two ECG signals but also the dimensionality problems induced by all the 12-lead ECG signals. Therefore the 3-lead VCG signals are adopted in this investigation of spatiotemporal nonlinear profiles.

Traditionally, most previous investigations focused on discovering cardiac pathological behaviors from the ECG signals. A large amount of algorithms were developed to delineate the ECG patterns pertinent to cardiac disorders in time domain [4], frequency domain [5], time-frequency domain [6] or state space domain [7]. In contrast, this paper presents a mathematical model (i.e., multiscale adaptive basis function model) for the representation of real-world spatiotemporal VCG signals. Such mathematical models is applicable to simulate the VCG signals from a specific cardiovascular disease, to evaluate biomedical signal processing algorithms, to characterize space-time patterns of cardiac electrical signals and to study specific regions of interest pertaining to cardiac pathological behaviors.

However, there are four practical issues pertinent to the construction of VCG basis function models. (1) What kind of function form should be used? There exist different types of basis functions including polynomials, Hermite function [8], Fourier bases [9], Gaussian function [10] and wavelet basis functions [7]. The selection of function form will directly impact the model parameters needed and fitting performances. (2) How to adaptively estimate model parameters based on the morphology of VCG signals? It is well known that cardiac electrical activity is initiated at the sinoatrial (SA) node, conducted in both atria, and then relayed through the atrioventricular (AV) node to further propagate through bundle of His and Purkinje fibers toward ventricular depolarization and repolarization. Such a sequential function provides an advantage to adaptively estimate the model parameters based on the patterns of VCG characteristic waves. (3) How many parameters should be involved in the model? The model complexity[2] depends on the number of basis functions involved. Based on the principles of parsimony, a desirable modeling strategy is to choose simpler models with sufficient explanatory

[2] *Model complexity refers to the number of unknown parameters involved in the model. Equivalently, it is pertinent to the number of basis functions and parameters in each basis function.*

power (i.e., simpler models are generally better than complex ones when the modeling performances are equal,). (4) How does the model performance (i.e., goodness of fit) vary with respect to uncertainty factors? It may also be noted that real-world ECG signals often show vastly different patterns because of cardiac conditions, human subject differences and other distortion factors. A robust model should yield the persistence of modeling characteristic behaviors in real-world ECG signals under perturbations or other uncertainty conditions.

The objective of this paper is to develop a multiscale adaptive basis function model of spatiotemporal VCG signals. We will experimentally investigate the four aforementioned practical issues for the efficacy, robustness, and generality in the mathematical modeling of real-world VCG signals. The remainder of this paper is organized as follows: Section II presents literature review of previous modeling studies, the developed research methodology is presented in Section III, Section IV contains the ECG database description and experimental design, Section V presents the experimental results, and Section VI includes the discussion and conclusions arising out of this investigation.

2.2 Research Background

Many recent investigations have focused on the mathematical modeling of ECG signals [11-15]. It may be noted that PhysioNet [16, 17] provides open-source software, i.e., ECGSYN to generate the realistic ECG waveform. This software is based on the ECG dynamical model firstly proposed by McSharry and Clifford et al. [11]. This dynamical model is composed of three ordinary differential equations as follows:

$$\begin{aligned} \dot{x} &= \alpha x - \omega y \\ \dot{y} &= \alpha y + \omega x \\ \dot{z} &= -\sum_{i\epsilon\{P,Q,R,S,T\}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0) \end{aligned} \tag{1}$$

where $\alpha = 1 - \sqrt{x^2 + y^2}$, $\Delta\theta_i = (\theta - \theta_i) \bmod 2\pi$, $\theta = \text{atan2}(y, x)$, and ω is the angular velocity of the trajectory. This model simulate the 1-dimensional ECG signals $z(t)$ around the limit cycle formed by x and y. The baseline wander $z_0(t)$ is simulated as a sinusoid function of the respiratory frequency f_2, $z_0(t) = A\sin(2\pi f_2 t)$ and $A = 0.15\text{mV}$.

Furthermore, Clifford et al. [12, 13] showed that the T wave is often asymmetrical and needs two Gaussian functions to correctly model this asymmetry. Hence, they added an extra parameter to the T feature, thereby adopting the convention of 6 events as $i \epsilon \{P, Q, R, S, T^-, T^+\}$. The nonlinear least-square optimization was also employed to fit the model with real-world ECG signals, but an eighteen-dimensional gradient descent search is required to solve this optimization problem. Sayadi et al. [15] proposed the wave-based dynamical model of ECG temporal dynamics and used separate state variables for each characteristic waves, namely, the P wave, QRS complex and the T wave. The model has 7 events, i.e., $i \epsilon \{P^-, P^+, Q, R, S, T^-, T^+\}$ and two additional Gaussian waves are used to deal with the asymmetry in P waves. Sameni et al. [14] extended the McSharry's model to 3- dimensional VCG signals together with an autoregressive (AR) model simulating the ECG noises. Sameni et al. also generalize their model to analyze maternal and fetal ECG mixtures recorded from the abdomen of pregnant women in single and multiple pregnancies.

It may be noted that Gaussian function form is unanimously adopted in the previous investigations. The four aforementioned practical issues have not been fully addressed for the mathematical modeling of ECG signals. It may also be noted that most of previous investigations focused on modeling the temporal dynamics of ECG signals. This presented paper is aimed at developing the basis function models of both temporal and spatial cardiac electrical dynamics. Given N time points in $\Re 1$: t1, ..., tN, and their corresponding target values of cardiac vectors in

$\Re d$: v_1, ..., v_N, we are seeking a function :$\Re 1 \rightarrow \Re d$, so that $\forall 1 \leq i \leq N$, $\vec{v}(t,w) = \vec{w}_0 + \sum_{j=1}^{M} \vec{w}_j \vec{\varphi}_j\left((t-\mu_j)/\sigma_j\right) + \varepsilon$, where $\varphi(t)$ is the general basis function form. Spatiotemporal model will not only help us gain a deeper understanding of underlying heart dynamics but also characterize space-time ECG patterns pertinent to pathological cardiac behaviors. The main contributions of this present paper and its differences from previous studies are as follows:

1) Selection of basis function – We experimentally compare model performances with different types of basis function $\varphi(t)$ including Gaussian function, Mexican hat wavelet, customized wavelet and other wavelet basis functions. This facilitates the identification of an optimal basis function $\varphi(t)$ for the mathematical modeling of ECG signals. In addition, this paper presents the customized design of wavelet basis function for a specific patient.

2) Adaptive model parameter estimation – Matching pursuit is utilized to adaptively model the ECG signals based on the "best matching" projections of ECG characteristic waves onto the dictionary of nonlinear basis functions.

3) Balance between model complexity and model performance – It may be noted that most previous investigations proposed either 5, 6, 7 or 11 Gaussian functions to deal with asymmetrical events and real-world ECG variations. We conduct further experiments to compare the variations of model performances (i.e., goodness of fit) when the number of basis functions is varied from 1 to 20.

4) Model robustness to cardiac conditions – The characteristics of ECG signals heavily depend on the subjects as well as cardiac conditions. Even in the same category of cardiac condition, e.g., healthy control, human subjects can generate vastly different ECG patterns. The robustness of basis function model is statistically tested under six cardiac

conditions, i.e., healthy control, MI inferior, MI inferior-lateral, MI anterior, MI anterior-lateral, and MI anterior-septal.

2.3 Research Methodology

Cardiac electrical dynamics are initiated and propagated spatiotemporally. Such spatiotemporal activities are traditionally projected onto 1-dimensional time domain for the identification of cardiac pathological patterns. This kind of projection diminishes important spatial information underlying cardiac electrical activities. Hence, some medical decisions that are made can be significantly influenced by such an information loss. This present study develops new mathematical models of spatiotemporal VCG signals that can characterize specific regions of interest pertinent to space-time cardiac pathological behaviors.

2.3.1 Spatiotemporal Characterization

One-dimensional ECG signals capture the projected temporal view of space-time cardiac electrical activities from a particular measurement angle. Multiple lead ECG systems, for e.g., 12-lead ECG and 3-lead vectorcardiogram (VCG), are designed to obtain the multi-directional view and spatial information. Considering the cardiac electrical activity varying across both space and time, we denote the cardiac process as $\{V(s,t): s(t) \in R \subset \mathbb{R}^d, t \in T\}$, where the dependence of spatial domain R on time T symbolizes the condition where the spatial domain changes over time. It may be noted that traditional ECG algorithms focus heavily on the temporal domain signal analysis and tend to have limitations to capture the space and time correlations. The presented study takes advantage of the increasing availability of spatiotemporal cardiac signals, and model both spatial and temporal variations of cardiac electric events.

As shown in Figure 1 (a), we utilized the Poincaré section to characterize the spatiotemporal cardiac topology in the form of ensembles as well as detailed beat-to-beat

variations. Poincaré section is a dE-1 dimensional hyperplane intersecting with the dE dimensional state space trajectories (see Figure 1 (a)). The 3D display of cardiac vector loops from several cycles commonly show the near-periodic patterns but with hidden temporal variations between heart cycles. The aligned ECG ensembles contribute to the gathering of homogeneous majority clusters of ECG signals for selective averaging. For near-periodic signals, each strand emanating from an intersection point Pi of Poincaré section and lasting approximately till the next intersection Pi+1 along the trajectory may be treated as a realization of a stochastic process from an invariant probability space [7]. Heart rate variability makes some ensembles move faster, i.e., the two successive intersections occur over shorter intervals, compared to the others. Figure 4 (b-d) shows the aligned heart beats along the X, Y, Z axes. It may be noted that heart cycles in the same direction, i.e., any of three orthogonal directions X, Y, or Z, are sharing similar morphologies but there exist remarkable beat-to-beat variations due to the heart rate variability.

It has long been understood that a metronomic heart rate is pathological, and healthy heart is influenced by multiple neural and hormonal inputs that result in variations in interbeat (RR) intervals [18]. There were evidences showing that the underlying heart beat dynamics may have a multi-fractal temporal structure, which has complex interdependencies at different scales. Random cascade model was used for modeling the multifractal self-similar behaviors in the heart beat dynamics [18]. Heart rate time series r(t) is taken as a product of J cascade components: $r_J(t) = \prod_{j=1}^{J} \omega_j(t)$ and $\omega_j(t) = 1 + \xi_j$, where ξ_j's are independent Gaussian variables with $\langle \xi_j \rangle = 0$ and $\langle \xi_i \xi_j \rangle = \delta_{ij}\sigma_j^2$ (δ_{ij} is the Kronecker delta). In addition, $\omega_j(t)$ is only varied on discrete times $\left\{t_k^{(j)}\right\}: \omega_j(t) = \omega_j(t_k^{(j)})$ for $t_k^{(j)} \le t \le t_{k+1}^{(j)}$, where $t_k^{(j)} = \frac{kN}{2^j}, k = 1, \cdots, 2^j$ and $N = 2^j$ is the total number of the data points. The power law is simulated as

$\sigma_j = \sigma_0 2^{-\alpha(j-1)} \sim \tau_j^{\alpha}$, where $\tau_j = t_k^{(j)} - t_{k-1}^{(j)} = N/2^j$. Hence, RR interval time series is modeled as a multifractal self-similar process. The mathematical model of VCG morphology within each RR interval will be detailed in the next section.

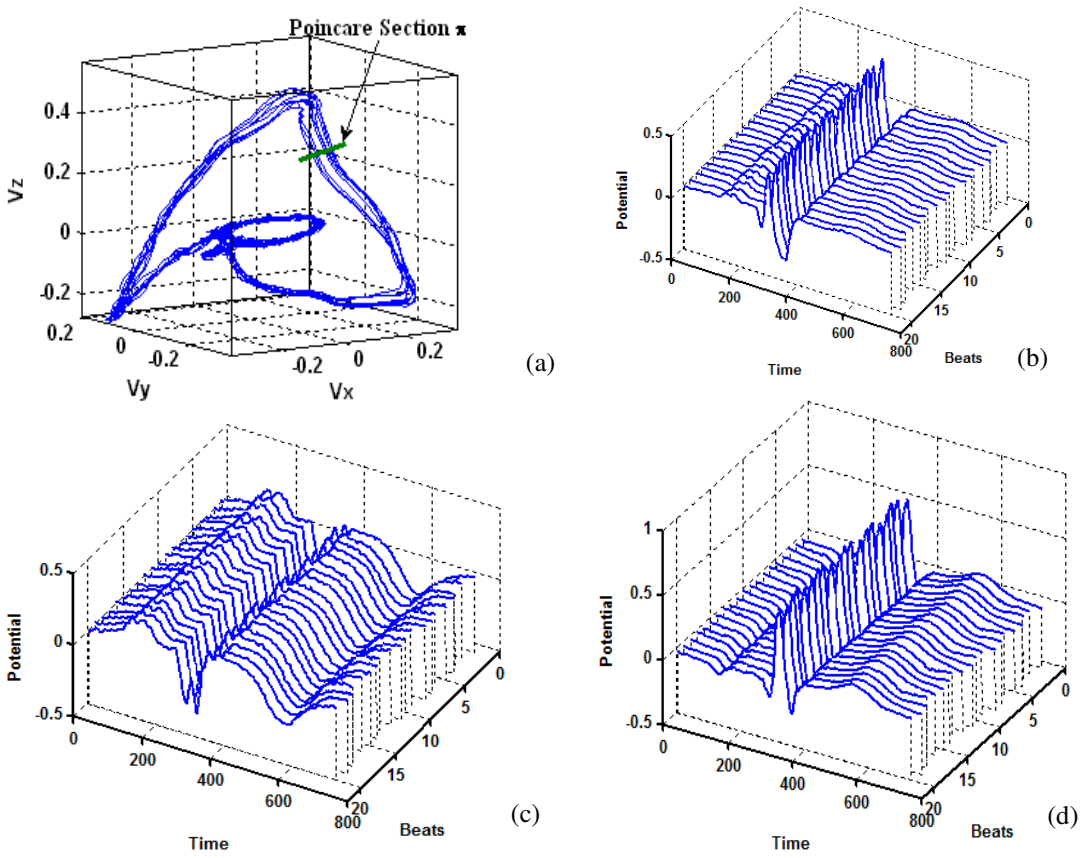


Figure 4 Poincaré sectioning of 3D VCG trajectory (a), and one-dimensional projected VCG ensembles along X axis (b), Y axis (c) and Z axis (d) [2]

2.3.2 Multiscale Basis Function Modeling

On the basis of McSharry's dynamical ECG model [11], we propose to model the 3-dimensional VCG morphology as the superposition of M multiscale basis functions: $\vec{v}(t,w) = \vec{w}_0 + \sum_{j=1}^{M} \vec{w}_j \vec{\varphi}_j\left((t-\mu_j)/\sigma_j\right) + \varepsilon$, where $\varphi(t)$ is the general basis function form that is not limited to Gaussian function, μ_j is the shifting factor and σ_j is the scaling factor. The objective is to minimize the representation error between VCG signals and basis function models

as $\operatorname{argmin}\left[\left\|\vec{v}(t)-\overrightarrow{w}_0-\sum_{j=1}^{M}\overrightarrow{w}_j\overrightarrow{\varphi}_j(t)\right\|^2,\{w,M,\varphi(t)\}\right]$. In the matrix form, this basis function model can be rewritten as $V=W^T\varphi$, where the corresponding weight matrix W and basis function matrix φ are

$$\boldsymbol{V}=\begin{bmatrix} v_{x1} & v_{x2} & \cdots & v_{xN} \\ v_{y1} & v_{y2} & \cdots & v_{yN} \\ v_{z1} & v_{z2} & \cdots & v_{zN} \end{bmatrix} \tag{2}$$

$$\boldsymbol{W}=\begin{bmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ \vdots & \vdots & \vdots \\ w_{1M} & w_{2M} & w_{3M} \end{bmatrix} \tag{3}$$

$$\boldsymbol{\varphi}=\begin{bmatrix} \varphi_1(t_1|\mu_1,\sigma_1) & \cdots & \varphi_1(t_N|\mu_1,\sigma_1) \\ \varphi_2(t_1|\mu_2,\sigma_2) & \cdots & \varphi_2(t_N|\mu_2,\sigma_2) \\ \vdots & \vdots & \vdots \\ \varphi_M(t_1|\mu_M,\sigma_M) & \cdots & \varphi_M(t_N|\mu_M,\sigma_M) \end{bmatrix} \tag{4}$$

This is analogous to express the spatiotemporal VCG signals $\vec{v}(t,w)$ using a combination of words (i.e., $\varphi_j(t)$) from the dictionary D. For a selected basis function $\varphi(t)$, the dictionary D is composed by shifting and scaling the basis function $D=\{\varphi_j((t-\mu_j)/\sigma_j),\ j=1,2,\ldots,N\}$. Mathematically, shifting a basis function $\varphi(t)$ by μ is equivalent to delay its onset and is represented by $\varphi(t-\mu)$. Scaling a basis function $\varphi(t)$ means either "stretching" or "shrinking" the function by the scale factor σ and is represented by $\varphi(t/\sigma)$. The smaller the scale factor is, the more "compressed" the wavelets are. The higher scales correspond to the most "stretched" basis function. The more stretched the basis function is, the longer the portion of the signal is modeled. In addition, if the dictionary D contains a set of orthonormal bases, e.g., orthogonal wavelet functions, then the signal representation will be sparser. In other words, the coefficient set $\{\overrightarrow{w}_j\}$ will has lower entropy and can be simply computed by the inner products of the basis function with the signal $<\varphi(t),v(t)>$. It is desirable that such a signal representation yields the least model complexity but higher accuracy.

In order to optimize the sparsity of representation, it requires a combinatorial search of M possible basis functions from the dictionary D to find the global minimum. As this problem is potentially intractable for a moderate size of dictionary, we have utilized an iterative procedure, i.e., Matching Pursuit algorithms [19, 20] to search the sub-optimal solution based on the characteristic wave patterns in the VCG/ECG signals. The VCG Matching Pursuit method is started from an initial approximation $s^{(0)} = 0$, residual $R^{(0)} = \vec{v}(t)$, and dictionary $D = \{\varphi_j(t),\ j = 1,2,\dots,N\}$. The first step identifies the basis function in the dictionary that best correlates with the residual, that is, finding $\gamma 0$ such that $\left|\langle R^{(0)}, \varphi^{(\gamma 0)}\rangle\right| = \max\left|\langle R^{(0)}, \varphi^{(\gamma)}\rangle\right|$, $\gamma \in \mathbb{N}$ and $\varphi^{(\gamma 0)} \in D$. Then the current approximation will be $s^{(1)} = s^{(0)} + \langle R^{(0)}, \varphi^{(\gamma 0)}\rangle\varphi^{(\gamma 0)}$ and the residual is defined as $R^{(1)} = R^{(0)} - \langle R^{(0)}, \varphi^{(\gamma 0)}\rangle\varphi^{(\gamma 0)}$. If the orthogonal wavelet bases are used, it may be noted that $\varphi^{(\gamma 0)}$ is orthogonal to $R^{(1)}$ because

$$\begin{aligned}\langle \boldsymbol{\varphi}^{(\gamma 0)}, \boldsymbol{R}^{(1)}\rangle &= \langle \boldsymbol{\varphi}^{(\gamma 0)}, \boldsymbol{R}^{(0)} - \langle \boldsymbol{R}^{(0)}, \boldsymbol{\varphi}^{(\gamma 0)}\rangle\boldsymbol{\varphi}^{(\gamma 0)}\rangle \\ &= \langle \boldsymbol{\varphi}^{(\gamma 0)}, \boldsymbol{R}^{(0)}\rangle - \langle \boldsymbol{\varphi}^{(\gamma 0)}, \langle \boldsymbol{R}^{(0)}, \boldsymbol{\varphi}^{(\gamma 0)}\rangle\boldsymbol{\varphi}^{(\gamma 0)}\rangle \\ &= \langle \boldsymbol{\varphi}^{(\gamma 0)}, \boldsymbol{R}^{(0)}\rangle - \langle \boldsymbol{R}^{(0)}, \boldsymbol{\varphi}^{(\gamma 0)}\rangle = 0\end{aligned}$$

Hence, $\langle R^{(0)}, \varphi^{(\gamma 0)}\rangle\varphi^{(\gamma 0)}$ is also orthogonal to $R^{(1)}$ so that

$$\left\|\boldsymbol{R}^{(0)}\right\|^2 = \left\|\boldsymbol{R}^{(1)}\right\|^2 + \left\|\langle \boldsymbol{R}^{(0)}, \boldsymbol{\varphi}^{(\gamma 0)}\rangle\boldsymbol{\varphi}^{(\gamma 0)}\right\|^2 \tag{5}$$

At step j+1, the residual $R^{(j+1)}$ is treated similarly as $R^{(0)}$ in the first step, yielding

$$\begin{aligned}\boldsymbol{R}^{(j+1)} &= \boldsymbol{R}^{(j)} - \langle \boldsymbol{R}^{(j)}, \boldsymbol{\varphi}^{(\gamma j)}\rangle\boldsymbol{\varphi}^{(\gamma j)} \text{ and } \boldsymbol{s}^{(j+1)} \\ &= \sum_{\boldsymbol{i=1}}^{\boldsymbol{j}} \langle \boldsymbol{R}^{(i)}, \boldsymbol{\varphi}^{(\gamma i)}\rangle\boldsymbol{\varphi}^{(\gamma i)}\end{aligned} \tag{6}$$

After M such steps, one has a representation of the form of additive decomposition.

$$\boldsymbol{v}(t) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{M-1}} \langle \boldsymbol{R}^{(i)}, \boldsymbol{\varphi}^{(\gamma i)}\rangle\boldsymbol{\varphi}^{(\gamma i)} + \boldsymbol{R}^{(M)} \tag{7}$$

An intrinsic feature of the matching pursuit algorithm is that when the dictionary has orthogonal bases, it works perfectly after a few steps yielding a sparse adaptive representation using only a few basis functions. Section II.C will detail the design of customized basis function

$\varphi(t)$ for ECG application, and the performance comparison with many standard wavelet basis functions will be shown in the section of results.

2.3.3 Customized Wavelet Basis Function Design

As mentioned in the section of Background, many previous ECG modeling algorithms unanimously selected the non-orthogonal Gaussian basis function $\varphi(t) = \exp(-\frac{\|t-\mu\|}{2\sigma^2})$. The shifting factor $\mu_j's$ are either adaptively distributed at the locations of ECG characteristic PQRST waves or uniformly distributed along the time axis. The table of parameters (i.e., weights, locations, and widths) was given based on the visual analysis of a segment of normal ECG signals [11]. These previous approaches show satisfactory results for the healthy ECG signals with apparent Gaussian shape PQRST waves. Furthermore, Clifford and Sameni [12-14] proposed to utilize the nonlinear least-square error (NLSE) method and radial basis neural network for the estimation of parameter $\mu_j's$ and $\sigma_j's$ using real-world ECG signals. However, ECG signals are often varied due to many different factors, e.g., cardiac conditions, human subjects, and locations of recording sensors. For example, P wave may be biphasic, or wavy in the case of atrial fibrillation. Ventricular fibrillation will totally recompose the Gaussian shape T wave into non-Gaussian forms. ECG characteristic waves are not necessarily shown to be bell-shaped waves along some measurement directions, e.g., frontal, transverse, and sagittal directions in 3-lead VCG. Although the sum of Gaussian functions is capable of fitting any curve shape, this requires an increasing number of Gaussian functions, thereby potentially leading to a more complex model.

In this paper, we hypothesized that wavelet functions, e.g., Mexican hat or customized wavelet functions based on intrinsic ECG patterns, will not only yield a sparser representation of spatiotemporal VCG signals but also increase robustness to the aforementioned variable factors.

Our previous investigation [7] detailed the theoretical study of customized wavelet basis function design for biomedical applications. It is generally agreed that the closer the wavelet basis functions $\varphi(t)$ match the ECG signal patterns $\zeta(t)$, the sparser the representation will be. Therefore, the design of customized wavelet functions is formulated as a minimization of the distance from $\varphi(t)$ to $\zeta(t)$, i.e., $\min\|\varphi(t)-\zeta(t)\|^2$, subject to the constraints that $\varphi(t)$ satisfies the wavelet admissibility conditions. The signal pattern $\zeta(t)$ is obtained from the ECG ensembles as shown in Figure 1 (b-d). The admissibility requirements for any valid real or complex-value continuous-time function $\varphi(t)$ to be a wavelet basis function are as follows: reconstruction, zero mean, finite energy and regularity constraints. The first three defines the wave, and the last condition determines the rate of decay or the let. A function satisfied with all the four conditions can be a valid wavelet for continuous wavelet transformation [7]. Since the ECG signal pattern $\zeta(t)$ is a finite length signal, the finite energy requirement is automatically met. The zero mean condition $\int_{-\infty}^{+\infty}\varphi(t)dt=0$ implies that the Fourier transform of $\varphi(t)$ vanishes at the zero frequency $\left|F_\varphi(\omega)\right|^2\Big|_{\omega=0}=0$, where $F_\varphi(\omega)$ stands for the Fourier transform of $\varphi(t)$. So, it will also make sure that $C_\varphi=\int_{-\infty}^{+\infty}\frac{\left|F_\varphi(\omega)\right|^2}{|\omega|}d\omega=0$ is finite to guarantee inverse continuous wavelet transform in the signal reconstruction. If the function's first i moments are zero $\int_{-\infty}^{+\infty}t^i\varphi(t)dt=0$, for 0<i<(k-1), then the number of vanishing moment of the function $\varphi(t)$ is k. By imposing the wavelet admissibility constraints in the objective function, the customized wavelet was designed to further optimize the mathematical modeling of spatiotemporal VCG signals. The customized wavelet will be adopted in the multiscale basis function modeling of spatiotemporal VCG signals and compared with other standard basis functions, e.g., Gaussian basis and Mexican hat basis.

2.3.4 Customized Lead Transformation

Dower et al. [2] previously demonstrated that 3-lead VCG can be linearly transformed to 12-lead ECG without significant losses of clinical information pertaining to heart dynamics. If we have the VCG data in matrix V (3×N), eight leads (I, II, v1-v6) E (8×N) can be derived by E=V*H except the mathematical calculated and augment leads (III, aVR, aVL, aVF). H is the Dower transformation coefficient matrix with size 8×3. However, our previous investigation [21] shows that the generalized Dower transformation matrix tends to have limitations to yield consistent results for a variety of healthy control (HC) and myocardial infarction (MI) subjects. Therefore, a customized transformation matrix is necessary for a given cluster of subjects, e.g., healthy control subjects in certain age and gender group. The linear affine transformation E=H0+ V*H =$\widetilde{\mathrm{V}} * \widetilde{\mathrm{H}}$ statistically customizes the $\widetilde{\mathrm{H}}$ for each individual through the maximum likelihood estimation $\widetilde{\mathrm{H}} = (\widetilde{\mathrm{V}}^{\mathrm{T}}\widetilde{\mathrm{V}})^{-1}\widetilde{\mathrm{V}}\mathrm{E}$. Thus, $\widetilde{\mathrm{H}}$ is the proposed linear affine transform matrix with dimension 8×4. The additional column of intercept term H0 is to compensate the baseline wander and other constant biases in the long ECG data streams, so that the resulting statistical transforms are more consistent and accurate. As such, this paper presents a customized modeling framework, including customized basis function model and linear affine transformation, for multiple lead ECG systems (i.e., 3-lead VCG and 12-lead ECG).

2.4 Materials and Experimental Design

A 3-way layout experiment was designed to test the variations of model performances due to three factor groups, i.e., model complexity, basis function, and cardiac conditions. As shown in Figure 5, the number of basis functions (i.e., model complexity) is varied from M=1 to M=20. The group of 10 basis functions includes Gaussian function, customized wavelet, Mexican hat, and hermitian wavelets (i.e., the nth derivative of Gaussian). Six cardiac conditions

included are healthy control (80), MI inferior (89), MI inferior-lateral (56), MI anterior (47), MI anterior-lateral (43), and MI anterior-septal (77), available in the PhysioNet PTB Database [28]. Each recording contains 15 simultaneous heart monitoring signals, i.e., 12-lead ECG and 3-lead VCG signals.

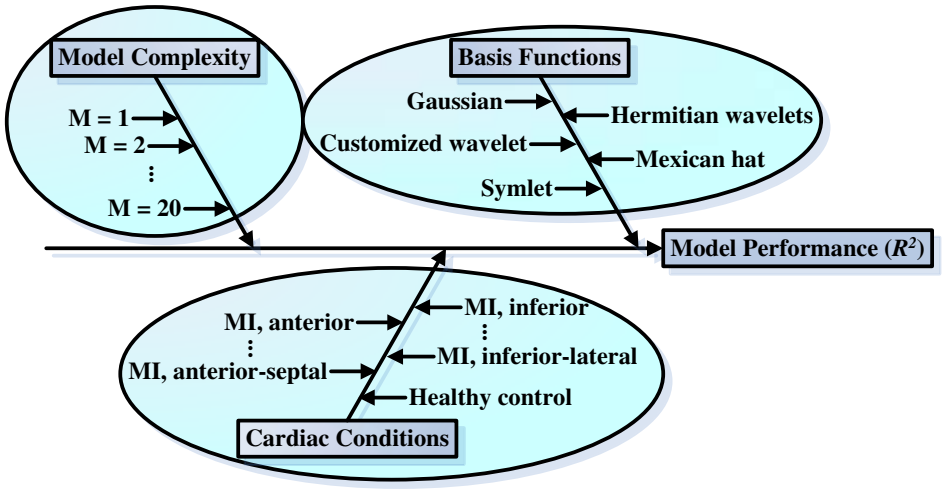


Figure 5 Cause-and-effect diagram for performance evaluation of multiscale adaptive basis function model of cardiac electrical signals

The R-square (R^2) is used to statistically measure the model performance, i.e., how well the basis function model fits the real-world VCG signals. The R^2 is defined as

$$\boldsymbol{R^2} = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum_i (v_i - \tilde{v}_i)^2}{\sum_i (v_i - \bar{v})^2} \tag{8}$$

where v_i is the real-world value of VCG signals, $\tilde{v}_i$ is the predicted value of basis function models, $\bar{v}$ is the average of real-world data, $SS_{residual}$ is the sum of squares of residuals, and SS_{total} is the total sum of squares. In other words, the R^2 is a statistic measuring the proportion of variability in a data set that can be explained by the basis function model, thereby providing the information about the goodness of fit. The range of $\boldsymbol{R^2}$ is from 0 to 1. When the value of $\boldsymbol{R^2}$ is approaching 1, the representation is closer to the original spatiotemporal and nonlinear profiles and vice versa.

2.5 Results

This present investigation made an attempt to address four practical issues pertinent to the construction of nonlinear basis function models of spatiotemporal VCG signals, i.e., adaptive parameter estimation and how the model performance (i.e., goodness of fit) is impacted by the basis function form, the number of basis function, and cardiac conditions. Furthermore, we have extracted and compared a clinic ECG metric (i.e., QT intervals) from both model representations and real-world VCG signals.

2.5.1 Multi-way Analysis of Variance

Based on the 3-way layout experiments, the data of model performances, R^2, is collected for 10 basis functions, 6 cardiac conditions and the number of basis functions from 1 to 20 (i.e., model complexity). The Analysis of Variance (ANOVA) is utilized to investigate which factor group has significant effects on the performance of basis function models. Results from ANOVA were obtained by separating the total variability of the R2, which was the sum of the squared deviations from the grand mean, into contributions by the cardiac conditions, basis function form, model complexity and random errors.

In the ANOVA table, the mean squares (MS) are computed by dividing the sums of squares (SS) by the corresponding degree of freedom (DOF). The F statistic is calculated by dividing the factor's mean squares by the residual mean square. The null hypothesis assumes that a factor has no significant effect on model performances. The F test statistically shows at a reasonable level of probability whether each of the three factor groups has a significant effect on the model performance. Generally, a smaller p value (<0.05) indicated that this factor has a significant effect. Table I shows the results of three-way ANOVA, indicating that the basis function and the model complexity have significant effects on model performances (i.e., p values

= 0, 0 and are less than the significant level 0.05) while the cardiac conditions are not significant (i.e., p value = 0.619). Therefore, it is important to investigate how the basis function and model complexity influence the performance of models (i.e., R^2 - goodness of fit).

Table 1 ANOVA Analysis of Model Performances (R^2)

Sources	SS	DOF	MS	F stat.	*p*
Cardiac Conditions	44.1	5	8.83	0.71	0.619
Basis Function	3183.4	9	353.71	28.29	0
Model Complexity	128732	19	6775.3	541.8	0
Residual	14579.2	1166	12.5		
Total	146539	1199			

2.5.2 Adaptive Basis Function Modeling

Figure 6 shows the 1-dimensional and 3-dimensional stepwise modeling of 3-lead VCG signals. The model performance R^2 is shown to monotonically increase with respect to the number of basis functions involved in the model. At each step, the matching-pursuit algorithm will scale and shift the basis function (i.e., Mexican hat function in Figure 3) to find the best-matching unit that adaptively captures the predominant wave in the VCG signals. As shown in Figure 6 (a), the matching-pursuit algorithm firstly captures the R wave with 1 basis function. This first step yields a model performance R^2 of 67.18%. It may be noted that P wave and T wave are sequentially captured in the 2nd and 3rd step. The model performance R^2 achieves 89.62% when 3 predominant waves (i.e., P, QRS, T waves) are captured. When the number of basis functions exceeds 6, the R^2 is greater than 97% showing a good representation. The matching-pursuit algorithm improves the modeling performance step by step, and captures the fine-grained details in 3-lead VCG when more and more basis functions are involved.

Furthermore, we have compared the model performance of 10 different basis functions (i.e., Gaussian function, customized wavelet, Mexican hat, and Hermitian wavelets). The model performance (R2) is calculated for 392 recordings in the PTB database. Figure 7 shows the mean and standard deviation of R2 for three basis functions, namely, customized wavelet (red), Mexican hat (blue), and Gaussian (green) when the number of basis functions is increased from 1 to 20. Gaussian basis function is widely used in previous research work, the Mexican hat is found to have a superior performance among standard basis functions. In addition, the customized wavelet basis is shown to yield better performance than Gaussian, Mexican hat and other basis functions.

It may be noted that the average of R2 is ascending, and the standard deviation of R2 is decreasing when more and more basis functions are included in the model. The experimental results show that the Mexican hat yields a better performance than Gaussian and other standard wavelet basis functions. However, the customized wavelet function evidently outperforms all other basis functions when the model complexity is within the range from 1 to 10. When the model complexity is greater than 15, the model performance (R2) is approximately the same for 3 basis functions (i.e., close to 100%). Because the customized wavelet is personalized designed for a specific subject, it captures more characteristic pattern in the 3-lead VCG than other basis functions. As a result, the model performance (R2) of customized wavelet is superior to all other basis functions when the model is sparse (i.e., from 1 to 10). When the model complexity is higher (e.g., >18), the differences of model performance (R2) are smaller among all the basis functions. It may be noted that a sparser model with sufficient explanatory power is preferable according to the principles of parsimony. Therefore, the customized wavelet is an optimal basis function for the modeling of space-time VCG signals.

2.5.3 Model Robustness to Cardiac Conditions

The 3-way ANOVA analysis shows that cardiac conditions do not have significant effects on the model performance (i.e., p value = 0.619). In other words, multiscale adaptive basis function models are robust to various cardiac conditions. In fact, the proposed model is not only applicable to the VCG signals, but also to various nonlinear and nonstationary signals.

Figure 8 shows the box plots of model performances (R2) for different cardiac conditions (i.e., healthy control - HC, MI anterior - AN, MI anterior-lateral - ANLA, and MI anterior-septal - ANSP, MI inferior - IN, MI inferior-lateral – INLA). The red line in the middle of boxplot represents the median, the blue box shows the lower quartile and upper quartile of performance distributions, and the black dash lines represent the most extreme values within 1.5 times the interquartile range.

As shown in Figure 8 (a), significant differences are found in the modeling performances for three basis functions (i.e., customized wavelet, Gaussian and Mexican hat) when the model complexity is 3. However, the modeling performances of each basis function show random variations among various cardiac conditions. When the model complexity reaches 6, the differences of modeling performances become smaller (approximately 2% ~ 6%) among 3 basis functions. As shown in Figure 8 (b), there are only random variations among various cardiac conditions for each basis function. The random variations across cardiac conditions are consistent for three basis functions in the model complexity of 3 and 6. As aforementioned in Figure 4, the differences of model performance (R2) are smaller (<0.5%) for all the basis functions when the model complexity is greater than 18. The results demonstrated the effectiveness and robustness of the proposed multiscale adaptive basis function modeling of spatiotemporal VCG signals for a variety of cardiac conditions.

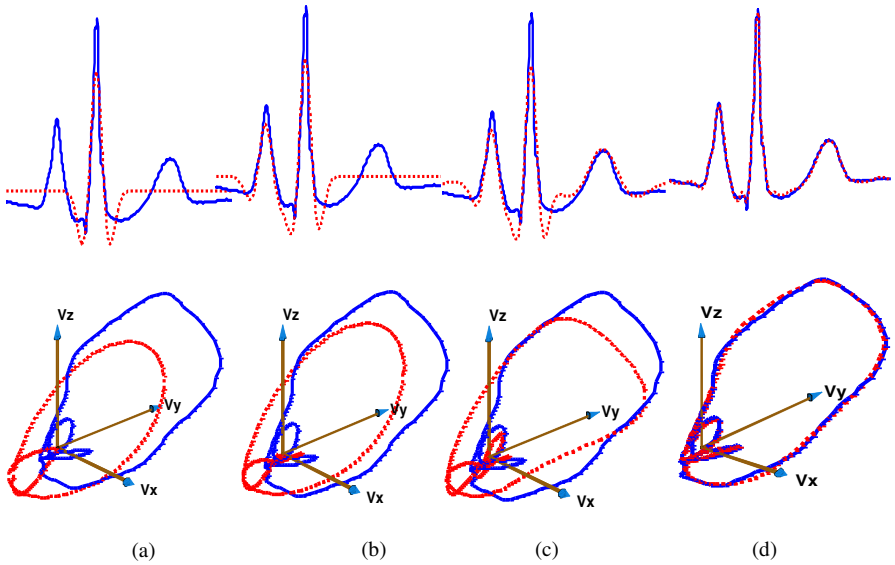


Figure 6 Stepwise performance of matching pursuit algorithms with the Mexican hat function for a MI Patient (patient006/S0064lre). (a) R^2 is 67.18% with 1 basis function; (b) R^2 is 81.20% with 2 basis function; (c) R^2 is 89.62% with 3 basis function; (d) R^2 is 99.27% with 10 basis function. The blue solid line denotes real-world VCG signals and the red dashed line is from model representation.

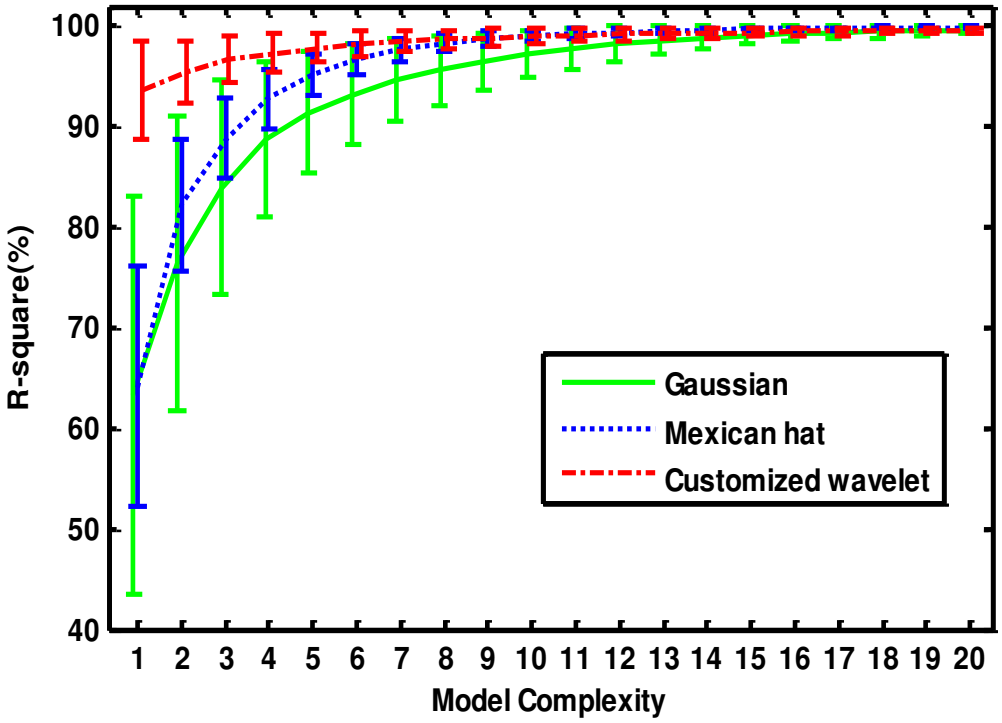


Figure 7 The mean and standard deviation of model performances for 3 basis functions (i.e., Gaussian, Mexican hat, and customized wavelet) when the model complexity is increased from 1 to 20.

Therefore, the increase of model complexity also makes the differences of modeling performances smaller and smaller for various cardiac conditions. It may be noticed that the proposed basis function model is robust to various cardiac conditions because it adaptively captures the characteristic patterns in the VCG signals. From Figure 8, it is seen that the values of R-square for Gaussian basis, Mexican hat and Customized wavelet are around 85%, 90% and 96% respectively at model complexity 3. And the values of R-square are around 92%, 96% and 98% at model complexity 6.

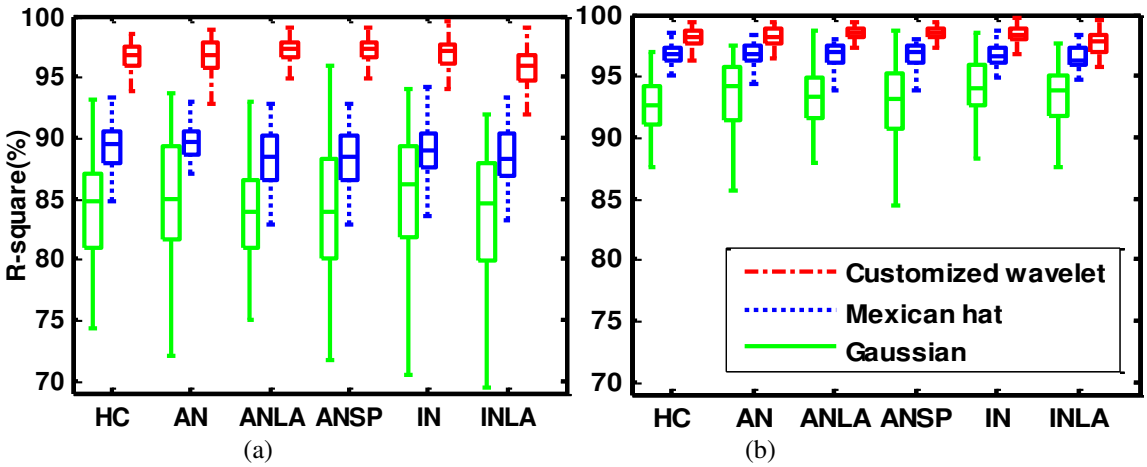


Figure 8 The variations of model performance (R2) with respect to cardiac conditions (i.e., healthy control - HC, MI anterior - AN, MI anterior-lateral - ANLA, and MI anterior-septal - ANSP, MI inferior - IN, MI inferior-lateral – INLA). (a) model complexity 3; (b) model complexity 6.

2.5.4 Model Validation with ECG Clinic Metrics

The performance metric (R2) gives the goodness-of-fit in the model representation. When the R2 is close to 100%, the model will achieve a perfect representation of real-world VCG signals. It may be noted that ECG metrics (e.g., QT intervals) are widely used in the clinical practice for the diagnostic purposes. Therefore, we have extracted the QT intervals from both real-world VCG signals and basis function representations for the model validation. The relative error is defined as

$$\delta Q = \frac{\Delta Q}{Q_A} = \frac{|Q_A - Q_M|}{Q_A} \quad (9)$$

where Q_A is the feature extracted from actual signals, Q_M is the feature extracted from model representations, and ΔQ is the absolute error between Q_A and Q_M. The QT interval represents the duration of ventricular depolarization and subsequent repolarization, and is a measure of the time between the onset of the Q wave and the offset of the T wave. A shortened or prolonged QT interval is recognized as biomarkers for the development of cardiac arrhythmias and sudden cardiac death.

As shown in Figure 9, the box plot is used to visualize the distributions of relative errors between QT intervals that are extracted from real-world VCG signals and model representation (i.e., customized wavelet, Mexican hat and Gaussian basis function models). It may be noted that the model is inadequate to capture 3 predominant waves (i.e., P, QRS and T) in the ECG signals when the number of basis functions is less than 3. Therefore, the relative errors of ECG features are large when the model complexity is smaller than 3. Figure 6 shows that the QT relative errors are decreasing when the model complexity is increased from 3 to 12. In addition, the QT relative errors of customized wavelet are smaller than Mexican hat and Gaussian when the model complexity is small (i.e., from 3 to 10). The QT relative errors of three basis functions are comparable (<5%) for all 392 subjects in PTB database when the model complexity is greater than 10. It may be noted that customized wavelet and Mexican hat are superior to the Gaussian function, which was widely adopted in many previous investigations. This experiment shows that multiscale basis function models yield small relative errors in the feature extraction from ECG signals under various cardiac conditions. It is suggested that the proposed multiscale adaptive basis function modeling effectively represents the spatiotemporal morphology of ECG signals and stores the critical information of clinic metrics.

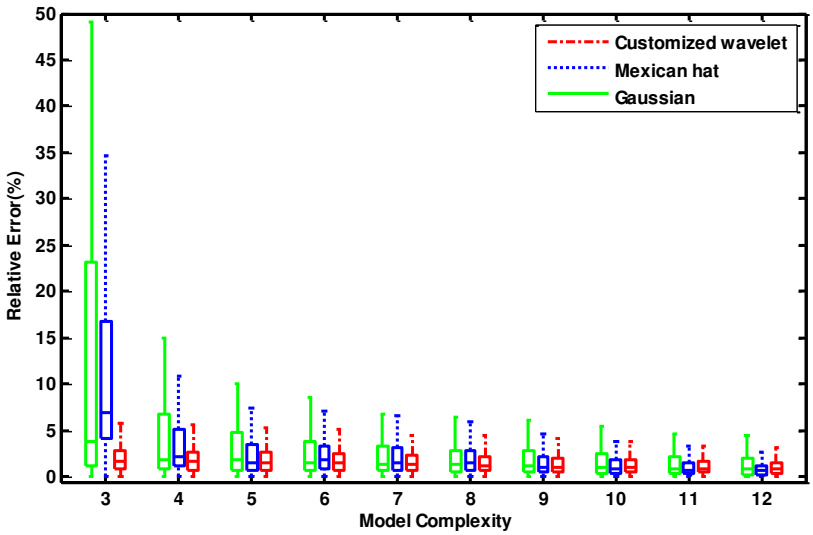


Figure 9 The mean and standard deviation of model performances for 3 basis functions (i.e., Gaussian, Mexican hat, and customized wavelet) when the model complexity is increased from 3 to 12.

2.6 Discussion and Conclusions

Mathematical modeling of cardiac electrical signals facilitates the simulation of realistic cardiac electrical behaviors, the evaluation of biomedical signal processing algorithms, and the characterization of underlying space-time patterns. However, there are practical issues pertinent to the construction of nonlinear basis function models of spatiotemporal VCG signals, i.e., adaptive parameter estimation and how the model performance (i.e., goodness of fit) is impacted by the basis function form, the number of basis function, and cardiac conditions.

This present investigation developed a multiscale basis function modeling approach to characterize not only temporal but also spatial behaviors of VCG signals. The proposed basis function models were experimentally validated using real-world VCG signals acquired from different cardiac conditions, including 80 healthy controls, 89 MI inferior, 56 MI inferior-lateral, 47 MI anterior, 43 MI anterior-lateral, and 77 MI anterior-septal. This 3-way layout ANOVA analysis shows the basis function and the model complexity have significant effects on model performances (i.e., p values = 0, 0 and are less than the significant level 0.05) while cardiac

conditions are not significant (i.e., p value = 0.619). In addition, the matching pursuit algorithms are developed to adaptively estimate the model parameters based on the "best matching" projections of ECG characteristic waves onto the dictionary of nonlinear basis functions. The model performance R^2 achieves 89.62% when 3 predominant waves (i.e., P, QRS, T waves) are captured with 3 basis functions. When the number of basis functions exceeds 6, the R^2 is greater than 97% showing a good representation.

Furthermore, the customized wavelet function is shown to outperform other basis functions when the model complexity is from 1 to 10. When the model complexity is greater than 15, the model performance (R2) is approximately the same for customized wavelet, Mexican hat and Gaussian basis functions (i.e., close to 100%). Also, the 3-way ANOVA analysis shows that the proposed basis function models are robust to various cardiac conditions, and the modeling performances of each basis function show random variations among various cardiac conditions. The comparison of a ECG metric (i.e., QT intervals) between model representations and real-world VCG signals shows that the relative errors with customized wavelet are smaller than Mexican hat and Gaussian when the model complexity is small (i.e., from 3 to 10). The relative errors of three basis functions are comparable (<5%) for all 392 subjects in PTB database when the model complexity is greater than 10.

The proposed model shows great potentials to model and analyze specific regions of interest that are related to space-time cardiac pathological behaviors. Such an effective model of 3-dimensional VCG topology will lead to the following benefits: (1) Feature extraction: The model parameters such as weights, shifting and scaling factors in the basis functions can be potentially used as features for the diagnostic application. As a result, large amount of VCG and ECG data is reduced to limited amount of features (i.e., model parameters) while preserve the

same information. (2) Data compression: It is well known that hundreds of gigabytes VCG and ECG data will be stored in the real-time cardiac monitoring. Since the basis function model yields a good representation (>99%) of real-world VCG signals, model parameters can be saved instead of the long-term VCG signals. (3) Algorithm evaluation: This proposed basis function model is data-driven and can be fitted to ECG signals from different kinds of cardiovascular diseases. The fitted model for different pathologies can generate large amount of VCG/ECG signals that can be used to test the algorithms of QRST cancellation, adaptive filtering, and classification etc. (4) Disease prognostics: Because the basis function model captures all the characteristics from the actual data, real-time ECG monitoring signals can be compared with the model representation trained in the healthy condition. The differences of pattern similarity can be used as a performance measure for the prognostic purpose. Our future research will focus on exploring the applications of multiscale adaptive basis function models of VCG signals.

2.7 References

[1] J. Malmivuo and R. Plonsey, Bioelectromagnetism. Oxford University Press, 1995.

[2] H. Yang, S. T. S. Bukkapatnam, and R. Komanduri, " Spatio-temporal representation of cardiac vectorcardiogram (VCG) signals," Biomedical Engineering Online, Vol.11, No. 16, 2012

[3] G. E. Dower, H. B. Machado and J. A. Osborne, "On deriving the electrocardiogram from vectorcardiographic leads," Clinical Cardiology, vol. 3, pp. 87-95, 1980.

[4] D. Dawson, H. Yang, M. Malshe, S. T. S. Bukkapatnam, B. Benjamin and R. Komanduri, "Linear affine transformations between 3-lead (Frank XYZ leads) vectorcardiogram (VCG) and 12-lead electrocardiogram (ECG) signals," Journal of Electrocardiology, vol. 42, pp. 622-630, 2009.

[5] N. P. Hughes, L. Tarassenko and S. J. Roberts, "Markov Models for Automated ECG Interval Analysis," Advances in Neural Information Processing Systems, vol. 16, 2003.

[6] R. H. Ireland, R. T. C. E. Robinson, S. R. Heller, J. L. B. Marques and N. D. Harris, "Measurement of high resolution ECG QT interval during controlled euglycaemia and hypoglycaemia," Physiology Measurement, vol. 21, pp. 295-303, 2000.

[7] P. Langley, F. E. Smith, S. T. King, D. Zheng, A. J. Haigh and A. Murray, "Fully automated computer measurement of QT interval from the 12-lead electrocardiogram," in Computers in Cardiology, 2006, 2006, pp. 345-348.

[8] S. T. S. Bukkapatnam, R. Komanduri, H. Yang, P. Rao, W. Lih, M. Malshe, L. M. Raff, B. Benjamin and M. Rockley, "Classification of atrial fibrillation episodes from sparse electrocardiogram data," Journal of Electrocardiology, vol. 41, pp. 292-299, 2008.

[9] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen and M. Shen Luo, "ECG Beat Detection Using Filter Banks," Biomedical Engineering, IEEE Transactions on, vol. 46, pp. 192-200, 1999.

[10] N. V. Thakor and Y. Zhu, "Applications of Adaptive Filtering to ECG Analysis: Noise Cancellation and Arrhythmia Detection," Biomedical Engineering, IEEE Transactions on, vol. 38, pp. 785-794, 1991.

[11] P. S. Addison, "Wavelet Transforms and the ECG: A Review," Physiological Measurement, vol. 26, pp. 155-199, 2005.

[12] D. Lemire, C. Pharand, J. Rajaonah, B. A. Dube and A. R. LeBlanc, "Wavelet time entropy, T wave morphology and myocardial ischemia," Biomedical Engineering, IEEE Transactions on, vol. 47, pp. 967-970, 2000.

[13] H. Yang, S. T. S. Bukkapatnam and R. Komanduri, "Nonlinear adaptive wavelet analysis of electrocardiogram signals," Physical Review E, vol. 76, pp. 026214, 2007.

[14] H. Yang, "Multiscale Recurrence Quantification Analysis of Spatial Cardiac Vectorcardiogram (VCG) Signals," Biomedical Engineering, IEEE Transactions on, vol. 58, pp. 339-347, 2011.

[15] F. M. Roberts and R. J. Povinelli, "Identification of ECG Arrhythmias Using Phase Space Reconstruction," Lecture Notes in Computer Science, pp. 411-423, 2001.

[16] F. M. Roberts, R. J. Povinelli and K. M. Ropella, "Rhythm Classification Using Reconstructed Phase Space of Signal Frequency Sub-bands," Computers in Cardiology, vol. 30, pp. 61-64, 2003.

[17] P. Strumillo and J. Ruta, "Poincare mapping for detecting abnormal dynamics of cardiac repolarization," Engineering in Medicine and Biology Magazine, IEEE, vol. 21, pp. 62-65, 2002.

[18] A. I. Rasiah, R. Togneri and Y. Attikiouzel, "Modelling 1-D signals using Hermite basis functions," Vision, Image and Signal Processing, IEE Proceedings on, vol. 144, pp. 345-354, 1997.

[19] A. Oppenheim, R. Schafer and J. Buck, Discrete-Time Signal Processing. Prentice Hall, 1999.

[20] T. Blu and M. Unser, "Wavelets, fractals, and radial basis functions," Signal Processing, IEEE Transactions on, vol. 50, pp. 543-553, 2002.

[21] M. T. Hagan, H. B. Demuth and M. H. Beale, Neural Network Design. PWS Pub, December 29, 1995.

[22] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," SIAM J. Sci. Comput., vol. 20, pp. 33-61, 01/00, 1998.

[23] P. E. McSharry, G. D. Clifford, L. Tarassenko and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," IEEE Transactions on Biomedical Engineering, vol. 50, pp. 289-294, 2003.

[24] G. D. Clifford, A. Shoeb, P. E. McSharry and B. A. Janz, "Model-based filtering, compression and classification of the ECG," International Journal of Bioelectromagnetism, vol. 7, pp. 158-161, 2005.

[25] G. D. Clifford, F. Azuaje and P. McSharry, Advanced Methods and Tools for ECG Data Analysis. Artech House Publishers, 2006.

[26] R. Sameni, G. D. Clifford, C. Jutten and M. B. Shamsollahi, "Multichannel ECG and Noise Modeling: Application to Maternal and Fetal ECG Signals," EURASIP Journal on Advances in Signal Processing, pp. 1-14, 2007.

[27] O. Sayadi, M. B. Shamsollahi and G. D. Clifford, "Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model," Physiological Measurement, vol. 31, pp. 1309, 2010.

[28] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation, vol. 23, pp. e215-e220, 2000.

[29] G. B. Moody, R. G. Mark and A. L. Goldberger, "PhysioNet: a Web-based resource for the study of physiologic signals," Engineering in Medicine and Biology Magazine, IEEE, vol. 20, pp. 70-75, 2001.

[30] D. C. Lin and R. L. Hughson, "Modeling Heart Rate Variability in Healthy Humans: A Turbulence Analogy," Phys. Rev. Lett., vol. 86, pp. 1650-1653, 2001.

[31] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," Signal Processing, IEEE Transactions on, vol. 41, pp. 3397-3415, 1993.

[32] G. Liu and H. Yang, "Multiscale Adaptive Basis Function Modeling of Spatiotemporal Vectorcardiogram Signals," Biomedical and Health Informatics, IEEE Journal of, vol. 17, pp. 484-492, 2013.

CHAPTER 3: MODEL-DRIVEN PARAMETRIC MONITORING OF NONLINEAR AND HIGH-DIMENSIONAL PROFILES[3]

In order to cope with system complexity and dynamic environments, modern industries are investing in a variety of sensor networks and data acquisition systems to increase information visibility. Multi-sensor systems bring the proliferation of high-dimensional functional profiles that capture rich information on the evolving dynamics of natural and engineered processes. This provides an unprecedented opportunity for online monitoring of operational quality and integrity of complex systems. However, the classical methodology of statistical process control is not concerned about high-dimensional sensor signals and is limited in the capability to perform multi-sensor fault diagnostics. It is not uncommon that multi-dimensional sensing capabilities are not fully utilized for decision making. This paper presents a new model-driven parametric monitoring strategy for the detection of dynamic fault patterns in high-dimensional functional profiles that are nonlinear and nonstationary. First, we developed a sparse basis function model of high-dimensional functional profiles, thereby reducing the large amount of data to a parsimonious set of model parameters (i.e., weight, shifting and scaling factors) while preserving the information. Further, we utilized the lasso-penalized logistic regression model to select a low-dimensional set of sensitive predictors for fault diagnostics. Experimental results on real-world data from patient monitoring showed that the proposed methodology outperforms

[3] This chapter was published on proceedings of 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei, Taiwan, 18-22 Aug, 2014 [24]. Permission is included in Appendix A.

traditional methods and effectively identify a sparse set of sensitive features from high-dimensional datasets for process monitoring and fault diagnostics.

3.1 Introduction

In order to cope with system complexity and dynamic environments, modern industries are investing in a variety of sensors and data acquisition systems to increase information visibility. For examples, multiple strain gauge sensors are often installed on stamping machines to collect tonnage signals for process quality improvement. Each cycle of tonnage signals measures the stamping force for producing one stamped part over a complete press stroke, indicating a series of operations such as draw, notch, blanking, cutoff and bulging [1, 2]. In addition, electrocardiogram (ECG) sensors are used to capture a wealth of dynamic information pertinent to cardiac function. Fig. 10 shows one cycle of ECG signals that corresponds to sequential stages of cardiac operations (i.e., P wave, QRS wave, and T wave) [3]. Each segmented wave is closely associated with specific physical activities of heart components. Notably, atrial depolarization (and systole) is represented by the P wave, ventricular depolarization (and systole) is represented by the QRS complex, and ventricular repolarization (and diastole) is represented by the T wave. However, a single sensor only captures 1-dimensional view of space-time dynamics of complex systems. Therefore, multi-sensor systems are usually designed to provide multi-directional views of the evolving dynamics of natural and engineered processes.

As such, multi-dimensional sensing, in days, months and even years, generates enormous amounts of data, which contains multifaceted information pertinent to the evolving dynamics of process operations. Indeed, both manufacturing and healthcare domains are facing spatially and temporally data-rich environments. Big data poses significant challenges for human experts (e.g.,

physicians, nurses, quality technicians) to accurately and precisely examine all the generated high-dimensional sensor signals for fault diagnosis and quality inspection. However, the proliferation of sensing data also provides an unprecedented opportunity to develop sensor-based methodologies for realizing the full potential of multi-dimensional sensing capabilities towards real-time process monitoring and fault diagnosis.

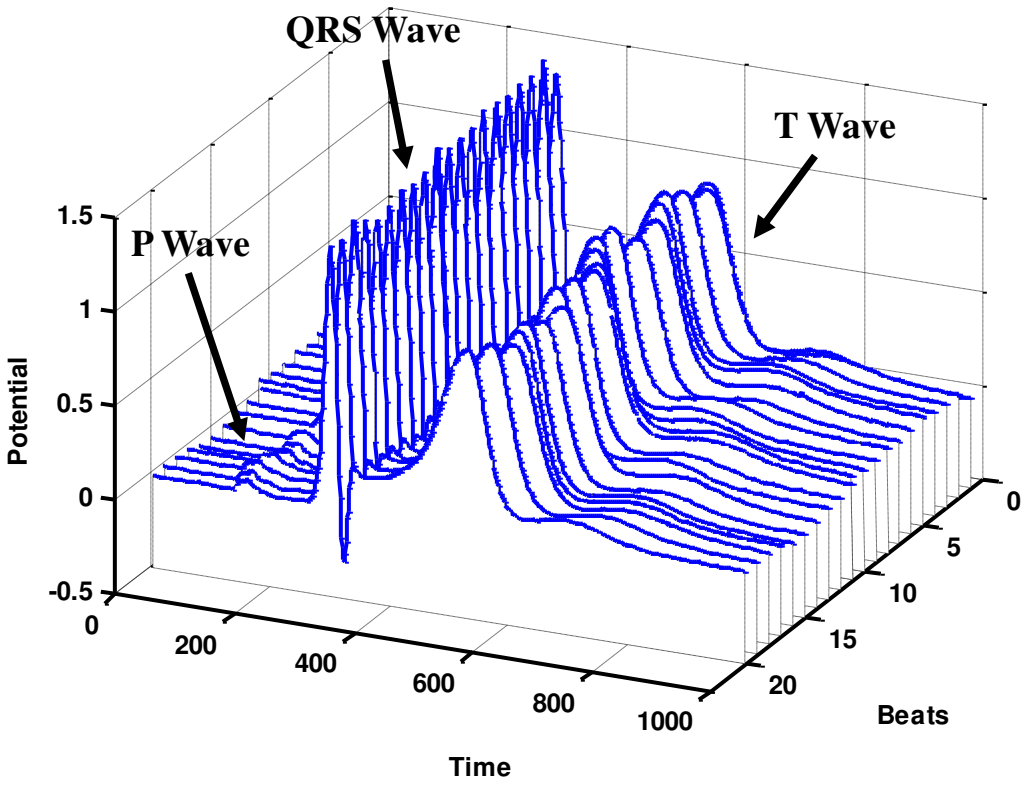


Figure 10 Examples of sensing functional profiles from human heart (electrical-mechanical biomachine) [3].

3.2 Monitoring of Functional Profiles

Existing methodologies tend to have limitations to address fundamental issues important to instantaneously transforming multi-sensor signals into useful knowledge for effective process monitoring and control. Traditional statistical process control (SPC) is not concerned with time-varying sensor signals but key product or process quality variables. Recently, the

advancement of sensing technology has fueled increasing efforts to extend SPC methods from monitoring individual data points to linear functional profiles to nonlinear functional profiles.

3.2.1 Monitoring of Linear Profiles

In the literature, extensive studies have been conducted on the monitoring of linear profiles. Kang and Albin used linear regression models to characterize linear profiles, and further monitored the variations of model parameters to describe process and product quality [4]. Kim, Mahmoud and Woodall studied the performance of various control charts for the monitoring of linear profiles and recommended optimal charts for Phase I or Phase II monitoring [5]. Further, Mahmoud and Woodall proposed linear structured model for Phase I analysis of linear profiles in calibration applications [6]. Zou et al. designed a change-point model for monitoring linear profiles by detecting the shifts in the slope, intercept and standard deviation of linear models [7]. In addition, Mahmoud et al. reported a change-point model for monitoring linear profiles by segmented regression approaches [8].

3.2.2 Monitoring of Nonlinear Profiles

Furthermore, monitoring nonlinear profiles received increasing attentions as manufacturing processes generate more and more nonlinear data that cannot be adequately represented by linear models. Koh and Shi et al. developed a series of methods that integrate engineering knowledge with statistical methods (i.e., wavelet transformation) for tonnage signal analysis and fault detection in stamping processes [1, 2]. Jin and Shi developed "feature-preserving" data compression of stamping tonnage signals using wavelets [9], and further decompose press tonnage signals to obtain individual station signals in transfer or progressive die processes [10]. Huang and Kim et al. used principal curves and latent variable modeling methods to analyze tonnage signals for in-line monitoring of forging processes [11].

In addition, Zhou et al. developed an SPC monitoring system for cycle-based waveform signals that not only detect a process change, but also identify the location and estimate the magnitude of the process mean shift within the signal [12]. Zou et al. employed non-parametric regression, generalized likelihood ratio test, and multivariate exponentially weighted moving average (EWMA) to detect the changes in nonlinear profiles [13]. Paynabar and Jin investigated both within-profile and between-profile variations using mixed-effect model and wavelet transformation [14].

3.2.3 Limitations of Previous Methods

However, most of previous work primarily focused on sample-based profiles in discrete-part manufacturing, but did not consider time-varying profiles in nonlinear dynamic processes (e.g., cardiovascular systems). Existing methods are not well suited for the detection of dynamic fault patterns in high-dimensional functional profiles from biological systems that are highly nonlinear and nonstationary.

It may be noted that cardiac ECG signals possess some common characteristics: a) Within one cycle, the waveform shows nonlinear variations and different segments change significantly that correspond to different stages of cardiac operations. b) Between cycles, the waveform is similar to each other but with variations. c) Cardiac electrical activity is varying across space and time. Current practice predominantly utilizes time-domain projections (e.g., temporal ECG tracing) of space-time cardiac electrical activity. Such a projection does not fully utilize multi-dimensional sensing capabilities for decision making [15].

As shown in Fig. 11, vectorcardiogram (VCG) signals monitor cardiac electrical activity along three orthogonal X, Y, Z planes of the body, namely, frontal, transverse, and sagittal [15-18]. Notably, VCG trajectories of myocardial infarction (red/dashed) yield a different spatial

path from the healthy controls (blue/solid). However, most previous works focused on the patterns (e.g., heart rate, ST segment, QT interval) in time-domain ECG signals, but overlooked spatiotemporal VCG signals. Therefore, in this study, we proposed the model-driven monitoring strategy for spatiotemporal functional profiles.

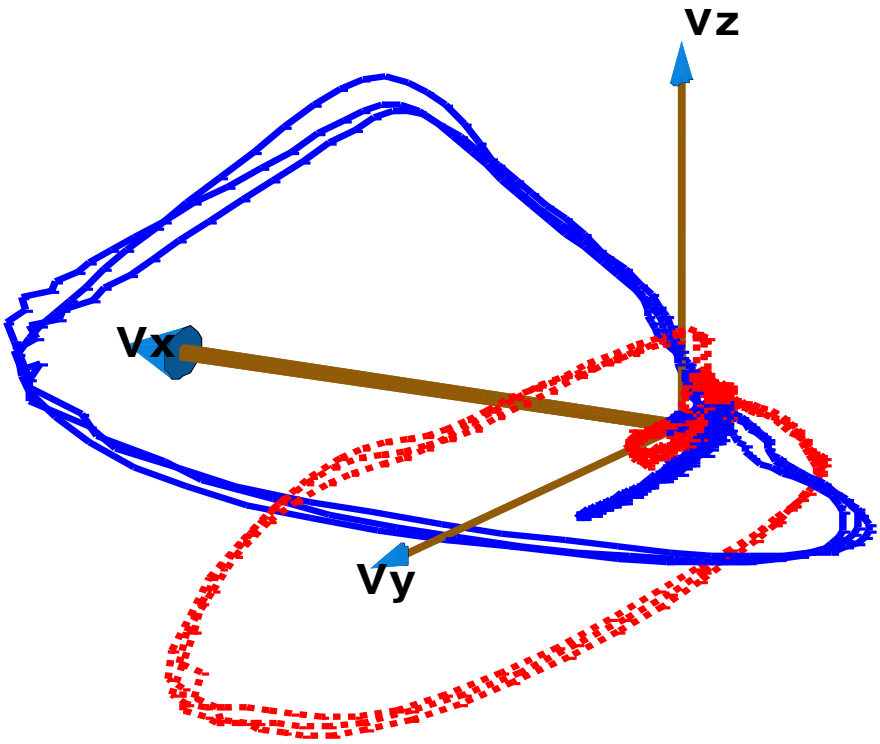


Figure 11 VCG signals of control (blue/solid) and diseased subjects (red/dashed) [15].

3.2.4 Proposed Model-driven Strategy

In this chapter, we developed a new model-driven parametric monitoring strategy for the detection of dynamic fault patterns in high-dimensional functional profiles that are both nonlinear and nonstationary.

1) Sparse modeling of high-dimensional nonlinear profiles: A sparse basis function model is developed to represent high-dimensional functional profiles, which minimizes the number of basis functions involved but maintains sufficient explanatory power. As such, large amounts of data are reduced to a parsimonious set of model parameters (i.e.,

weight, shifting and scaling factors in basis functions) while preserving the information. The model parameters and their derivatives can be used as features for the detection of process faults. However, the dimensionality of these features is high and can potentially lead to sensitive predictive models.

2) Lasso-penalized feature selection for process monitoring and fault diagnosis: Therefore, we further utilize lasso-penalized logistic regression model to investigate "redundancy" and "relevancy" properties between these parameter-based features and fault patterns, so as to identify a sparse set of sensitive predictors from the large number of features for process monitoring and fault diagnostics.

This chapter is organized as follows: Section II introduces the research methodology. Section III presents the materials and experimental results, and Section IV includes the discussion and conclusions arising out of this investigation.

3.3 Research Methodology

This paper is aimed at developing a new approach of model-driven parametric monitoring of high-dimensional nonlinear functional profiles that are nonlinear and nonstationary. Notably, the information hidden in large amount of sensing data is preserved in a parsimonious set of model parameters. In other words, nonlinear functional profiles can be reconstructed with the parameters and model structures. As such, this provides a great opportunity to take advantage of sparse model parameters for the objectives of process monitoring and fault detection.

Fig. 12 shows the overall flow chart of the proposed model-driven parametric monitoring methodology that is supported by sparse basis function modeling and lasso-penalized feature selection. Notably, multi-dimensional sensing gives rise to process data that are high-dimensional, nonlinear and nonstationary (also see Figs. 1 and 2). The present investigation

is embodied by two core components focusing on the development of model-driven parametric monitoring methodology. First, a sparse basis function model is designed to represent high-dimensional nonlinear functional profiles, thereby reducing large amount of data into a sparse set of parameters. Second, we utilized the lasso-penalized logistic regression model to investigate the "redundancy" and "relevancy" properties between features and fault patterns, thereby identifying a sparse set of sensitive predictors for process monitoring and fault diagnostics. It is remarkable that the proposed method is efficient in detecting the changes between cycles of the nonstationary profiles when it is applied to one single subject for real-time monitoring. In this paper, the 2-stage model is static because we aimed at classifying the faults from controls and the average profile is adopted.

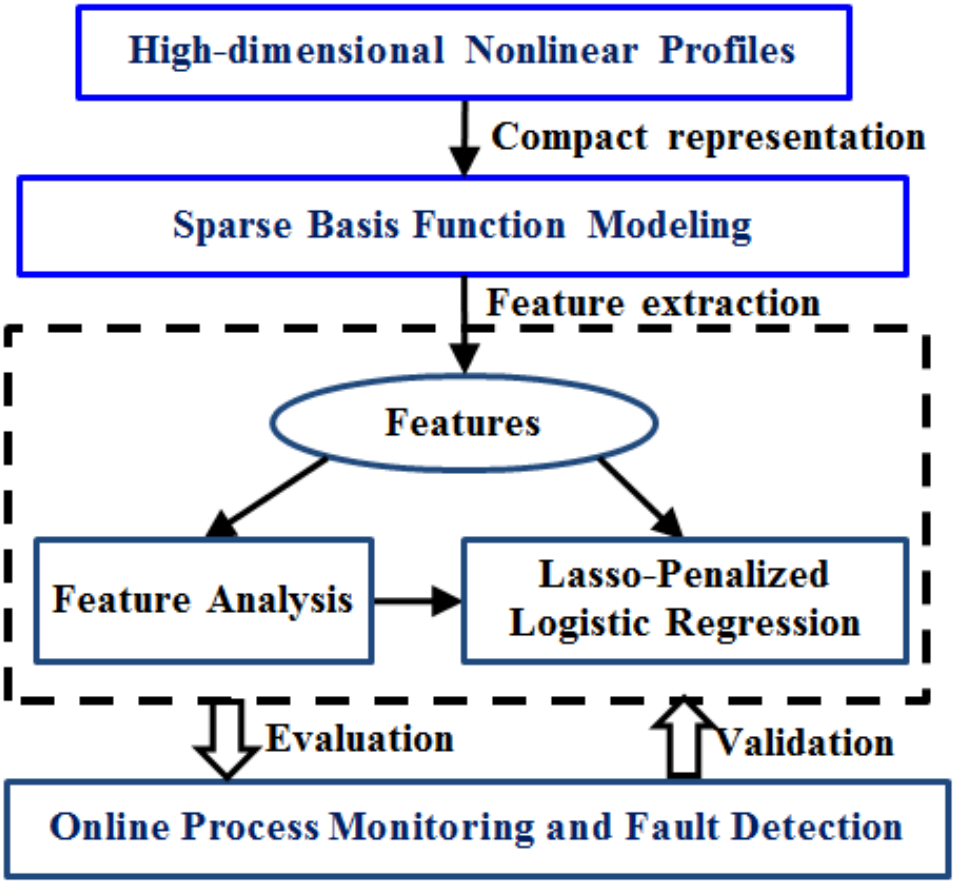


Figure 12 Flow diagram of the research methodology.

3.3.1 Model-driven Parametric Features

Multi-sensor systems provide multi-directional views of the evolving dynamics of natural and engineered processes, thereby giving rise to high-dimensional functional profiles. If d sensors are used to record the cycle profile of length T, then the dimensionality of functional profiles is $d \times T$. Notably, human heart is near-periodically beating to maintain vital living organs, and stamping machines are cyclically forming sheet metals during production. Monitoring process quality and fault conditions are more concerned with the variations of nonlinear waveforms between cycles. Here, we propose to represent the high-dimensional nonlinear profiles with the basis function model, and then monitor the variations of low-dimensional model parameters instead of the big data itself. In order to capture intrinsic characteristics in the data, we modeled the high-dimensional nonlinear profiles (see Fig. 4) as the superposition of M basis functions:

$$\boldsymbol{v}(t, \boldsymbol{w}) = \boldsymbol{w}_0 + \sum_{j=1}^{M} \boldsymbol{w}_j \boldsymbol{\psi}_j\big((\boldsymbol{t} - \boldsymbol{\mu}_j)/\boldsymbol{\sigma}_j\big) + \boldsymbol{\varepsilon} \tag{1}$$

where $\psi(t)$ is the basis function, w_j is the weight factor, μ_j is the shifting factor and σ_j is the scaling factor. Mathematically, shifting a basis function $\psi(t)$ by μ means delaying its onset and is represented by $\psi(t-\mu)$. Scaling a basis function $\psi(t)$ means either "stretching" or "shrinking" the function by a scale factor σ, i.e., $\psi(t/\sigma)$.

The objective is to optimize the representation of high-dimensional nonlinear profiles with a sparse basis function model:

$$\text{argmin}\left[\left\|\boldsymbol{v}(t) - \boldsymbol{w}_0 - \sum_{j=1}^{M} w_j \boldsymbol{\psi}_j(\boldsymbol{t})\right\|^2, \{\boldsymbol{w}, M, \boldsymbol{\psi}(\boldsymbol{t})\}\right] \tag{2}$$

Compact topological representation calls upon the minimization of the number of basis functions M and the optimal placement of basis function $\psi(t)$. Model parameters w, μ, σ are adaptively estimated by the "best matching" projections of characteristic waves of high-dimensional profiles onto a dictionary of nonlinear basis functions (see Fig. 13). Our previous work has detailed the optimization algorithms to develop a sparse basis function representation of high-dimensional profiles [3]. Such a sparse representation reduces large amount of data to a limited number of model parameters while preserving the same information. Our previous experiments show that model goodness-of-fit is greater than 99.9% (R^2) with a parsimonious set of 20 basis functions [3]. In this study, these parameters, i.e., weight, shifting and scaling factors, will be further investigated for the applications of process monitoring and fault detection.

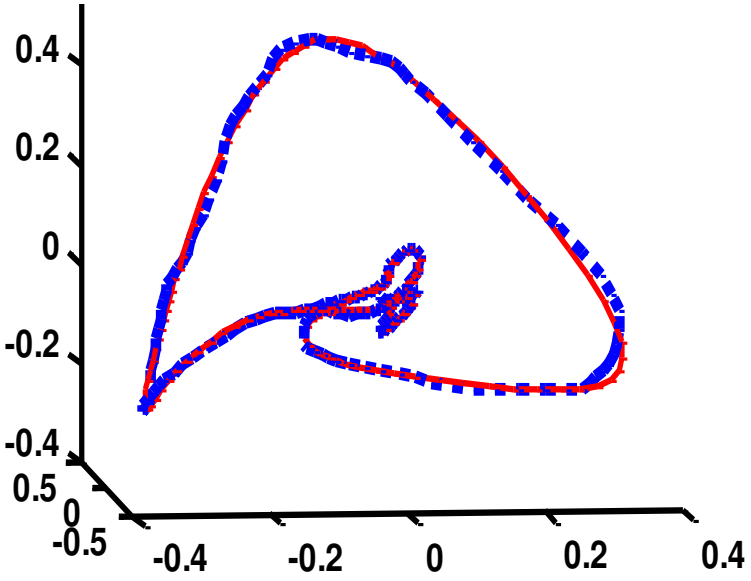


Figure 13 3D trajectory of VCG signals from basis function model (red/solid) and real-world data (blue/dashed) [3].

This present paper focuses on the extraction of parametric features from the sparse basis function model, and their further applications for fault detection. If M basis functions are used

to represent functional profiles of the dimensionality $\mathrm{d} \times \mathrm{T}$, then the set of parameters is $\{\mathrm{w}_{\mathrm{d}\times\mathrm{M}}, \mu_{\mathrm{d}\times\mathrm{M}}, \sigma_{\mathrm{d}\times\mathrm{M}}\}$. The total number of parameters will be $3 \times \mathrm{d} \times \mathrm{M}$. In this study, the number of sensors d is 3 and the number of basis function M is selected to be 20 in order to achieve >99.9% goodness-of-fit. Hence, we have a total of 180 parameters that are adaptively estimated from the 3D VCG trajectory with $\mathrm{d} \times \mathrm{T} = 3 \times 1000$ data points. In addition, we added the absolute values of weights, residual sum of squares (RSS) and the RR interval (i.e., heart rate) in this present investigation. The absolute weights indicate the amplitudes of a heartbeat. They provide different information from the original weights which is combined with the directions. The residual sum of squares (RSS) indicates the remaining pattern after the model representation. It is the complement to the basis function modeling and they together form the entire original profiles. The RR interval shows the temporal beat-to-beat time of profiles. Thus the complete feature matrix is:

$$\boldsymbol{X} = \left\{\boldsymbol{w}_{3\times20,}\boldsymbol{\mu}_{3\times20,}\boldsymbol{\sigma}_{3\times20,}|\boldsymbol{w}|_{3\times20,}\boldsymbol{RSS}_{3\times1}, RR_{1\times1}\right\} \tag{3}$$

The lasso-penalized logistic regression model will be detailed in the next section for selecting a sparse set of sensitive predictors for process monitoring and fault detection.

3.3.2 Lasso-penalized Logistic Regression

Due to the high dimensionality of features x in the vector form of $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{p})^\mathrm{T}$, there is an urgent need to select a sparse set of predictors that are sensitive to the process fault, i.e., the binary response variable y (0 or 1). Let $\mathrm{p}(\mathrm{x}, \beta)$ be the probability for y to be a success $(\mathrm{y} = 1)$ and thus $1 - \mathrm{p}(\mathrm{x}, \beta)$ is the probability for y to be a fault $(\mathrm{y} = 0)$, where $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_\mathrm{p})^\mathrm{T}$ is the coefficient vector. The logistic regression model is:

$$log\left(\frac{p(\boldsymbol{x},\boldsymbol{\beta})}{1-p(\boldsymbol{x},\boldsymbol{\beta})}\right) = \boldsymbol{\beta}^T\boldsymbol{x} \tag{4}$$

The likelihood function of $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ given the observation data $X = (x_1, x_2, .., x_n)^T, y = (y_1, \ldots, y_n)^T$ is

$$\prod_{i=1}^{n} p(\boldsymbol{x}_i, \boldsymbol{\beta})^{y_i}(1 - p(\boldsymbol{x}_i, \boldsymbol{\beta}))^{1-y_i} \tag{5}$$

As such, the log likelihood function becomes:

$$\begin{aligned} L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) &= \sum_{i=1}^{n} [y_i \log\big(p(\boldsymbol{x}_i, \boldsymbol{\beta})\big) + (1 - y_i) \log\big(1 - p(\boldsymbol{x}_i, \boldsymbol{\beta})\big)] \\ &= \sum_{i=1}^{n} [y_i \boldsymbol{\beta}^T \boldsymbol{x_i} - \log\big(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x_i}}\big)] \end{aligned} \tag{6}$$

The lasso-panelized logistic regression is formulated to minimize the following objective function with the constraint that the upper limit of L_1-norm of β is less than C,

$$\begin{aligned} &\min_{\boldsymbol{\beta}} -L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) \\ &\text{subject to } \|\boldsymbol{\beta}\|_1 \leq C \end{aligned} \tag{7}$$

This is equivalent to solve the following unconstrained optimization problem with λ be the regularization parameter:

$$\min_{\boldsymbol{\beta}, \lambda} -L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) + \lambda \|\boldsymbol{\beta}\|_1 \tag{8}$$

Notably, there is a one-to-one correspondence between C in equation (7) and λ in equation (8). The optimal solution β of the unconstrained optimization problem given λ also solves (7) with $C = \|\beta\|_1 = \sum_{i=1}^{p} |\beta_i|$. To solve this constrained optimization problem, let's first obtain the solution to the unregularized logistic regression model. The objective function of unregularized logistic regression model is:

$$\min_{\boldsymbol{\beta}} -L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) \tag{9}$$

From the Newton-Raphson algorithm, it may be noted that the update of parameters is obtained by approximating the objective function with the second-order Taylor expansion. Let $\beta^{(k)}$ be

the current parameters, then Newton-Raphson method finds the new set of parameters $\gamma^{(k)}$ based on the quadratic approximation as follows:

$$\boldsymbol{\gamma}^{(k)} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z} \tag{10}$$

where $z = X\beta + W^{-1}(y - p)$ and W is the diagonal matrix with $(W)_{ii} = p(x_i, \beta)\big(1 - p(x_i, \beta)\big)$. As such, solving for $\gamma^{(k)}$ is equal to find the solution to the following weighted least squares problem:

$$\boldsymbol{\gamma}^{(k)} = arg\min_{\boldsymbol{\gamma}} \left\| \left(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X} \right)\boldsymbol{\gamma} - \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{z} \right\|_2^2 \tag{11}$$

For lasso-penalized logistic regression, there is a need to add the L_1 constraint to the unregularized logistic regression so as to ensure $\|\gamma\|_1 \leq C$, i.e.,

$$\begin{gathered} \min_{\boldsymbol{\gamma}} \left\| \left(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X} \right)\boldsymbol{\gamma} - \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{z} \right\|_2^2 \\ \text{subject to } \|\boldsymbol{\gamma}\|_1 \leq C \end{gathered} \tag{12}$$

As a result, the lasso-penalized logistic regression is transformed to an iteratively reweighted least square problem [19]. At each iteration, we update the $W^{\frac{1}{2}}X$ and $W^{\frac{1}{2}}z$ based on the new estimate of coefficients. After $\gamma^{(k)}$ is obtained, we update $\beta^{(k)}$ by

$$\boldsymbol{\beta}^{(k+1)} = (1 - \theta)\boldsymbol{\beta}^{(k)} + \theta\boldsymbol{\gamma}^{(k)} \tag{13}$$

where $\theta \in [0,1]$ is the learning rate for the parameter update. In this study, we adopted the coordinate descent algorithm [20] to solve the regularized problem in equation (12). If we write $W^{\frac{1}{2}}X = \breve{X}$ and $W^{\frac{1}{2}}z = \breve{y}$, only one β_j is changed at each time while the other parameters β_k $(k \neq j)$ stay the same. Table 2 summarizes the lasso-penalized logistic regression algorithms used in this study. It is suggested that the weighted average is the fastest way to find the next most appropriate vector of coefficients.

Table 2 Lasso-penalized Logistic Regression Algorithm

Step 1: Initialize $\beta^{(0)} = 0$

Step 2: Start from $k = 1$, compute matrix W and vector z

Step 3: Coordinate descent algorithm to solve the L_1 constrained least squares problem in equation (12) and find $\gamma^{(k)}$

3.1 Standardize the predictors, the columns of $\breve{X}$ to have mean zero and unit norm.

3.2 Initialize the coefficient vector $\gamma = 0$.

3.3 For $j = 1$ to p

$$\gamma_j^{lasso} = S(\gamma_j^{ols}, \delta) = \begin{cases} \gamma_j^{ols} - \delta, & \text{if } \gamma_j^{ols} > 0 \text{ and } \delta < |\gamma_j^{ols}| \\ \gamma_j^{ols} + \delta, & \text{if } \gamma_j^{ols} < 0 \text{ and } \delta < |\gamma_j^{ols}| \\ 0, & \text{if } \delta \geq |\gamma_j^{ols}| \end{cases}$$

where γ_j^{ols} is the least squares estimation of γ_j and $\gamma_j^{ols} = \sum_{i=1}^{n} \breve{x}_{ij}(\breve{y}_i - \breve{y}_i^{(j)})$ in which $\breve{y}_i^{(j)} = \sum_{k \neq j} \breve{x}_{ik} \gamma_k$.

3.4 Repeat step 3.3 until converge

Step 4: Set $\beta^{(k+1)} = (1 - \theta)\beta^{(k)} + \theta\gamma^{(k)}$

Step 5: Evaluate the objective function in equation (7) at $\beta^{(k+1)}$

Step 6: Stop if the stopping criterion is satisfied; otherwise, repeat step 2 to step 5 until the stopping criterion is satisfied.

3.4 Materials and Experimental Results

In this present investigation, we used the 3-dimensional nonlinear profiles from 388 subjects (79 controls and 309 faults), available in the PhysioNet Database [21, 22]. Each functional profile recorded near-periodic electrical activity of human heart, and was digitized at 1 kHz sampling rate with a 16-bit resolution over a range of 16.384 mV. We have developed the sparse basis function model for each functional profile. For modeling details, see our previous investigation [3]. Model-driven parametric features (see equation 3) include the weights w_i for each basis, the shifting factor μ_i and the scaling factor σ_i of each basis, the absolute value of weights $|w_i|$, the residual sum squares (RSS) and the RR interval. In total, there are 244 features for each subject. As shown in the following sections, we will further delineate the structures inherent to these features and identify a parsimonious set of sensitive predictors for process monitoring and fault diagnosis.

3.4.1 Parametric Feature Analysis

First, we tested the statistical significance of parametric features using the Kolmogorov-Smirnov (KS) test. Two-sample KS test is utilized to test the differences in cumulative distribution function (CDF) between Controls and Faults [23]. Let $F_i(x)$and $F_j(x)$ denote cumulative distribution functions of the ith and jth groups respectively and the hypotheses of KS test are:

$$H_0: F_i(x) = F_j(x) \text{ or } H_1: F_i(x) \neq F_j(x)$$

The test statistics (KS stat.) and critical values (KS crit.) are shown in the Fig. 5. The KS test statistic is compared with the corresponding critical value given by $c_\alpha\sqrt{\frac{n_1+n_2}{n_1 n_2}} = 1.36\sqrt{\frac{79+309}{79\times309}} = 0.17$, where n1 and n2 are the number of independent observations in corresponding groups, the significant level is $\alpha = 0.05$, and $c_\alpha$is approximated as 1.36. If the KS statistic is greater than the critical value, the null hypothesis H_0 will be rejected and the cumulative distribution functions of two groups are declared to be different at significant level of 0.05. Hence, the bigger the KS statistic, the more significant the feature is.

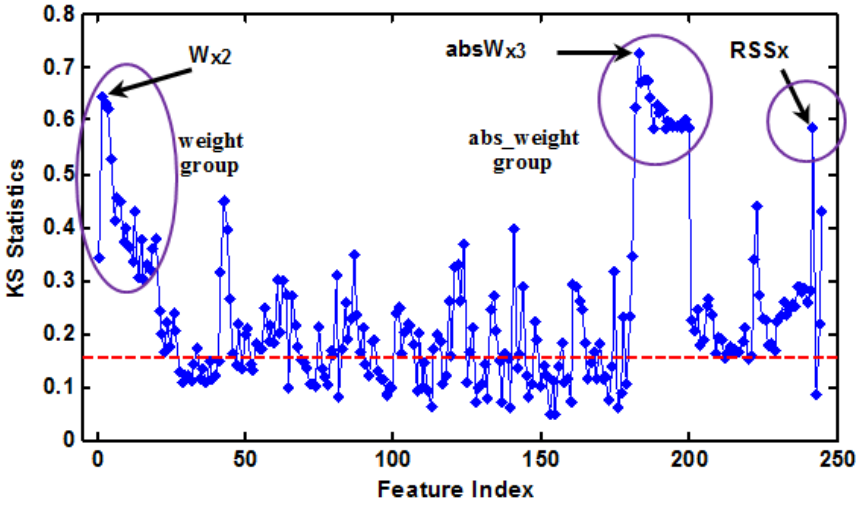


Figure 14 The KS statistics for model-driven parametric features.

As shown in Fig. 14, there are more than 60% of features that have the KS statistic greater than 0.17. In particular, three significant groups have the KS statistic greater than 0.5 that are highlighted in purple circles. It may be noted that three features, namely W_{X2} (the weight of the second basis in X direction), $ABSW_{X3}$ (the absolute value of weight of the third basis in X direction) and RSS_x (Residual sum squares in X direction), yield the KS statistic > 0.58. The results of KS test show that most of model-driven parametric features are significantly different between control and fault conditions. It is worth mentioning that weight factors are the most significant group of features among all parametric features.

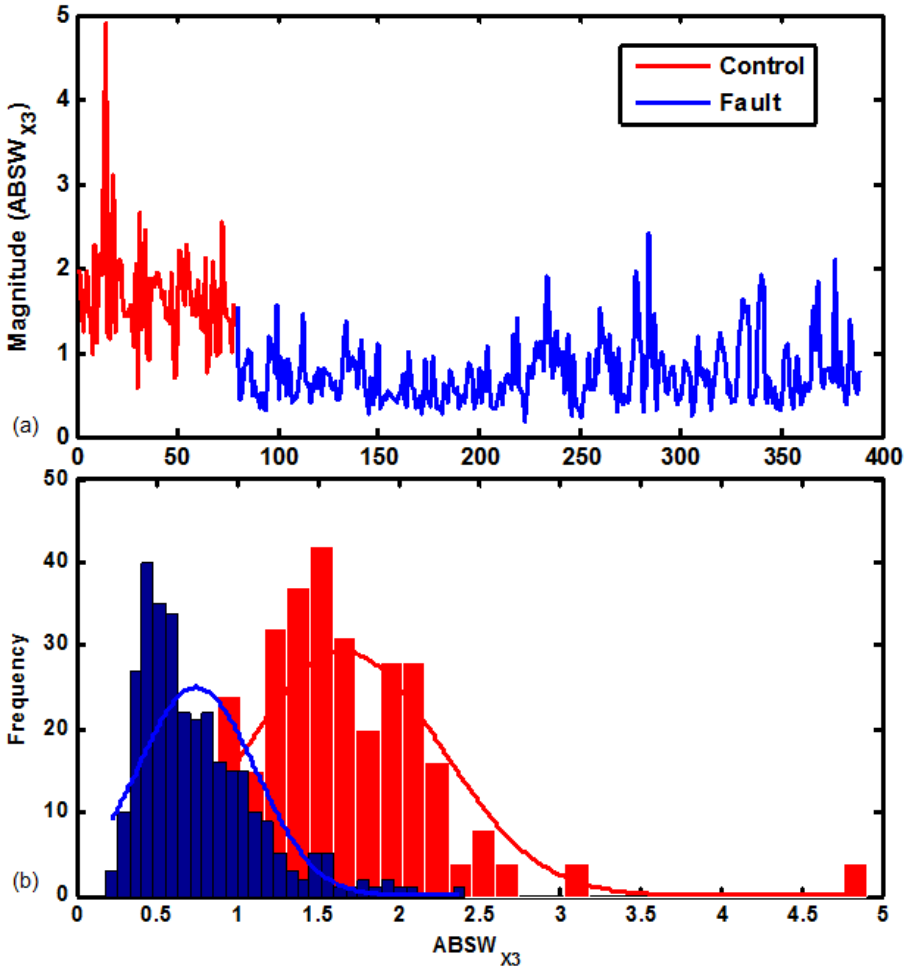


Figure 15 The visualization of $ABSW_{X3}$ (i.e., the feature with highest KS statistic 0.73) in the forms of (a) scatter plot and (b) histogram.

Fig. 15 shows the visualization of $ABSW_{X3}$ (i.e., the feature with highest KS statistic 0.73) in the forms of time-series plot (a) and histogram (b). The values of $ABSW_{X3}$ from the control group are marked in red, and the fault group is marked in blue. Both plots show distinct differences between the control and fault groups. Notably, the control group yields a bigger mean and variation than the fault group. It is remarkable because biomachine is operating in a vastly different way from mechanical machines. When the heart is healthy, it tends to be more active and dynamic. On the contrary, if the heart is diseased, degree of freedom is lost and its activity will not be as versatile as the healthy one.

3.4.2 Feature Selection via Lasso-penalized Logistic Regression

However, a large number of predictors tend to bring the "curse of dimensionality" problem, as well as the overfitting for the predictive modeling. Therefore, we adopted the lasso-penalized logistic regression model to shrink the number of predictors and identify a sparse set of sensitive features. By regularized learning, lasso-penalized logistic regression can also increase the model interpretability, as opposed to the transformed features with dimensionality reduction methods (e.g., principal component analysis). Fig. 16 shows the coefficient paths of all the predictors when the $L1$ constraint of parameters is increased. Three bolded paths indicate the first three predictors entered the model, i.e., $ABSW_{X3}$, W_{X2} and $ABSW_{X6}$. This selection is consistent with the results of Kolmogorov–Smirnov test in section III.A. If we only use these 3 selected predictors, the logistic regression model yields an accuracy of 86.77% (with sensitivity 88.90% and specificity 84.63%). The vertical blue line indicates an optimal regularization parameter that is identified using cross validation.

Fig. 17 shows the variations of prediction errors with respect to the regularization parameter. Notably, as the number of selected features increases, the prediction error decreases to

a certain point and then increases. The optimal regularization parameter λ_{opt} is identified as the point with minimal cross-validation error plus one standard deviation, which is indicated in Fig. 8 as the blue dashed line and the blue circle. Notably, the optimal regularization parameter λ_{opt} suggested the selection of 81 features, and achieved the accuracy of 97.13% (sensitivity 94.68% and specificity 99.60%). In addition, we also experimented to select the set of 10 features using lasso-penalized logistic regression and achieve the accuracy of 91.15% (sensitivity 89.70% and specificity 92.61%). It may also be noted that if we use all the 244 features and the logistic regression model without lasso penalization, the prediction accuracy is about 88.77%. It is evident that feature selection via Lasso penalization yields not only a simpler model but also much better performances.

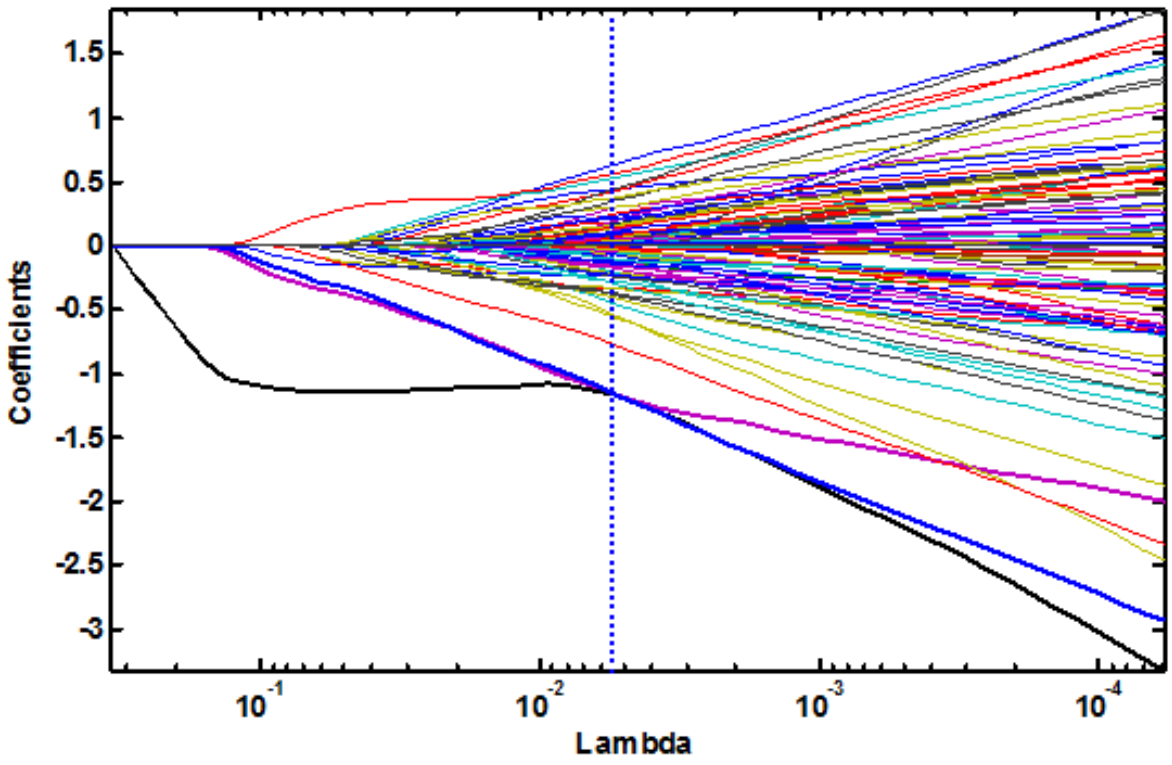


Figure 16 Coefficient path for lasso-penalized logistic regression model.

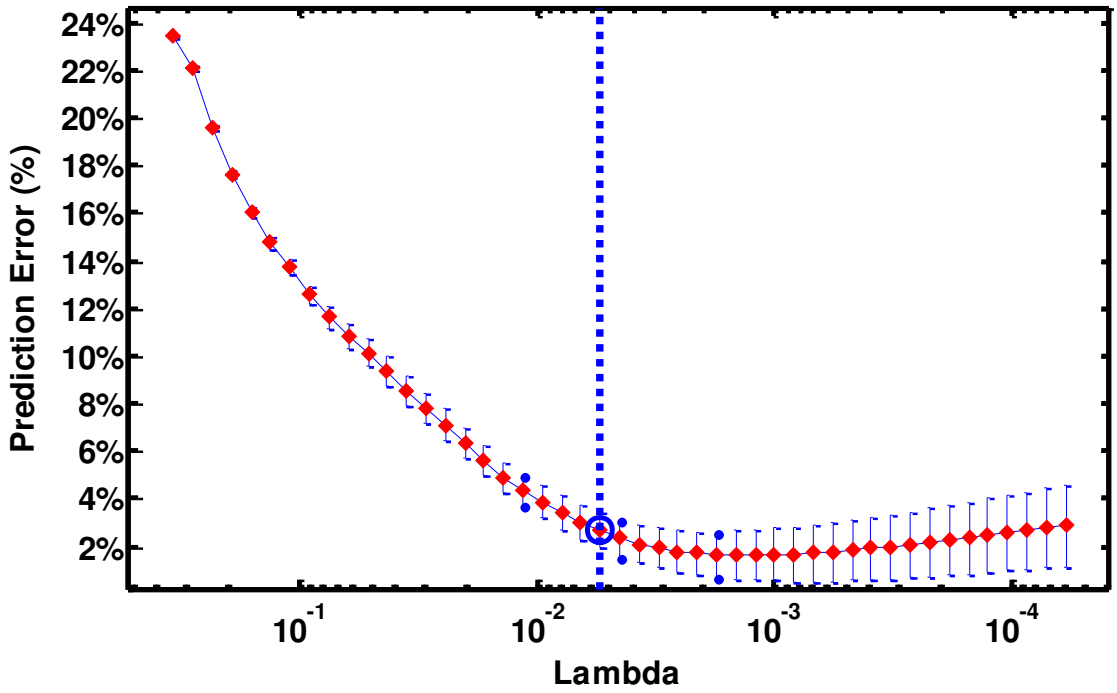


Figure 17 The variations of prediction errors vs. the regularization parameter in lasso-penalized logistic regression.

3.5 Conclusion and Discussion

Few, if any, previous works considered model-driven parametric monitoring strategy for high-dimensional nonlinear functional profiles generated from the operation of biomachine (e.g., human heart) that is highly nonlinear and nonstationary. Current practice predominantly utilizes time-domain projections (e.g., temporal ECG tracing) for diagnostic and prognostic applications. In addition, traditional SPC methods are not concerned about high-dimensional sensor signals and are limited in the capability to perform multi-sensor fault diagnostics. Notably, multi-dimensional sensing capabilities are not fully utilized for real-time process monitoring and fault diagnosis.

This present study developed a new approach of model-driven parametric monitoring of high-dimensional nonlinear functional profiles. First, a sparse basis function model is designed to represent high-dimensional nonlinear functional profiles, thereby reducing large amount of data

into a sparse set of parameters. Second, we utilized the lasso penalized logistic regression model to investigate the "redundancy" and "relevancy" properties between features and fault patterns, thereby identifying a sparse set of sensitive predictors for fault diagnostics. Experimental results show that there are more than 60% of features that have the KS statistic greater than the critical value 0.17, indicating significantly differences between control and fault conditions. Further, the lasso-penalized logistic regression model yields a superior accuracy of 97.13% with a parsimonious set of 81 features. The developed sensor-based methodology for model-driven parametric monitoring facilitates the modeling and characterization of high-dimensional nonlinear profiles and provides effective predictors for real-time fault detection, thereby promoting the understanding of fault-altered spatio-temporal patterns in the complex natural and engineered systems.

3.6 References

[1] C. Koh, J. Shi and J. Black, "Tonnage Signature Attribute Analysis for Stamping Process," NAMRI/SME Transactions, vol. 23, pp. 193-198, 1996.

[2] C. Koh, J. Shi and W. Williams, "Tonnage Signature Analysis Using the Orthogonal (Harr) Transforms," NAMRI/SME Transactions, vol. 23, pp. 229-234, 1995.

[3] G. Liu and H. Yang, "Multiscale adaptive basis function modeling of spatiotemporal cardiac electrical signals," IEEE Journal of Biomedical and Health Informatics, vol. 17, pp. 484-492, 2013.

[4] L. Kang and S. L. Albin, "Online monitoring when the process yields a linear profile," Journal of Quality Technology, vol. 32, pp. 418-426, 2000.

[5] K. Kim, M. A. Mahmoud and W. H. Woodall, "On the monitoring of linear profiles," Journal of Quality Technology, vol. 35, pp. 317, 2003.

[6] M. A. Mahmoud and W. H. Woodall, "Phase I Analysis of Linear Profiles With Calibration Applications," Technometrics, vol. 46, pp. 380-391, 2004.

[7] C. Zou, Y. Zhang and Z. Wang, "A control chart based on a change-point model for monitoring linear profiles," IIE Transactions, vol. 38, pp. 1093-1103, 2006.

[8] M. A. Mahmoud, P. A. Parker, W. H. Woodall and D. M. Hawkins, "A change point method for linear profile data," Quality and Reliability Engineering International, vol. 23, pp. 247-268, 2007.

[9] J. Jin and J. Shi, "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets," Technometrics, vol. 41, pp. 327-339, 1999.

[10] J. Jin and J. Shi, "Press Tonnage Signal Decomposition and Validation Analysis For Transfer or Progressive Die Processes," ASME Transactions, Journal of Manufacturing Science and Engineering, vol. 127, pp. 231-235, 2005.

[11] J. Kim, Q. Huang, J. Shi and T. -. Chang, "Online Multi-Channel Forging Tonnage Monitoring and Fault Pattern Discrimination Using Principal Curve," Transactions of the ASME, Journal of Manufacturing Science and Engineering, vol. 128, pp. 944-950, 2006.

[12] S. Zhou, B. Sun and J. Shi, "An SPC Monitoring System for Cycle-Based Waveform Signals Using Haar Transform," IEEE Transactions on Automation Science and Engineering, vol. 3, pp. 60-72, 2006.

[13] C. Zou, F. Tsung and Z. Wang, "Monitoring profiles based on nonparametric regression methods," Technometrics, vol. 50, pp. 512-526, 2008.

[14] K. Paynabar and J. Jin, "Characterization of non-linear profiles variations using mixed-effect models and wavelets," IIE Transactions, vol. 43, pp. 275-290, 2011.

[15] H. Yang, C. Kan, G. Liu and Y. Chen, "Spatiotemporal differentiation of myocardial infarctions," IEEE Transactions on Automation Science and Engineering, vol. 10, pp. 938-947, 2013.

[16] H. Yang, "Multiscale Recurrence Quantification Analysis of Spatial Cardiac Vectorcardiogram (VCG) Signals," Biomedical Engineering, IEEE Transactions on, vol. 58, pp. 339-347, 2011.

[17] H. Yang, S. T. S. Bukkapatnam and R. Komanduri, "Spatio-temporal representation of cardiac vectorcardiogram (VCG) signals," Biomedical Engineering Online, vol. 11, pp. 16, 2012.

[18] H. Yang, S. T. S. Bukkapatnam, T. Le and R. Komanduri, "Identification of myocardial infarction (MI) using spatio-temporal heart dynamics," Medical Engineering & Physics, vol. 34, pp. 485-497, 2011.

[19] S. Lee, H. Lee, P. Abbeel and A. Y. Ng, "Efficient L1 regularized logistic regression," in Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), 2006, pp. 1-8.

[20] J. Friedman, T. Hastie, H. Hofling and R. Tibshirani, "Pathwise coordinate optimization," The Annals of Applied Statistics, vol. 1, pp. 302-332, 12, 2007.

[21] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation, vol. 23, pp. e215-e220, 2000.

[22] G. B. Moody, R. G. Mark and A. L. Goldberger, "PhysioNet: a Web-based resource for the study of physiologic signals," Engineering in Medicine and Biology Magazine, IEEE, vol. 20, pp. 70-75, 2001.

[23] C. Kan and H. Yang, "Dynamic spatiotemporal warping for the detection and location of myocardial infarctions," in Automation Science and Engineering (CASE), 2012 IEEE International Conference on, 2012, pp. 1046-1051.

[24] G. Liu, C. Kan, Y. Chen and H. Yang, "Model-driven parametric monitoring of high-dimensional nonlinear functional profiles," in Automation Science and Engineering (CASE), 2014 IEEE International Conference on, 2014, pp. 722-727.

CHAPTER 4: SELF-ORGANIZING NETWORK FOR VARIABLE CLUSTERING AND PREDICTIVE ANALYTICS

Rapid advancement of sensing and information technology brings the big data, which presents a gold mine of the 21st century. However, big data also brings significant challenges for data-driven decision making. In particular, it is not uncommon that a large number of variables (or features) underlie the big data. Complex interdependence structures among variables challenge the traditional framework of predictive modeling. This paper presents a new methodology of self-organizing network for variable clustering and predictive modeling. Specifically, we developed a new approach, namely nonlinear coupling analysis to measure nonlinear interdependence structures among variables. Further, all the variables are embedded as nodes in a complex network. Nonlinear-coupling forces move these nodes to derive a self-organizing topology of network. As such, variables are clustered as sub-network communities in the space. Experimental results on simulation studies and real-world data demonstrated that the proposed methodology not only outperforms traditional variable clustering algorithms such as hierarchal clustering and oblique principal component analysis, but also effectively identify interdependent structures among variables and further improves the performance of predictive modeling. The proposed new idea of self-organizing network is generally applicable for predictive modeling in many disciplines that involve a large number of highly-redundant variables and the complex interdependence structures among the variables, e.g., nonlinear and asymmetric interdependence.

This paper is motivated by the need to handle complex interdependence structures among a large number of variables that underlie the big data. This paper presents a novel approach of self-organizing variable clustering that embeds each variable as a node in the complex network, and then leverages nonlinear-coupling forces to derive a self-organizing network. As a result, sub-network communities delineates the cluster communities delineates the cluster structure of variables. Experimental results in both simulation studies and real-world case studies demonstrated effectiveness and robustness of the proposed model to improve the performance of predictive modeling.

4.1 Introduction

Predictive analytics leverages information and patterns extracted from large amounts of data to predict outcomes and drive decisions or actions. It is extensively used in a variety of areas, e.g., business, healthcare and manufacturing. Predictive outcomes can be categorized into three different types: rankings, decisions or estimations. In business, online retailers strive to increase sales by using predictive analytics to rank the profitable levels of customers and further provide them with favorite products [1]. In healthcare, professionals use predictive analytics to extract information from clinical data and exploit data-driven patterns for better medical decision making [2, 3]. In manufacturing, predictive analytics helps estimate the degradation trajectory so as to prevent potential failures of manufacturing equipment and defective products [4]. Indeed, predictive analytics is critical to increasing the companies' profits, improving the health of our society, and enhancing the performance of manufacturing systems.

As a result, modern industries are investing a variety of sensing technology, data centers and information systems to increase the information visibility. For examples, retail companies collect large amounts of transactional data about their customers, suppliers and store operations

to facilitate marketing research and drive revenue improvements [1]. Large-scale manufacturing systems deploy thousands of networked sensors to record intrinsic details of production operations for performance improvements [4]. Healthcare systems in the 21st century increasingly adopt electronic health records, advanced biomedical sensing and patient monitoring systems to assist in the process of clinical decision making [2, 3]. Advanced sensing and data acquisition technology bring the new era of big data. This provides an unprecedented opportunity for predictive analytics.

However, high dimensionality and complex structures of big data pose significant challenges on traditional methodologies in predictive analytics. Realizing the full potentials of big data for predictive analytics hinges upon the development of new methodologies that effectively handle the high dimensionality of variables and complex interdependencies among variables (predictors). It is not uncommon that a large number of variables underlie the big data, which brings the issue of "curse of dimensionality" in predictive analytics. When the dimensionality increases, large amounts of training data are required to learn predictive models. Notably, the "curse of dimensionality" increases means squared errors (MSE) and the bias of predicted responses [5]. On the other hand, complex interdependence structures among variables pose a new challenge on predictive analytics. It is well known that a higher correlation (>0.90) between variables (collinearity) leads to more sensitive estimations of parameters in predictive models (i.e., increased variances of estimation) [6]. In addition, complex systems often exhibit nonlinear coupling and synchronization behaviors [7]. As such, there are nonlinear interdependence structures among process variables of complex systems. Linear and nonlinear redundancies among variables impact the performance of predictive analytics. Therefore a predictive strategy considering interdependent structures is in urgent need.

In the literature, variable selection and variable clustering are widely used to address these challenges. For example, generalized linear models (GLM) are often integrated with shrinkage methods to optimize model sparsity and improve the prediction accuracy. Examples of shrinkage and selection methods include best-subset selection [8], ridge regression [9], LASSO [10], least angle regression [11] and elastic net [12]. It is worth mentioning that most of existing variable selection methods focus on the relevancy between predictors and response variables. Interdependent structures among predictors are often overlooked, or not explicitly investigated. On the other hand, variable clustering depends on similarity measurements between variables such as linear correlation or mutual information [13]. Yet, linear correlation cannot capture nonlinear interdependences among variables. Mutual information characterizes linear and nonlinear correlation, but requires the stationarity assumption [13]. Latent-variable methods are also commonly used for variable clustering, e.g., oblique principal component clustering (OPCC) [14]. However, Oblique rotation and principal component analysis are based on linear projections of variables. As such, nonlinear interdependences among variables are not fully considered. Notably, variable clustering mainly focuses on the redundancy among variables with neglecting the relevancy between predictors and the response variables. New methodologies that integrate variable clustering with variable selection to improve effectiveness and efficiency of predictive analytics are urgently needed.

In this paper, we developed a new methodology of self-organizing network that leverages advantages of variable clustering and variable selection to investigate both redundancy and relevancy among variables for improving the performance of predictive modeling. Specifically, we developed a new approach, namely nonlinear coupling analysis, to measure nonlinear interdependence structures among variables. Further, we embedded these variables are nodes in a

complex network. Nonlinear-coupling forces move these nodes to derive a self-organizing topology of network. As such, variables are clustered as sub-network communities in the space. Experimental results in both simulation studies and real-world case studies demonstrated that the proposed methodology not only outperforms traditional variable clustering algorithms such as hierarchal clustering and oblique principal component analysis, but also effectively identify interdependent structures among variables and further improves the performance of predictive modeling at the same time.

The reminder of this chapter is organized as follows: Section II reviews the research background; Section III presents the methodology; Section IV contains experimental design and results of simulation study; Section V shows the results of a real-world case study that extract model parameters from vectorcardiogram (VCG) signals for the identification of myocardial infarctions; and Section VI includes the conclusions arising out of this investigation.

4.2 Research Background

Variable clustering is aimed at detecting subsets of homogeneous variables and then clustering them into the same group, in which variables have stronger interrelations to each other than to those in other groups. It is worth mentioning that variable clustering is different from data clustering. As shown in Fig. 18 (a), data clustering separates samples into clusters, where each point represents a data sample, X-axis is the dimension of variable 1 and Y-axis is the dimension of variable 2. Here, each data point has two coordinates, e.g., (1.26, -3.36) represents a data sample in the 2-dimensional space. Data samples are clustered based on the distance measure, e.g., Euclidean distance. Data clustering is an unsupervised method to group data samples into homogeneous clusters based on distance measures. However, Fig. 18 (b) illustrates the clustering results for 21 variables, each of which has 1000 data samples. For example, the variable X_{18}

represents a series of 1000 data samples. Notably, each point in Fig. 18 (b) is a variable instead of a data sample. Variable clustering considers the interdependence structure among variables, e.g., correlation or mutual information. The Pearson's correlation between variables between variables x_1 and x_2 is

$$\rho_{x_1,x_2} = \frac{\mathrm{cov}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} = \frac{E\left[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})\right]}{\sigma_{x_1}\sigma_{x_2}} \tag{1}$$

where $\mathrm{cov}(x_1, x_2)$ is the covariance between x_1 and x_2, σ_{x_1} is the variance of x_1, σ_{x_2} is the variance of x_2, E is the expected value operator, μ_{x_1} and μ_{x_2} are means of x_1 and x_2. The Pearson's correlation only measure the linear relationship between x_1 and x_2. Further, mutual information [13] is widely used to quantify nonlinear correlation between variables, i.e.,

$$MI_{x_1,x_2} = \sum_{x_2}\sum_{x_1} Pr(x_1, x_2)\log\left(\frac{Pr(x_1, x_2)}{Pr(x_1)Pr(x_2)}\right) \tag{2}$$

In the literature, both Pearson's correlation and mutual information were integrated with hierarchical clustering (HC) [15] for variable clustering. This clustering procedure is either done in the agglomerative way or in the divisive way. For example, each variable is a singleton cluster in the first step of agglomerative HC. Then two closest clusters are merged into one cluster. The recursive merging continues to move up along the hierarchy until the stopping criteria is satisfied, e.g., the maximum number of clusters or the maximum group-average (GA) dissimilarity. The criteria of group average measures the intergroup dissimilarity as the average dissimilarity between two clusters, i.e.,

$$D_{GA}(C_i, C_j) = \frac{1}{N_{C_i}N_{C_j}} \sum_{x_i \in C_i} \sum_{x_j \in C_j} D_{x_i x_j} \tag{3}$$

where N_{C_i} and N_{C_j} are the number of variables in the cluster C_i and C_j, variables x_i and x_j are in the clusters of C_i and C_j, respectively, $D_{x_i x_j}$ is the dissimilarity between variables x_i

and x_j, which is usually calculated as $1-\rho_{x_i,x_j}$ or $1-MI_{x_i,x_j}$ depending on the measurement of dissimilarity.

However, linear correlation cannot adequately capture nonlinear interdependence among variables. Also, both linear correlation and mutual information measure symmetric interdependence between variables. In other words, x_1 and x_2 in equation (1) or equation (2) can be placed interchangeably without impacting the correlation measures. In fact, HC is only applicable when dissimilarity measures are symmetric. However, it is not uncommon that the interdependence structure between two variables are asymmetric, e.g., $\Pr(x_1|x_2) \neq \Pr(x_2|x_1)$. In other words, information transfer between x_1 and x_2 is not necessarily symmetric. The presence of nonlinear and asymmetric interdependence structures poses a significant challenge for variable clustering. Further, HC is not a dynamic approach. In other words, we cannot relocate the variables once the merge is done for two closest clusters. If two variables are 'incorrectly' clustered at the early stage, there is no adaptive step in the later stage to make corrections.

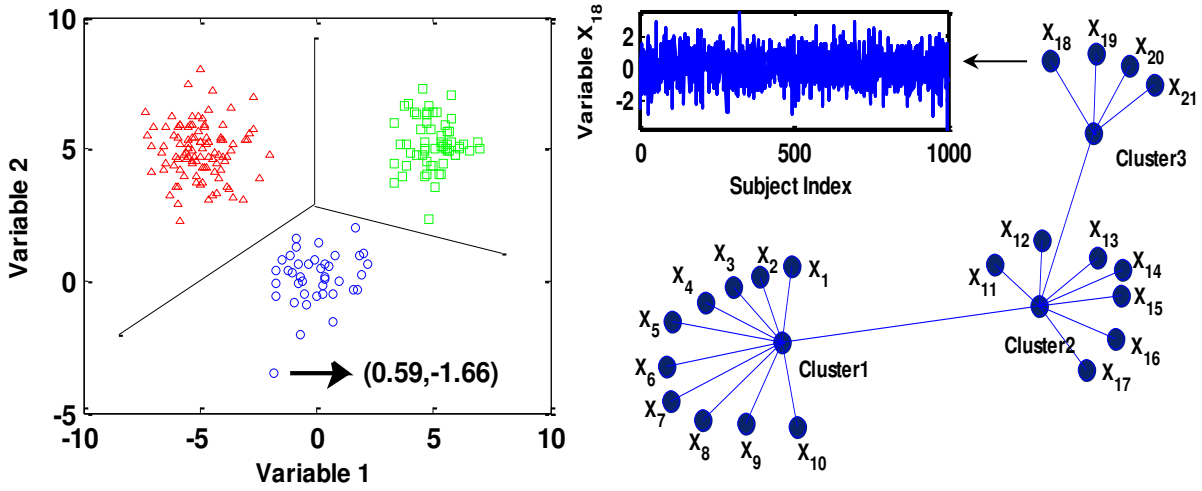


Figure 18 (a) Data clustering with each point representing a data sample and (b) variable clustering with each point representing a variable.

In addition, latent-variable methods such as oblique principal component clustering (OPCC) [14] are widely used for variable clustering. Suppose $\mathrm{X_{n\times p}} = [\mathrm{x_1}, \mathrm{x_2}, \ldots \mathrm{x_p}]$, $\mathrm{x_i} = [\mathrm{x_{i1}}, \mathrm{x_{i2}}, \ldots, \mathrm{x_{in}}]^{\mathrm{T}}$ is the data matrix of n rows representing n observations and p columns representing p variables. Without the loss of generality, we standard the variables in data matrix X to have zero mean and standard deviation. PCA transforms the data matrix into the orthogonal space, where a sparse set of $(\mathrm{q} \leq \mathrm{p})$ principal components (PCs) preserve most of information in original data [16]. These PCs are latent variables which are linear projections of original variables. The k-th $(k = 1,2, \ldots, q)$ PC is calculated as

$$\boldsymbol{X}\boldsymbol{z}_k = \boldsymbol{z}_{k1}\boldsymbol{x}_1 + \boldsymbol{z}_{k2}\boldsymbol{x}_2 + \cdots + \boldsymbol{z}_{kp}\boldsymbol{x}_p = \sum_{j=1}^{p} \boldsymbol{z}_{kj}\boldsymbol{x}_j \tag{4}$$

where $\mathrm{z_k} = [\mathrm{z_{k1}}, \mathrm{z_{k2}}, \ldots, \mathrm{z_{kp}}]^{\mathrm{T}}$ is the k-th eigenvector with a unity norm. The eigenvector $\mathrm{z_k}$ is derived by maximizing the variance of the k-th PC (i.e., $\boldsymbol{X}\boldsymbol{z}_k$), while meeting with the constraints: (1) eigenvectors are orthogonal to each other; (2) PCs are ordered according to the magnitude of variances

$$\begin{gathered} \underset{\boldsymbol{z_k}}{\operatorname{argmax}}\, \boldsymbol{var}\Big(\sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \boldsymbol{z_{kj}}\boldsymbol{x_j}\Big) = \underset{\boldsymbol{z_k}}{\operatorname{argmax}}\, \boldsymbol{z_k^T}\boldsymbol{\Sigma}\boldsymbol{z_k} \\ s.t. \quad \boldsymbol{z_k^T}\boldsymbol{z}_i = 0 \;\; \boldsymbol{z_k^T}\boldsymbol{z}_k = 1 \; (i = 1,2, \ldots, k-1) \end{gathered} \tag{5}$$

where $\Sigma = \mathrm{X^T X}$ is the covariance matrix of X.

Although PCA orthogonalizes the variables and tackles the multicollinearity issue, it is limited in the capability to interpret data matrix in the original input space. Such an interpretation is critical to cluster variables in the input space. Therefore, the OPCC method was further developed to enhance the interpretability of principal components and identify the cluster structure of variables. The OPCC method rotates the eigenvector matrix Z to obtain a new one

$B = Z\Omega$ that has a simple structure. In other words, oblique rotation is aimed at obtaining a sparse matrix B in which most of the elements are close to 0. The Varimax criterion is to find the oblique rotation matrix Ω that maximizes the function:

$$\max_{\Omega} \sum_{i=1}^{p} \left[\sum_{j=1}^{q} {b_{ij}}^4 - \left(\sum_{j=1}^{q} {b_{ij}}^2 \right)^2 \right] \tag{6}$$

where b_{ij} are the element in the ith row and jth column of new loading matrix B. See details of oblique rotation, for example, in [17]. After the oblique rotation, the simple structure of B facilitates the identification of cluster structures of variables. The steps for oblique principal component clustering (OPCC) are as follows:

1) Perform principal component analysis of all variables and find the first two PCs.

2) Oblique rotation of eigenvectors, Z, to obtain the B.

3) Calculate the correlation between all variables and the rotated component, and then assign each variable to one of the two clusters based on the higher squared correlation.

4) Repeat the binary split for each cluster.

5) Stop the recursive split when the second eigenvalue is less than 1.

Many previous studies showed that OPCC achieves better clustering results than HC. However, both methods are based on linear transformation and are limited in the capability to handle nonlinear interdependences among variables. Here, we show a motivating example to evaluate the performance of HC and OPCC for variable clustering. Four clusters of variables are generated as follows:

$\{\boldsymbol{x}_1, \quad \boldsymbol{x}_2 = |\boldsymbol{x}_1|, \qquad \boldsymbol{x}_3 = {\boldsymbol{x}_1}^2, \qquad \boldsymbol{x}_4 = {\boldsymbol{x}_1}^3, \qquad \boldsymbol{x}_5 = {\boldsymbol{x}_1}^4\};$

$\{\boldsymbol{x}_6, \quad \boldsymbol{x}_7 = |\boldsymbol{x}_6|, \qquad \boldsymbol{x}_8 = {\boldsymbol{x}_6}^2, \qquad \boldsymbol{x}_9 = {\boldsymbol{x}_6}^3, \qquad \boldsymbol{x}_{10} = {\boldsymbol{x}_6}^4\};$

$\{\boldsymbol{x}_{11}, \ \boldsymbol{x}_{12} = \boldsymbol{x}_{11}(t+3), \quad \boldsymbol{x}_{13} = \boldsymbol{x}_{11}(t+5), \quad \boldsymbol{x}_{14} = \boldsymbol{x}_{11}(t+7), \quad \boldsymbol{x}_{15} = \boldsymbol{x}_{11}(t+9)\};$

$\{\boldsymbol{x}_{16},\ \boldsymbol{x}_{17} = \boldsymbol{x}_{16}(t+10),\quad \boldsymbol{x}_{18} = \boldsymbol{x}_{16}(t+20),\quad \boldsymbol{x}_{19} = \boldsymbol{x}_{16}(t+30),\quad \boldsymbol{x}_{20} = \boldsymbol{x}_{16}(t+40)\}.$

where x_1 and x_6 are independent standard normal variables, x_{11} is a nonlinear variable sampled from logistic map $x_{11}(n+1) = 3.8x_{11}(n)(1-x_{11}(n))$, x_{16} is a second-order autoregressive variable that is nonlinearly coupled with x_{Lorenz}, $x_{16}(n) = 1.095x_{16}(n-1) - 0.4x_{16}(n-2) + 0.7\varepsilon_n + 0.3x_{Lorenz}^2$, where ε_n is Gaussian noise, x_{Lorenz} is the x-component of a Lorenz system: $x' = 10(y-x), y' = x(28-z) - y, z' = xy - \frac{8}{3}z$ with time step 0.01. The sample size of each variable is 1000.

Fig. 19 (a) shows the dendrogram of hierarchical clustering for the motivating example. Fig. 19 (b) shows the clustering results of OPCC. It may be noted that both HC and OPCC cannot identify the cluster structure of variables. This is mainly due to the fact that nonlinear and asymmetric interdependence structures among variables are not considered. Very little work has been done to cluster a large number of variables with complex structures of nonlinear and asymmetric interdependences. In order to tackle these issues and fill the gap, we propose a new strategy that integrates nonlinear coupling analysis with self-organizing networks for variable clustering and predictive modeling.

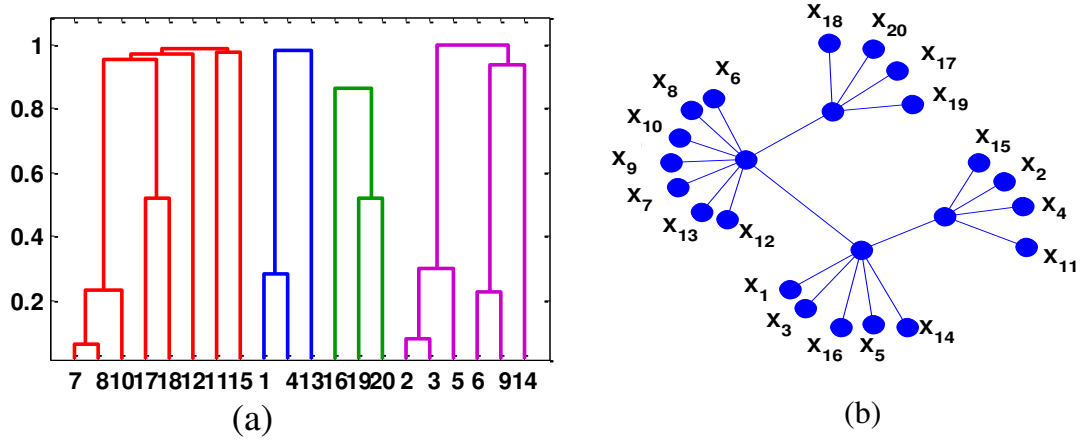


Figure 19 Clustering results for motivating example using (a) HC and (b) OPCC.

4.3 Research Methodology

In this section, we will first introduce nonlinear coupling analysis to characterize and measure nonlinear interdependence structures among variables. Second, we develop a self-organizing network algorithm to cluster variables that involve nonlinear and asymmetric interdependences. Finally, we orthogonalize variables in each of the self-organized clusters and then integrated them with group elastic-net model to improve the performance of predictive modeling. We will illustrate the proposed methodology using the motivating example introduced in section II.

4.3.1 Nonlinear Coupling Analysis

In this investigation, we propose to characterize and quantify nonlinear interrelationship among variables. Traditionally, such interrelationships are estimated with methods such as correlation and mutual information. As aforementioned, correlation is a second-order quantity evaluating merely linear dependency among data. Mutual information quantifies both linear and nonlinear dependency between variables but requires stationarity in the computation. Both of them are limited in the ability to handle nonlinear and asymmetric interdependence structures.

Therefore, we performed nonlinear coupling analysis by exploiting cross recurrences between two variables in the feature space [18, 19]. Denote $\mathcal{T}(x_1(m))$ as the recurrence neighborhood of $x_1(m)$ containing the k closest neighbors of$x_1(m)$. Let us assume that their indices are $m \in \{n_1, n_2, \dots, n_k\}$. If there are some relations between two variables x_1 and x_2, then the recurrence of x_1 will also imply a recurrence of x_2, at least with a greater than zero probability. In other words, $x_2(m)$ with the same indices $m \in \{n_1, n_2, \dots, n_k\}$ should also be closer to $x_2(m)$ than a average randomly chosen vector. Therefore, the quantity of nonlinear interdependence is defined as:

$$I_{x_1x_2} = \langle \frac{r_m(\boldsymbol{x}_2) - d_m(\boldsymbol{x}_2|\boldsymbol{x}_1)}{r_m(\boldsymbol{x}_2) - d_m(\boldsymbol{x}_2)} \rangle_n \tag{7}$$

where $\langle \cdot \rangle_{\mathrm{m}}$ is the ergodic average over the range of variable $\boldsymbol{x}_2$, and $d_m(\boldsymbol{x}_2|\boldsymbol{x}_1)$ is the average conditional distance:

$$d_m(\boldsymbol{x}_2|\boldsymbol{x}_1) = \frac{1}{k} \sum_{i \in \{n_1, n_2, \ldots, n_l\}} (\boldsymbol{x}_2(m) - \boldsymbol{x}_2(i))^2 \tag{8}$$

where $\{n_1, n_2, \ldots, n_l\}$ is the indices determined from the variable $\boldsymbol{x}_1$. In addition, $d_m(\boldsymbol{x}_2)$ is the average distance of the *k* closest neighbors of $\boldsymbol{x}_2(m)$:

$$d_m(\boldsymbol{x}_2) = \frac{1}{k} \sum_{\boldsymbol{x}_2(i) \in \mathcal{V}(\boldsymbol{x}_2(m))} (\boldsymbol{x}_2(m) - \boldsymbol{x}_2(i))^2 \tag{9}$$

and $\mathcal{V}(\boldsymbol{x}_2(m))$ is the true neighborhood of $\boldsymbol{x}_2(m)$. The average distance of *k* randomly chosen $\boldsymbol{x}_2(i)$ to $\boldsymbol{x}_2(m)$ is $r_m(\boldsymbol{x}_2) = \langle [\boldsymbol{x}_2(m) - \boldsymbol{x}_2(i)]^2 \rangle_k$. If $I_{x_1x_2}$ is small (close to zero), then there is no evident interdependence between variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, because the true neighbors of $\boldsymbol{x}_2(m)$ are much closer to $\boldsymbol{x}_2(m)$ than those neighbors based on the recurrences in the $\boldsymbol{x}_1$ process. When $I_{x_1x_2}$ is close to unity, there is a strong interdependence between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Fig. 20 shows the matrices of both linear correlations and nonlinear interdependences among variables that are computed from the motivating example. The red color represents a high interdependence, while the blue color indicates no interdependence. Note that nonlinear interdependence in Fig. 20 (b) is significantly different from linear correlation in Fig. 20 (a). Nonlinear coupling analysis provides a better characterization of complex interdependence structure (i.e., nonlinear and asymmetric) among variables than linear correlations. It is remarkable that the nonlinear interdependence reveals some interdependent structures that are not found by correlation.

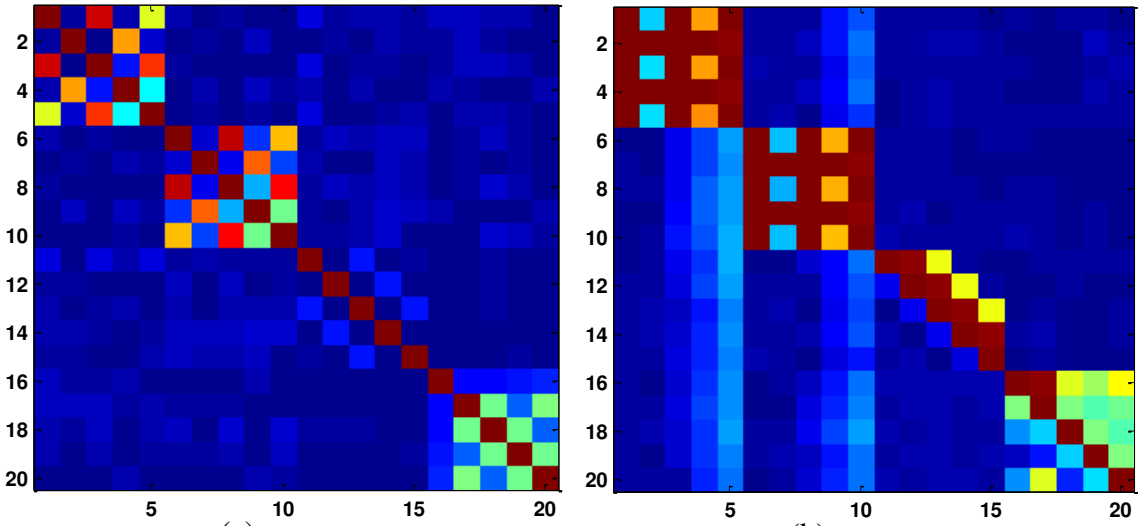


Figure 20 Matrices of (a) linear correlations and (b) nonlinear interdependences.

4.3.2 Self-organizing Network for Variable Clustering

Fig. 20 (b) shows that nonlinear interdependence is not symmetric. Traditional similarity-based clustering algorithms are not applicable. Latent-variable methods using oblique PCA or factor analysis do not fully consider nonlinear interdependences among variables. To tackle these challenges, we develop a self-organizing network algorithm to cluster variables that involve nonlinear and asymmetric interrelationships. Notably, this present investigation extends our previous work from self-organizing topology of recurrence networks [20] to self-organizing clustering of highly-redundant variables.

In the literature, very little work has been done to cluster variables with complex nonlinear and asymmetric interdependences. We propose to treat variables as nodes in the network and the nonlinear interdependences between variables are treated as the weights of links, which is varying from 0 to 1. Let $G = \{V, E\}$ be the directed and weighted network, where V is the set of nodes and E is the set of edges. The spring-electrical model assigns two forces, i.e., attractive and repulsive forces between nodes. The repulsive force exists between any pair of

nodes while the attractive force exists between the nodes that have a higher relation characterized by nonlinear interdependence.

The repulsive force is defined as

$$f_r(i,j) = -\frac{1}{\|\boldsymbol{s}(i) - \boldsymbol{s}(j)\|^2} * \frac{1}{e^{\alpha\left|I_{x_i x_j}\right|}} \tag{10}$$

where α is a system parameter, $\boldsymbol{s}(\mathrm{i})$ and $\boldsymbol{s}(\mathrm{j})$ are spatial locations of node i and node j. The repulsive force is inversely proportional to the nonlinear interdependence between two nodes (variables), because a bigger repulsive force is expected to separate two nodes when they have a smaller interdependence. The attractive force is defined as

$$f_a(i,j) = \|\boldsymbol{s}(i) - \boldsymbol{s}(j)\|^2 * e^{\gamma\left|I_{x_i x_j}\right|}, I_{\boldsymbol{x}_i \boldsymbol{x}_j} \neq 0 \tag{11}$$

where γ is the system parameter. The attractive force is proportional to the nonlinear interdependence between two nodes (variables), because a bigger attractive force will pull two nodes closer when they have a higher interdependence. The combined force on a node i is the summation of all repulsive forces and attractive forces on the node:

$$\begin{aligned} f(i,\boldsymbol{s},\alpha,\gamma) = \sum_{i \neq j} -\frac{\frac{1}{e^{\alpha\left|I_{x_i x_j}\right|}}}{\|\boldsymbol{s}(i) - \boldsymbol{s}(j)\|^3}\left(\boldsymbol{s}(i) - \boldsymbol{s}(j)\right) \\ + \sum_{i \leftrightarrow j} e^{\gamma\left|I_{x_i x_j}\right|}\|\boldsymbol{s}(i) - \boldsymbol{s}(j)\|\left(\boldsymbol{s}(i) - \boldsymbol{s}(j)\right) \end{aligned} \tag{12}$$

where $\boldsymbol{s}(\mathrm{i}) - \boldsymbol{s}(\mathrm{j})$ is the force-directional vector, which is separated from $\mathrm{f_r(i,j)}$ and $\mathrm{f_a(i,j)}$ to define the direction of combined force $\mathrm{f(i,}\boldsymbol{s}\mathrm{,}\alpha\mathrm{,}\gamma\mathrm{)}$. The attractive and repulsive forces drive the network to self-organize and form a topological structure. The objective of self-organizing process is to identify spatial locations of nodes by minimizing the energy of the network, i.e., the summation of squared combined forces on each node:

$$\boldsymbol{s}* = \arg\min_{\boldsymbol{s}} \left\{ \sum_{i \in \boldsymbol{N}} f^2\,(i, \boldsymbol{s}, \alpha, \gamma) \right\} \tag{13}$$

The self-organizing process derives a topological structure of the network of variables by minimizing the total energy, thereby clustering homogeneous variables into sub-network communities. It is remarkable that the variations of system parameters α and γ will not change the structure of the clustering but yield a similar structure only in different scales. Our previous investigation detailed the algorithms to derive the self-organizing network topology by minimizing the total energy function [20].

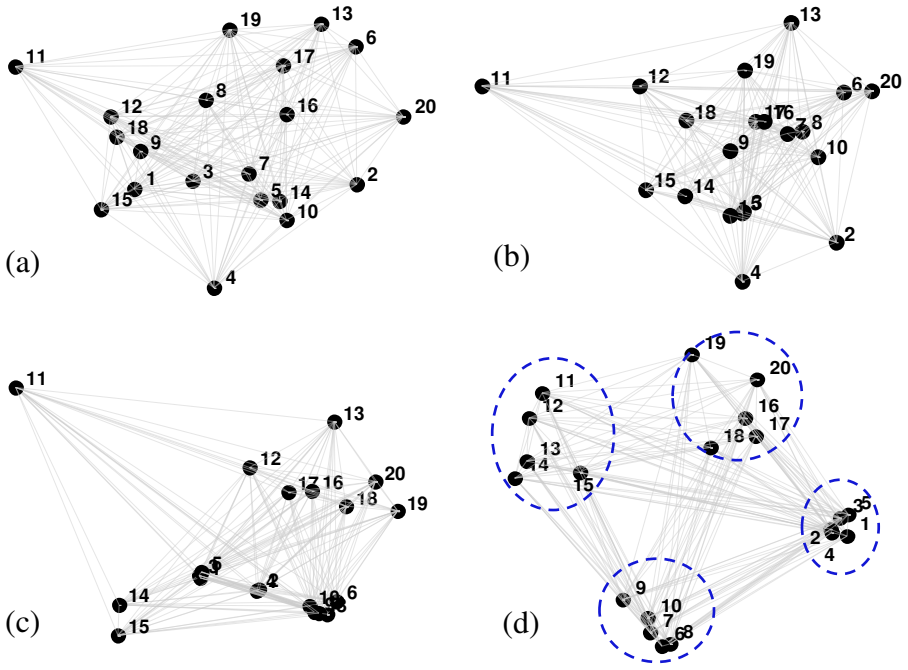


Figure 21 Self-organizing network for clustering 20 variables: (a) Initial topological structure, (b) topological structure after 200 iterations, (c) topological structure after 400 iterations and (d) final topological structure.

Fig. 21 shows the self-organizing organizing process for clustering 20 variables in the motivating example. At the beginning, 20 variables are randomly distributed in a 3-dimensional space (see Fig. 21 (a)). The topological structures after 200 and 400 iterations are shown in Fig.

21 (b) and (c). The final structure after 600 iterations is shown in Fig. 21 (d). After 600 iterations, the self-organizing process converges and identifies the underlying cluster structures of 20 variables, which demonstrates the superior performance of self-organizing networks over previous methods.

Our proposed self-organizing network shares some similarities with minimum energy design [21] in the field of design of experiments and self-organizing map [22, 23] in the domain of neural network. However, the spatial location of a design point will not change in the minimum energy design when the experiment has been conducted at this setting. The algorithm will optimize the spatial location of the next design point given that spatial locations of previous design points are fixed. In addition, the proposed approach of self-organizing variable clustering is vastly different from self-organizing map in neural network, which learns self-organizing positions of neurons based on distance measures in the data space. Nonetheless, our proposed research seeks to self-organize the data space of variables.

4.3.3 Predictive Modeling with Highly-redundant Variables

The self-organizing network drives highly-redundant variables into sub-network clusters. The variables in each cluster bring the redundant information. It is necessary to delineate the structure of latent variables hidden in each cluster of homogeneous variables. As such, we propose to minimize the redundancy within the same cluster before grouped variables are used in predictive models. Assume we have M clusters and there are K variables, $x_{m1}, x_{m2}, \dots, x_{mK_m}$, in the m-th cluster. The Gram-Schmidt orthonormalization (GSO) minimizes the redundant information by transforming original variables ($x_{m1}, x_{m2}, \dots, x_{mK_m}$) into orthonormal set of new variables ($w_{m1}, w_{m2}, \dots, w_{mK_m}$) in each cluster. It begins by normalizing x_{m1},

$$\mathrm{v_{m1}} = \mathrm{x_{m1}}; \quad \mathrm{w_{m1}} = \frac{\mathrm{v_{m1}}}{\|\mathrm{v_{m1}}\|} \tag{14}$$

where $\mathrm{w_{m1}}$ is the normalized variable of $\boldsymbol{x}_{\mathrm{m1}}$. The second orthogonal vector $\mathrm{v_{m2}}$ is obtained as denoted by,

$$\boldsymbol{v}_{m2} = \boldsymbol{x}_{m2} - \langle \boldsymbol{x}_{m2}, \boldsymbol{w}_{m1} \rangle \boldsymbol{w}_{m1}; \quad \boldsymbol{w}_{m2} = \frac{\boldsymbol{v}_{m2}}{\|\boldsymbol{v}_{m2}\|} \tag{15}$$

where $\boldsymbol{w}_{\mathrm{m2}}$ is the second orthonormalized vector. The process is repeated to obtain the *k*-th orthogonal vector $\boldsymbol{v}_{\mathrm{mk}}$

$$\boldsymbol{v}_{mk} = \boldsymbol{x}_{mk} - \sum_{i=1}^{k-1} \langle \boldsymbol{x}_{mk}, \boldsymbol{w}_{\boldsymbol{mi}} \rangle \boldsymbol{w}_{\boldsymbol{mi}}; \quad \boldsymbol{w}_{mk} = \frac{\boldsymbol{v}_{mk}}{\|\boldsymbol{v}_{mk}\|} \tag{16}$$

where $\mathrm{w_{mk}}$ is the k-th orthonormalized vector. Then we leverage orthonormalized variables in each cluster to develop a group elastic-net model:

$$\max_{\beta} \sum_{i=1}^{n} [y_i \log\left(h_\beta(w,i)\right) + (1-y_i)\log(1-h_\beta(w,i))]^2$$

$$h_\beta(w,i) = \frac{1}{1+\exp\left[-\left(\beta_0 + \left(\sum_{m=1}^{M}\sum_{k=1}^{\mathrm{K_m}} w_k(i)\beta_k\right)\right)\right]} \tag{17}$$

$$s.t. \quad \sum_{m=1}^{M}\sum_{k}^{K_m}(\alpha \boldsymbol{\beta}_{mk}{}^2 + (1-\alpha)|\boldsymbol{\beta}_{mk}|) \leq \lambda$$

where θ and λ are penalization parameters. Further, we will evaluate and validate the proposed approach using the experimental study, as detailed in the next section.

4.4 Simulation Study

In this section, simulation experiments were designed to evaluate the performance of self-organizing variable clustering (SOC) algorithms. On the basis of 20 variables in the motivating example, we further utilized 3 b-spline basis functions to expand each variable to 3 derived variables. Therefore, a total of 60 variables are generated in 4 groups to evaluate the proposed methodology.

4.4.1 Data Generation

The logistic function $h_\beta(x)$ is computed using a sparse set of variables selected from the original 20 variables:

$$h_\beta(x) = \frac{1}{1+exp[-(\sum_{i=1}^{20\times p_s} x_i\beta_i + c\varepsilon)]} \quad (18)$$

where p_s denotes the sparsity level (i.e., the percentage of variables involved to derive the response), c is the amplitude of random noises, and β_i is the model parameter that follows the distribution of $\mathrm{N}(\mu_\mathrm{j}, 0.1)$, $\mu_\mathrm{j} = 1 + (\mathrm{j} - 1) \times 0.3$, $\mathrm{j} = 1,2,\ldots,20$. The decision boundary $h_\beta(x) = 0.5$ is used to generate the binary response variable y.

4.4.2 Experimental Design

As shown in Fig. 22, a 3-way layout experiment was designed to evaluate the performance of SOC algorithms with three factor groups, i.e., signal-noise-ratio (SNR), sparsity, and sample size. The sample sizes of training set n_1, validation set n_2 and testing set n_3 are varied in three levels (i.e., 50/50/100, 100/100/200 and 200/200/400). The sparsity level p_s is changed from 10%, 40% to 70%. The SNR level is varied from no noise, 1, 2, to 3, where SNR is the power ratio between the signal and the background noise, i.e., $\mathrm{var}(\sum_{i=1}^{20\times p_s} \boldsymbol{x}_i\beta_i)/\mathrm{var}(c\varepsilon)$. As a result, we generated 36 treatment levels of experimental factors and further evaluate and validate the performance of clustering algorithms. In this present investigation, we compared the performance of self-organizing variable clustering (SOC) with no clustering, hierarchical clustering (HC) and Oblique Principal Components Clustering (OPCC). The training dataset is used to train the predictive model, and the validation dataset is to optimally select the penalization parameters in equation (17). Model performance is only computed from the testing dataset. Each treatment level is replicated for 100 times.

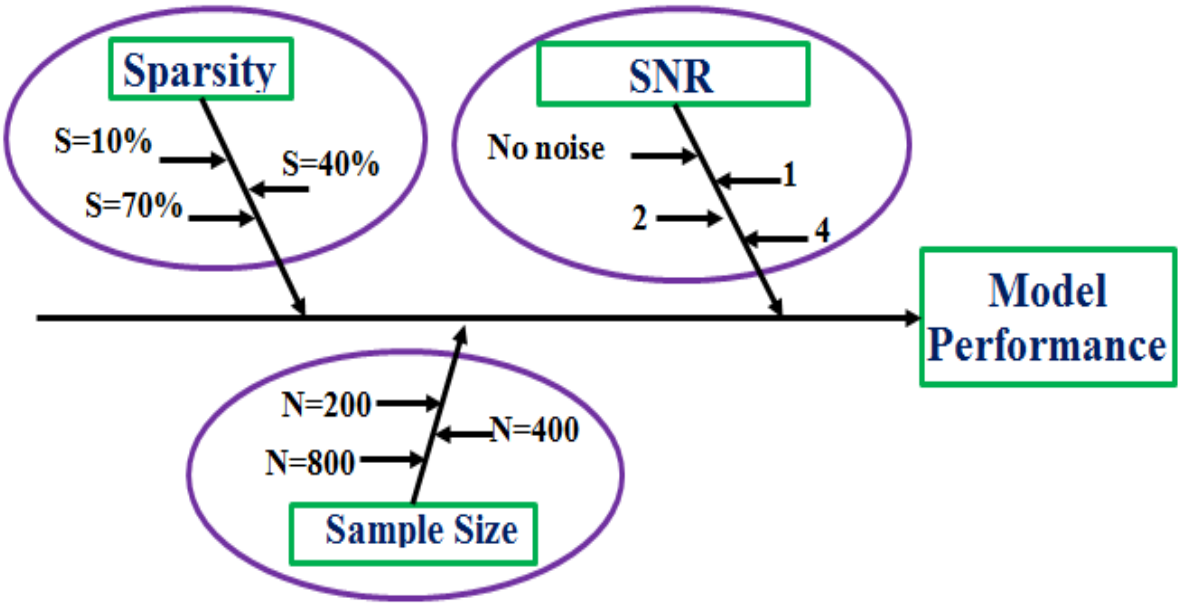


Figure 22 Cause-and-effect diagram for performance evaluation of the proposed self-organized variable clustering algorithm.

4.4.3 Predictive Modeling with Variable Clustering

Table III summarizes the averages and standard deviations of prediction errors from 100 replicates with GSO of variables in each cluster. Note that the numbers in the parenthesis are the standard errors over 100 replicates. As shown in Table 1, all variable clustering methods yield better performance than no clustering of variables. In particular, self-organizing variable clustering outperforms the other two methods and achieves a relatively better performance. No clustering yields the highest prediction errors at all the 36 treatment levels. The comparison results are not surprising because Fig.2 and Fig.4 show that self-organizing clustering identifies cluster structures better than HC and OPCC. The proposed self-organizing algorithm achieves better performance in 23 out 36 treatment levels in the experiments. Characterizing nonlinear interdependence structures among variables helps improve the performance of predictive modeling. In addition, Table 3 shows that the prediction error decreases as more data samples are available for variable clustering and predictive modeling. Adding noises to data deteriorates the

predictive performance of models. Finally, it is worth mentioning that model performance is better when the sparsity level is lower.

Table 3 Averages and standard deviation of prediction errors for simulation study from 100 replications with Gram-Schmidt orthonormalization

SNR	Sample Size	No Clustering			GSO-HC			GSO -OPCC			GSO-SOC		
		S=0.1	S=0.4	S=0.7	S=0.1	S=0.4	S=0.7	S=0.1	S=0.4	S=0.7	S=0.1	S=0.4	S=0.7
1	n1 = 50, n2=50,n3=100	0.361	0.489	0.508	0.278	0.413	0.406	0.259	**0.335**	0.381	**0.252**	0.340	**0.381**
		(0.06)	(0.05)	(0.05)	(0.05)	(0.09)	(0.09)	(0.05)	**(0.07)**	(0.09)	**(0.03)**	(0.09)	**(0.06)**
	n1 = 100, n2=100,n3=200	0.351	0.456	0.430	0.276	0.304	0.299	**0.251**	**0.301**	**0.295**	0.278	0.335	0.307
		(0.05)	(0.04)	(0.06)	(0.03)	(0.04)	(0.04)	**(0.03)**	**(0.04)**	**(0.03)**	(0.03)	(0.06)	(0.03)
	n1 = 200, n2=200,n3=400	0.283	0.328	0.337	0.270	0.298	0.280	0.264	0.294	**0.255**	**0.259**	**0.266**	0.283
		(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	**(0.02)**	**(0.02)**	**(0.02)**	(0.02)
2	n1 = 50, n2=50,n3=100	0.203	0.345	0.438	0.157	0.271	0.263	0.161	0.280	0.255	**0.120**	**0.192**	**0.230**
		(0.05)	(0.11)	(0.08)	(0.03)	(0.10)	(0.08)	(0.03)	(0.08)	(0.05)	**(0.02)**	**(0.05)**	**(0.05)**
	n1 = 100, n2=100,n3=200	0.214	0.263	0.284	0.147	0.171	0.188	0.154	**0.165**	0.194	**0.136**	0.199	**0.180**
		(0.02)	(0.03)	(0.05)	(0.02)	(0.02)	(0.03)	(0.02)	**(0.03)**	(0.03)	**(0.02)**	(0.03)	**(0.02)**
	n1 = 200, n2=200,n3=400	0.175	0.236	0.227	0.154	0.171	0.174	**0.130**	0.172	**0.167**	0.131	**0.170**	0.185
		(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	**(0.01)**	(0.02)	**(0.01)**	(0.01)	**(0.01)**	(0.02)
3	n1 = 50, n2=50,n3=100	0.197	0.355	0.436	0.116	0.169	0.240	0.098	0.210	0.247	**0.097**	**0.165**	**0.236**
		(0.04)	(0.10)	(0.08)	(0.03)	(0.05)	(0.07)	(0.03)	(0.08)	(0.05)	**(0.03)**	**(0.04)**	**(0.06)**
	n1 = 100, n2=100,n3=200	0.156	0.262	0.269	0.100	0.154	0.168	0.105	0.137	0.158	**0.099**	**0.135**	**0.157**
		(0.02)	(0.03)	(0.04)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	**(0.02)**	**(0.02)**	**(0.02)**
	n1 = 200, n2=200,n3=400	0.133	0.194	0.187	0.084	0.138	0.136	0.088	0.150	**0.114**	**0.072**	**0.116**	0.133
		(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	**(0.01)**	**(0.01)**	**(0.01)**	(0.01)
No Noise	n1 = 50, n2=50,n3=100	0.127	0.304	0.345	0.058	0.166	0.151	**0.051**	0.135	0.146	0.065	**0.109**	**0.145**
		(0.03)	(0.10)	(0.09)	(0.02)	(0.08)	(0.05)	**(0.03)**	(0.08)	(0.03)	(0.02)	**(0.04)**	**(0.04)**
	n1 = 100, n2=100,n3=200	0.109	0.199	0.206	0.043	0.063	0.101	**0.043**	**0.056**	**0.095**	0.052	0.062	0.099
		(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	**(0.01)**	**(0.02)**	**(0.03)**	(0.02)	(0.02)	(0.03)
	n1 = 200, n2=200,n3=400	0.072	0.145	0.163	0.026	0.054	0.067	0.030	0.056	0.066	**0.023**	**0.044**	**0.066**
		(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.018)	(0.02)	**(0.01)**	**(0.01)**	**(0.02)**

Table 4 shows the averages and standard deviations (i.e., the numbers in the parenthesis) of prediction errors from 100 replicates in each treatment level with PCA of variables in each cluster. Similar to the results in Table I, the proposed SOC yields better performance than HC and OPCC algorithms in 23 out of 36 treatment levels in the experiments. The performances of HC and OPCC are close to each other, which is because linear correlation and PCA processing of variables are utilized in both cases. In terms of experimental factors (i.e., sample size, noise and

sparsity), Table II shows consistent results as in Table I. The prediction error decreases as the sample size increases. Adding noise to data deteriorates the predictive performance of models. When the sparsity level is lower, the model performance is better.

Table 4: Averages and standard deviation of prediction errors for simulation study from 100 replications using the principal components

SNR	Sample Size	PCA-HC			PCA-OPCC			PCA-SOC		
		S=0.1	S=0.4	S=0.7	S=0.1	S=0.4	S=0.7	S=0.1	S=0.4	S=0.7
1	n1 = 50, n2=50,n3=100	**0.2228**	0.3316	**0.3690**	0.2442	0.3963	0.4271	0.2530	**0.3298**	0.3938
		(0.033)	(0.072)	**(0.071)**	(0.033)	(0.080)	(0.066)	(0.038)	**(0.078)**	(0.085)
	n1 = 100, n2=100,n3=200	**0.2466**	**0.2750**	0.3359	0.3266	0.3727	0.3070	0.2470	0.3279	**0.2888**
		(0.027)	**(0.024)**	(0.051)	(0.045)	(0.063)	(0.0364)	(0.027)	(0.049)	**(0.032)**
	n1 = 200, n2=200,n3=400	0.2621	0.2927	0.2844	0.2220	**0.2740**	**0.2733**	**0.2151**	0.2957	0.2809
		(0.018)	(0.017)	(0.021)	(0.016)	**(0.019)**	**(0.016)**	**(0.015)**	(0.019)	(0.020)
2	n1 = 50, n2=50,n3=100	0.1542	0.2362	0.2623	0.1577	0.2364	**0.2175**	**0.1398**	**0.2269**	0.2659
		(0.032)	(0.074)	(0.065)	(0.031)	(0.056)	**(0.047)**	**(0.028)**	**(0.065)**	(0.062)
	n1 = 100, n2=100,n3=200	0.1518	0.2039	0.2007	0.1500	0.1759	**0.1834**	**0.1433**	**0.1600**	0.2082
		(0.021)	(0.026)	(0.028)	(0.020)	(0.025)	**(0.023)**	**(0.018)**	**(0.025)**	(0.037)
	n1 = 200, n2=200,n3=400	0.1395	0.1633	0.1698	0.1377	0.1852	0.1671	**0.1327**	**0.1577**	**0.1593**
		(0.014)	(0.015)	(0.015)	(0.014)	(0.017)	(0.018)	**(0.012)**	**(0.014)**	**(0.015)**
3	n1 = 50, n2=50,n3=100	0.1064	0.2010	0.2428	0.1253	**0.1518**	0.2409	**0.0960**	0.1653	**0.2330**
		(0.028)	(0.053)	(0.054)	(0.023)	**(0.053)**	(0.052)	**(0.026)**	(0.047)	**(0.057)**
	n1 = 100, n2=100,n3=200	0.0966	0.1378	0.1877	**0.0956**	**0.1352**	0.1710	0.1097	0.1419	**0.1691**
		(0.019)	(0.023)	(0.023)	**(0.017)**	**(0.020)**	(0.024)	(0.018)	(0.021)	**(0.027)**
	n1 = 200, n2=200,n3=400	0.0838	0.1261	0.1338	**0.0779**	**0.1236**	0.1397	0.0937	0.1317	**0.1255**
		(0.011)	(0.012)	(0.015)	**(0.011)**	**(0.013)**	(0.014)	(0.012)	(0.014)	**(0.014)**
No Noise	n1 = 50, n2=50,n3=100	0.0576	0.1458	0.1643	0.0605	0.1228	0.1607	**0.0547**	**0.1178**	**0.1470**
		(0.020)	(0.034)	(0.033)	(0.023)	(0.060)	(0.035)	**(0.022)**	**(0.061)**	**(0.039)**
	n1 = 100, n2=100,n3=200	0.0429	0.0686	0.1125	0.0425	0.0684	0.1052	**0.0388**	**0.0506**	**0.0982**
		(0.013)	(0.020)	(0.025)	(0.015)	(0.021)	(0.025)	**(0.013)**	**(0.019)**	**(0.030)**
	n1 = 200, n2=200,n3=400	0.0255	0.0515	0.0647	0.0289	0.0468	0.0612	**0.0205**	**0.0461**	**0.0610**
		(0.0085)	(0.014)	(0.018)	(0.010)	(0.013)	(0.015)	**(0.008)**	**(0.011)**	**(0.016)**

In addition, if we compare the GSO with PCA-based orthogonalization of variables within each cluster, it may be noted that their performance are similar. For 36 treatment levels, the GSO-SOC approach yields 18 setting that have lower prediction errors than the PCA-SOC approach. The differences between GSO-SOC and PCA-SOC approaches are not statistically significant. However, both Table I and Table II show that the self-organizing network algorithm

significantly decreases the errors in predictive models. This demonstrated that variable clustering is critical to improving the performance of predictive models. Further, experimental results showed variable clustering that considers complex interdependence structures (e.g., nonlinear and asymmetric) among variables yields better results than only considering linear correlations.

4.5 Case Study

Furthermore, we evaluated and validated the proposed methodology using a real-world case study that extracts parameters from representation models of vectorcardiogram (VCG) signals for the identification of myocardial infarctions. This present paper is an extension of our previous work that developed model-based representation of VCG signals using multiscale adaptive basis functions [24]. As shown in Fig. 23 (a) cardiac electrical activity is varying across space and time. VCG signals monitor cardiac electrical activity along three orthogonal X, Y, Z planes of the body, namely, frontal, transverse, and sagittal (see Fig. 23 (b)) [25-27]. Within one cycle, the VCG waveform shows nonlinear variations and different segments change significantly that correspond to different stages of cardiac operations. Between cycles, the VCG waveform is similar to each other but with variations. As shown in Fig. 23 (b), VCG trajectories of myocardial infarction (red/dashed) yield a different spatial path from the health controls (blue/solid) [28]. In order to reduce large amounts of VCG signals into a sparse set of parameters, we developed the basis function representation of VCG signals (see Fig. 23 (c)). This present paper leverages parametric features for predicting the incidence of myocardial infarction (see Fig. 23 (d) and (e)). Because the set of parametric features contains redundant information that inflates the variance of predictive models, this motivates our further development of the proposed methodology of self-organizing variable clustering. The results demonstrated the effectiveness and robustness to improve the prediction performance.

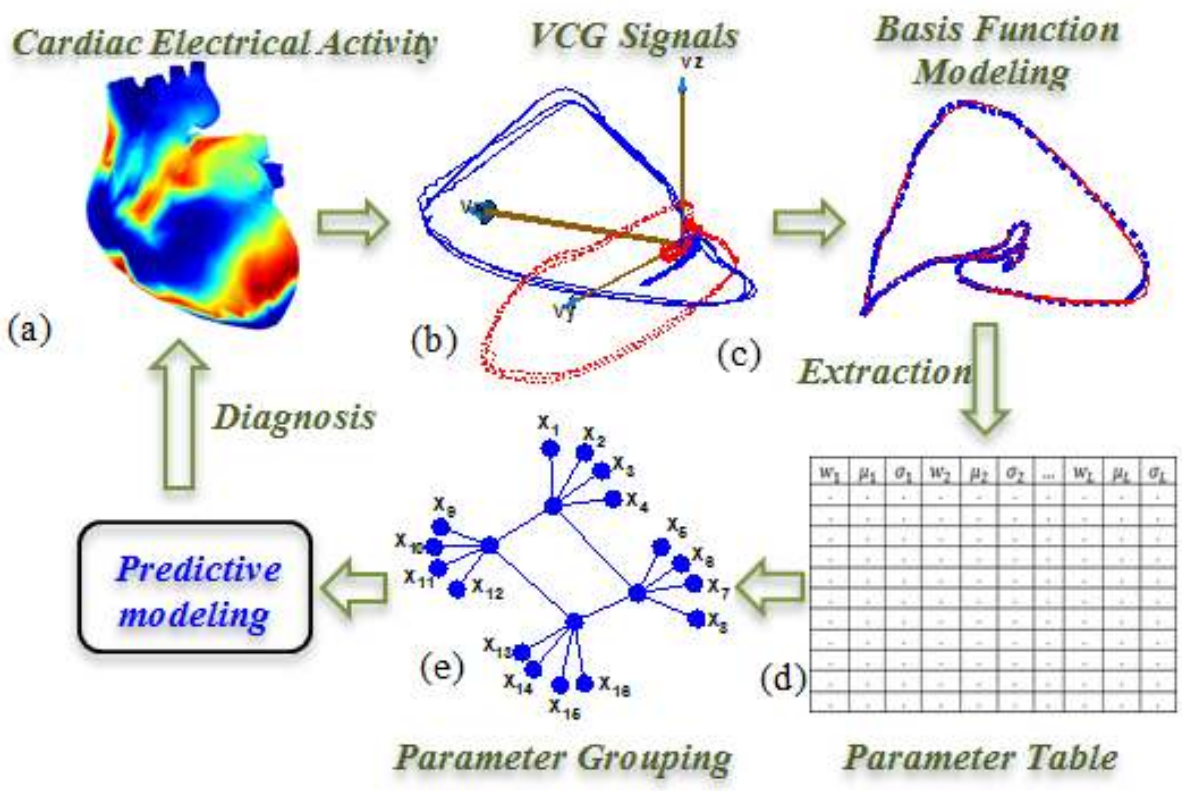


Figure 23 Flow chart of a real-world case study that extracts model parameters from VCG signals for the identification of myocardial infarctions.

4.5.1 Multiscale Adaptive Basis Function Modeling of VCG Signals

Our previous work developed a sparse basis function model to characterize and represent 3-dimensional VCG signals [24]. Such a sparse representation reduces large amounts of data to a limited number of model parameters while preserving the same information. This present paper will further develop predictive models of myocardial infarctions using the low-dimensional set of model parameters, as opposed to the original data itself. Fig. 6 (c) shows an example of the basis function model of 3D trajectories of VCG signals. In order to capture intrinsic characteristic of cardiac electrical activity, we modeled VCG signals as the superposition of M basis functions:

$$v(t,w) = w_0 + \sum_{j=1}^{M} w_j \psi_j\big((t-\mu_j)/\sigma_j\big) + \varepsilon \tag{19}$$

where $\psi(t)$ is the basis function, w_j is the weight factor, μ_j is the shifting factor and σ_j is the scaling factor. The objective is to optimize the representation of 3D VCG signals with a sparse basis function model:

$$\text{argmin}\left[\left\|v(t) - w_0 - \sum_{j=1}^{M} w_j\psi_j(t)\right\|^2, \{w, M, \psi(t)\}\right] \quad (20)$$

Compact topological representation calls the minimization of the number of basis functions M and the optimal placement of basis function $\psi(t)$. Model parameters w, μ, σ are adaptively estimated by the "best matching" projections of VCG signals onto a dictionary of nonlinear basis functions. Our previous work detailed the modeling algorithms to develop a sparse basis function representation of spatiotemporal VCG signals [24]. In addition, our previous experiments show that model goodness-of-fit is greater than 99.9% (R^2) with a parsimonious set of 20 basis functions for a variety of cardiac conditions. In this present study, model parameters, i.e., weight, shifting, scaling factors and residuals will be further investigated for the identification of myocardial infarctions.

4.5.2 Predictive Modeling of Myocardial Infarction

This case study focuses on the extraction of parametric features from the sparse basis function model, and their further applications for predictive modeling of myocardial infarctions. If M basis are used to represent 3-lead VCG signals, then the set of parameters (i.e., weight, shifting and scaling factors) is $\{w_{3\times M}, \mu_{3\times M}, \sigma_{3\times M}\}$. The total number of parameters will be $3 \times 3 \times M$. Our previous study showed that 20 basis functions yield >99.9% goodness-of-fit in the modeling performance for a variety of cardiac conditions. Hence, we have a total of 180 parameters that are adaptively estimated from the 3D VCG trajectory. In addition, we added the absolute values of weights, residual sum of squares (RSS) and the RR interval (i.e., heart rate) in

this present investigation. Thus the feature matrix is:

$$\boldsymbol{X} = \{\boldsymbol{w}_{3\times 20,}\boldsymbol{\mu}_{3\times 20,}\boldsymbol{\sigma}_{3\times 20,}|\boldsymbol{w}|_{3\times 20,}\boldsymbol{RSS}_{3\times 1}, RR_{1\times 1}\}$$

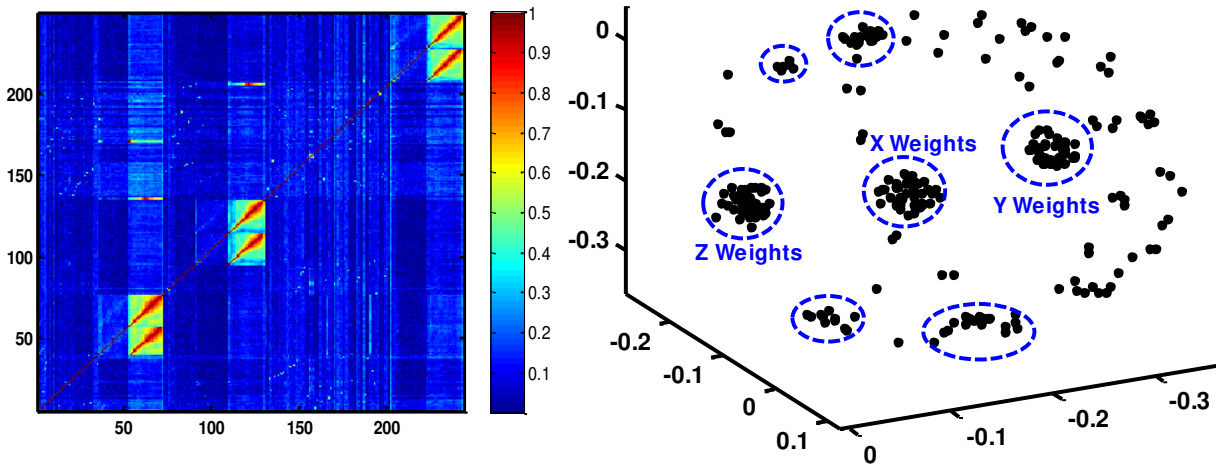
where the absolute values of weights $|\mathrm{w}|_{3\times 20}$ describes the amplitudes of each basis function, which provide local strengths of a heartbeat. The residual sum of squares $\boldsymbol{RSS}_{3\times 1}$ describes the variations that cannot be adequately explained by the model representation. The RR interval characterizes temporal beat-to-beat variations of cardiac electrical activity. In total, we have 244 parameter-based features that provide effective measures of original VCG signals. Notably, model representation reduces the high-dimensional set of VCG signals with a big amount of data points into a sparse set of feature matrix.

In this present investigation, we used the 3-lead VCG signals from 388 subjects (79 controls and 309 faults), available in the PhysioNet Database [29]. VCG signals were digitized at 1 kHz sampling rate with a 16-bit resolution over a range of 16.384 mv.

Our previous study showed that most of model-driven parametric features are statistically significant between healthy controls and myocardial infarctions [30]. Specifically, our experimental results showed that more than 146 features have the Kolmogorov-Smirnov statistic greater than the critical value 0.17, indicating significantly differences between control and diseased conditions [30].

It is worth mentioning that weight factors are the most significant group of features among all parametric features. However, a large number of predictors tend to bring the "curse of dimensionality" problem, as well as the overfitting for the predictive modeling. Therefore, our previous study utilized the lasso-penalized logistic regression model to shrink the number of predictors and identify the cases of myocardial infarction [30]. Nonetheless, our previous study [30] focused on the relevancy between predictor variables and the response variables, but did not

specifically consider complex interdependence structures among predictor variables. Prior research showed that a higher correlation (>0.90) between variables (collinearity) leads to more sensitive estimations of parameters in predictive models (i.e., increased variances of estimation) [6]. This present paper further investigates the nonlinear correlations between variables and then identifies the cluster structures for improving the performance of predictive modeling. Fig. 24 (a) shows the plot of nonlinear and asymmetric interdependence structures among variables. It may be noted that there are three groups of variables with stronger interdependences, and also some groups of variables with weaker interdependence relationships. However, few, if any previous work has explicitly considered such relationships among variables before predictive modeling. Fig. 24 (b) shows sub-network communities in the self-organized clustering of model-based parametric features. Notably, self-organizing algorithms derive a topological structure of the network of variables based on the matrix of nonlinear and asymmetric interdependences in Fig. 24 (a). As a result, homogeneous variables are clustered into sub-network communities.



(a) (b)

Figure 24 (a) Nonlinear interdependence matrix; (b) Self-organized clustering of model-based parametric features.

Furthermore, we minimized the redundancy of variables within each self-organized cluster through Gram-Schmidt orthonormalization and then integrated this new set of clustered variables with group elastic-net model to improve the performance of predictive modeling. The available data is divided into three parts: a training set (25% of samples), a validation set (25% of samples) and a test set (50% of samples).

The training set is to train the predictive model, and the validation set is to cross-validate the model and optimally determine the penalization parameters (also see equation 17). Model performance is only computed from the test dataset. The experiments were replicated for 100 times. Fig.8 shows the averages and standard deviation of prediction errors in the real-world case study. “Without clustering” represents the results from the lasso-penalized logistic regression model in our previous study [30], while “with clustering” denotes the results from the present study with self-organizing variable clustering that specifically considers nonlinear and asymmetric interdependence structures among predictor variables.

As shown in Fig. 25, self-organizing variable clustering yields smaller standard deviations of performance metrics (i.e., accuracy, sensitivity, and specificity) than “without clustering”. In addition, the accuracy of predictive model is improved from 89.50% to 94.04%, the sensitivity is increased from 94.33.8% to 97.22% and the specificity is improved from the original 84.80% to 90.84%. It is remarkable that experimental results demonstrated that the proposed methodology outperforms traditional models that do not explicitly consider complex interdependent structures among prediction variables. It strongly suggests that the nonlinear interdependence structures among the variables should be considered for better predictive performances. The results also imply that the self-organized variable cluster algorithm offers a reasonable grouping strategy for group variable selection.

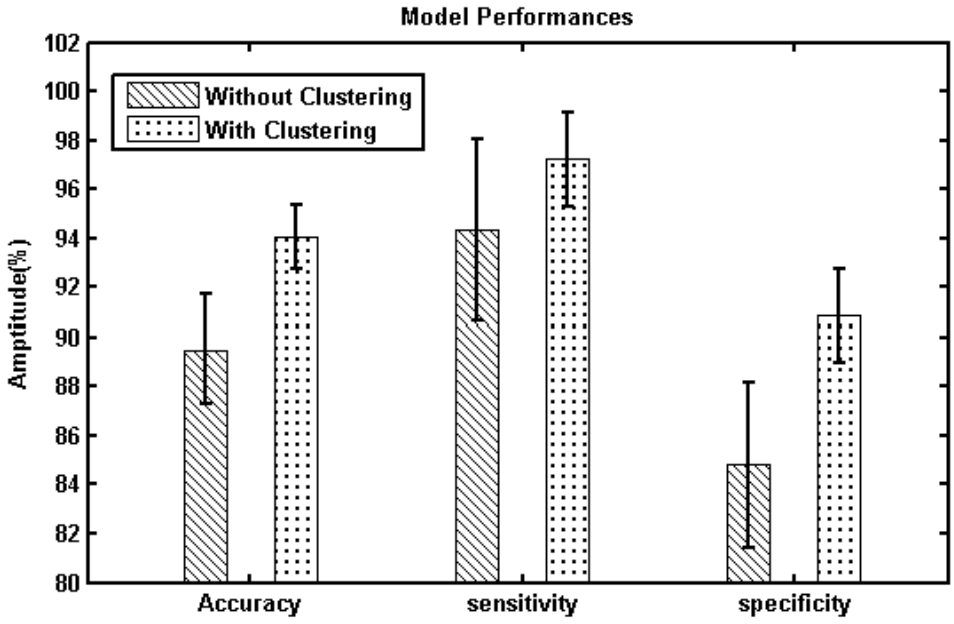


Figure 25 Averages and standard deviation of prediction errors in the real-world case study that extracts model parameters from VCG signals for identification of myocardial infarction patients.

4.6 Conclusions

Advanced sensing and real-time data acquisition bring Big Data which provides an unprecedented opportunity to move forward the new frontier of innovation. However, it is common that big data involves large amounts of variables with complex interdependence structures, which pose significant challenges on traditional methodologies in predictive analytics. To tackle these challenges, variable selection and variable clustering are widely used in the literature. Nonetheless, variable selection focuses primarily on the relevancy between predictors and the response variable and does not explicitly consider the redundancy (i.e., interdependence structures) among variables. The variable clustering, on the other hand, focuses only on the issues of variable redundancy while neglecting the relevancy between variables and the response. New methodologies that integrate variable clustering with variable selection to improve effectiveness and efficiency of predictive analytics are urgently needed.

This paper presents a new strategy that integrates the advantages of both variable clustering and variable selection. Complex interdependence structures among variables are characterized and quantified using nonlinear coupling analysis. Then, we developed a self-organizing network algorithm to effectively cluster variables that have nonlinear and asymmetric interdependences. This new method circumvents the limitations of existing methods of variable clustering. For examples, the HC algorithm is not applicable when the interdependence structure is asymmetric. The OPCC algorithm cannot adequately handle nonlinear interdependences. Further, the redundant information from related variables in self-organized clusters is minimized. Finally the self-organized clusters are integrated with group elastic net models to improve the performances of predictive models. As such, we handle the relevancy and redundancy among variables simultaneously. Experimental results in both simulation study and real-world case studies demonstrated that the proposed methodology not only outperforms traditional variable clustering algorithms such as HC and OPCC but also effectively identify cluster structures among variables, thereby improving the performance of predictive modeling. The proposed new idea of self-organizing algorithm is generally applicable for variable clustering and predictive modeling in many disciplines that involve a large number of highly-redundant variables.

4.7 References

[1] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu and C. H. Mason, "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," J. Market. Res., vol. 43, pp. 204-211, 05/01; 2015/04, 2006.

[2] Y. Chen and H. Yang, "Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units," in Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, Chicago, IL, 2014, pp. 4310-4314.

[3] H. Yang and E. Kundakcioglu, "Healthcare Intelligence: Turning Data into Knowledge," Intelligent Systems, IEEE, vol. 29, pp. 54-68, 2014.

[4] Y. Ding, E. A. Elsayed, S. Kumara, J. Lu, F. Niu and J. Shi, "Distributed Sensing for Quality and Productivity Improvements," Automation Science and Engineering, IEEE Transactions on, vol. 3, pp. 344-359, 2006.

[5] J. Friedman, "On Bias, Variance, 0/1-Loss, and the Curse-of-Dimensionality," Data Mining and Knowledge Discovery, vol. 1, pp. 55-77, 1997.

[6] T. Nas and B. H. Mevik, "Understanding the collinearity problem in regression and discriminant analysis," Journal of Chemometrics, vol. 15, pp. 413-426, 2001.

[7] H. Wang, X. Zhang, K. Ashok and Q. Huang, "Nonlinear Dynamics Modeling of Correlated Functional Process Variables for Condition Monitoring in Chemical–Mechanical Planarization," Semiconductor Manufacturing, IEEE Transactions on, vol. 22, pp. 188-195, 2009.

[8] P. M. Narendra and K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection," Computers, IEEE Transactions on, vol. C-26, pp. 917-922, 1977.

[9] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," American Society for Quality, vol. 12, pp. 55-67, 1970.

[10] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, pp. 267-288, 1996.

[11] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani, "Least Angle Regression," The Annals of Statistics, vol. 32, pp. 407-499, 2004.

[12] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Statist. Soc. B, vol. 67, pp. 301-320, 2005.

[13] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Physical Review A, vol. 33, pp. 1134-1140, 1986.

[14] T. Lee, D. Duling, S. Liu and D. Latour, "Two-stage Variable Clustering for Large Data Sets," Proceeding of SAS Global Forum 2008, pp. 1-14, 2008.

[15] Z. Li, "An parallel hierarchical clustering algorithm based on SIMD-EREW," in Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on, 2012, pp. 658-660.

[16] H. Yang and Y. Chen. Heterogeneous recurrence monitoring and control of nonlinear stochastic processes. Chaos: An Interdisciplinary Journal of Nonlinear Science 24(1), pp. 013138. 2014.

[17] H. F. Kaiser, "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, vol. 23, pp. 187-200, 1958.

[18] H. Kantz and T. Schreiber, "Coupling and synchronisation of nonlinear systems," in Nonlinear Time Series Analysis, 2nd ed.Anonymous Cambridge, UK: Cambridge University Press, 2003, pp. 292-299.

[19] J. Arnhold, P. Grassberger, K. Lehnertz and C. E. Elger, "A robust method for detecting interdependence: application to intracranially recorded EEG," Physica D, vol. 134, pp. 419-430, 1999.

[20] H. Yang and G. Liu, "Self-organized topology of recurrence-based complex networks," Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 23, pp. 043116, 2013.

[21] V. R. Joseph, T. Dasgupta, R. Tuo and C. F. Jeff Wu, "Sequential Exploration of Complex Surfaces Using Minimum Energy Designs," Technometrics, vol. 57, pp. 64-74, 2014.

[22] T. Kohonen, "The self-organizing map," Proceedings of the IEEE, vol. 78, pp. 1464-1480, 1990.

[23] Y. Chen and H. Yang, "Self-organized neural network for the quality control of 12-lead ECG signals," Physiological Measurement, vol. 33, pp. 1399-1418, 2012.

[24] G. Liu and H. Yang, "Multiscale Adaptive Basis Function Modeling of Spatiotemporal Vectorcardiogram Signals," Biomedical and Health Informatics, IEEE Journal of, vol. 17, pp. 484-492, 2013.

[25] H. Yang, S. T. Bukkapatnam and R. Komanduri, "Spatio-temporal representation of cardiac vectorcardiogram (VCG) signals," Biomedical Engineering Online, vol. 11, pp. 16, 2012.

[26] H. Yang, "Multiscale Recurrence Quantification Analysis of Spatial Cardiac Vectorcardiogram (VCG) Signals," Biomedical Egnineering, IEEE Transactions on, vol. 58, pp. 339-347, 2011.

[27] H. Yang, S. T. Bukkapatnam, T. Le and R. Komanduri, "Identification of myocardial infarction (MI) using spatio-temporal heart dynamics," Medical Engineering & Physics, vol. 34, pp. 485-497, 2011.

[28] H. Yang, C. Kan, G. Liu and Y. Chen, "Spatiotemporal differentiation of myocardial infarctions," IEEE Transactions on Automation Science and Engineering, vol. 10, pp. 938-947, 2013. .

[29] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioTollkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation, vol. 23, pp. e215-e220, 2000.

[30] G. Liu, C. Kan, Y. Chen and H. Yang, "Model-driven parametric monitoring of high-dimensional nonlinear functional profiles," in Automation Science and Engineering (CASE), 2014 IEEE International Conference on, 2014, pp. 722-727.

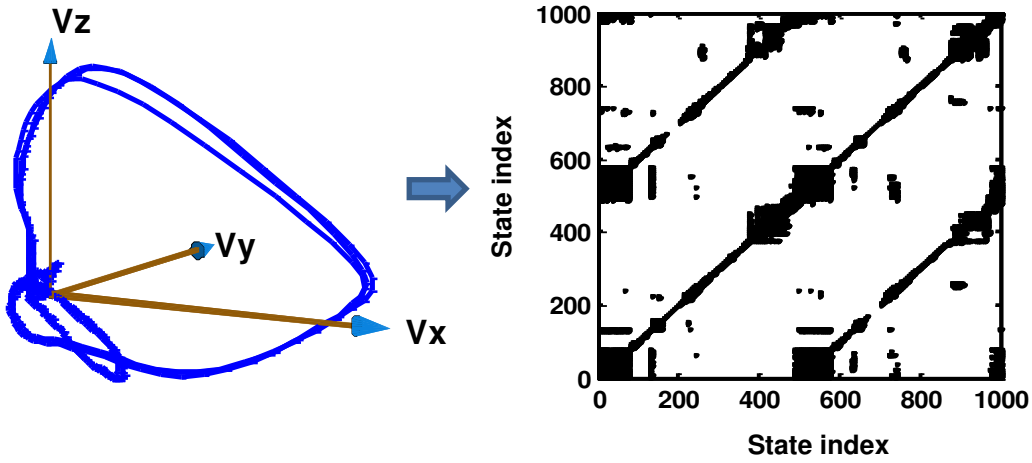
CHAPTER 5: SELF-ORGANIZED RECURRENCE NETWORKS

Network theory leads to a new way to investigate the dynamics of complex systems. As a result, many methods were proposed to construct a network from nonlinear time series. However, most previous works focused on deriving the adjacency matrix to represent the complex network and extract network-theoretic measures. Although the adjacency matrix provides connectivity information of nodes and edges, the network geometry can take variable forms. The definite network topology remains unknown. This paper develops a self-organizing approach to derive the steady geometric structure of a network from the adjacency matrix. Consequentially, novel network-theoretic measures will be achieved based on actual node-to-node distances in the self-organized network topology.

5.1 Introduction

Recurrence is a common behavior of complex systems. The recurrence plot was firstly introduced by Eckmann et al. [1] to characterize recurrence behaviors of dynamical systems. It provides a convenient means to capture the topological relationships existing in the state space in the form of 2D images. As shown in Fig. 26, if two states are located close to each other in the m-dimensional state space (e.g., 3D space in Fig. 26a), the color code is black in the recurrence plot. It is a graphical illustration of the state space and the recurrence behaviors. The recurrence plot characterizes the proximity of state vectors, i.e., whether or not the state-space distance between two "state" $\vec{x}(i)$ and $\vec{x}(j)$ is below a certain recurrence threshold r. It is mathematically defined as $R(i,j) \coloneqq \Theta(r - \|\vec{x}(i) - \vec{x}(j)\|)$, where Θ is the Heaviside function

and $||\cdot||$ is a distance measure. If they are located farther apart, the color is white. The structures of a recurrence plot have distinct topology and texture patterns. The ridges locate the nonstationarity and/or the switching between local behaviors. The parallel diagonal lines indicate the near-periodicity of the system behaviors [2]. Recently, recurrence methods have been successfully applied in different fields, e.g., cardiovascular system, biology, economy, manufacturing system, geophysics, and neuroscience.



(a) (b)

Figure 26 Graphical illustration of (a) the state space and (b) its recurrence plot

Recurrence-based networks were developed to explore the recurrence characteristics of dynamical systems from the perspective of network theory. Examples of previous approaches to construct a recurrence network include the partitioning of state space [3], cycle segmentation of pseudo-periodic time series [4, 5], correlation network of state vectors [6], visibility graph [7], and k-nearest neighbor network [8]. Further, it was argued that these existing approaches suffer from certain methodological limitations or a lack of general applicability [9]. Note that the recurrence plot (Fig. 1b) is a binary matrix with 0 and 1, which is analogous to the adjacency matrix of a complex network. A unifying framework to define recurrence network is based on the recurrences in the state space [9]. The recurrence matrix is considered as the adjacency matrix,

i.e., $A_{i,j} = R_{i,j} - I_{i,j}$, where $I_{i,j}$ is the Kronecker delta for avoiding self-loops. As a result, network-theoretic measures (e.g., average path length, clustering coefficient, and degree centrality) provide new means to characterize the complexity of a dynamical system. It may also be noted that network structure can be characterized at different scales or levels. The sub-graphs, also called network motifs [8], were shown to have different connectivity-patterns and occurring-frequencies for specific types of dynamics.

Although the recurrence-based adjacency matrix provides connectivity information of nodes and edges, the network geometry can take variable forms. In other words, the relevant spatial locations among nodes cannot be determined from the adjacency matrix. However, not only the connectivity but also spatial locations of nodes are critical in the functionality of a complex network. Few studies have been done to reconstruct a complex network with steady geometric structures from recurrence matrices. Note that most previous works focus on extracting network-theoretic measures from the adjacency matrix in recurrence-based complex networks. For example, the length of a shortest path $d_{i,j}$ is defined as the minimum number of edges that connect the node i and node j. The shortcoming of this measure is that it ignores the geometric distance between two nodes. If a steady geometry can be automatically organized for a recurrence network, the actual distances of network edges will be known. Hence, novel network-theoretic measures (e.g., average path length, diameter, efficiency and proximity ratio) will be achieved based on actual node-to-node distances in the network.

This paper presents a self-organized approach to derive the steady geometric structure of a recurrence-based complex network. The structure of the paper is organized as follows: Section II reviews the relevant literature on recurrence-based complex networks. Section III introduces the methodology of force-directed recurrence networks. Experimental design is presented in

section IV. Section V shows the experimental results and section VI derives the conclusions arising out of this present investigation.

5.2 Recurrence Networks

With the rapid technological advancement, network is almost everywhere in our daily life, e.g., social network, transportation network, computer network, sensor network. The adjacency matrix A is often used to represent the node-to-node connectivity in a network. In an undirected adjacency matrix, the element A_{ij} is 1 if node i is linked to node j and 0 otherwise. Statistical measures are widely used to exploit meaningful information in a network, e.g., node degree, link density, average path length, network diameter and clustering coefficient. A comprehensive review of network measures can be found in [10]. The node degree k_i, also named the degree centrality, is defined as the number of neighboring nodes of node i, i.e., $k_i = \sum_{j=1}^{n} A_{ij}$, where n is the number of nodes in the network. The link density ρ is the ratio of the number of edges to the number of possible edges, i.e., $\rho = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} A_{ij}$. The distance $d_{i,j}$ is the minimal number of edges to travel from node i to node j. The average path length L is $L = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{i,j}$. The network diameter D is the longest of all shortest paths, i.e., $D = \mathrm{Max}\{d_{i,j}\}$. The clustering coefficient of a node is the probability that two neighbors of a node i are also neighbors. It may be noted that most of the metrics are calculated from the adjacency matrix of a network. The success of network theory has fueled increasing interests in analyzing nonlinear time series from the perspective of complex network.

First, pseudoperiodic time series was extracted to construct a complex network for the investigation of nonlinear dynamics. Each cycle (i.e., the signal segment between two neighboring minima or maxima) is treated as a node in the undirected network. Two nodes are connected if their distance is less than a threshold. Further, the lag-reconstructed state space from

time series was used to derive the network. Each embedded state represents a node in the network. For the time series $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{N}\}^\mathrm{T}$, the state space, i.e., $\mathrm{x(i)} = (\mathrm{x_i}, \mathrm{x_{i+\tau}}, \dots, \mathrm{x_{i+\tau(M-1)}})$, $\mathrm{i} = 1, \dots, \mathrm{N} - \tau(\mathrm{M}-1)$ is a delay-embedded M-dimensional manifold. The embedding dimension M is determined by the false nearest neighbor algorithm, and the time delay parameter τ is estimated with the method of mutual information. Then, the Euclidean distance between nodes are calculated as

$$D_{i,j} = \|\boldsymbol{x}(i) - \boldsymbol{x}(j)\| = (\textstyle\sum_{m=0}^{M-1}(x_{i+m\tau} - x_{j+m\tau})^2)^{\frac{1}{2}} \quad (1)$$

A node $\mathrm{x(i)}$ is connected to its k nearest neighbors, but excluding the nodes in the same strand of the trajectory, i.e., $|\mathrm{j} - \mathrm{i}| > \Delta\mathrm{t}$. Therefore, the adjacency matrix of the k-nearest neighbor (KNN) network is defined as:

$$A_{ij} = \begin{cases} 1, & |j-i| > \Delta t \;\&\; j \in \{k \text{ nearest neighbors of } i\} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Based on the method of k-nearest neighbor network, the superfamily phenomena of the motifs have been further investigated to quantify the connectivity patterns of sub-networks [8]. Notably, a multiscale approach is developed to quantify the structure features of complex networks across different scales, i.e., from the local network to the global network.

Marwan et al. proposed to use the binary recurrence matrix to construct a complex network from the nonlinear time series [11]. Note that the recurrence of states is an important characteristic in the dynamical system. The recurrence matrix R characterizes the close proximity of two nodes, which is defined as $\mathrm{R_{i,j}} \coloneqq \Theta(\mathrm{r} - \|\mathrm{x(i)} - \mathrm{x(j)}\|)$, where Θ is the Heaviside function, r is the recurrence threshold and $||\cdot||$ is a distance measure. Marwan et al. constructed the undirected and unweighted network by treating the delay-embedded states as nodes and the recurrence as a link. The adjacency matrix A is obtained from the recurrence

matrix by removing the diagonal identities:

$$A_{i,j} = R_{i,j} - I_{i,j} \tag{3}$$

In addition, they proposed to use the network-theoretic measures as new quantifiers of nonlinear dynamical systems, e.g., degree centrality, clustering coefficient and average path length. Donner et al. [9] reviewed various approaches to transform nonlinear time series into complex networks, and showed that the concept of recurrence networks yields a unifying framework. Further, Donner et al. [12] investigated the geometrical properties of dynamical systems from the perspective of complex networks. Notably, they proposed to define two graph theoretical properties, i.e., local clustering coefficient and global (network) transitivity, which can be used to effectively define two new local and two new global measures of dimension in phase space.

It may also be noted that the adjacency matrix of KNN network is similar to the recurrence matrix. The KNN network uses the criterion of a fixed number of neighbors, instead of a fixed neighborhood size in the recurrence network proposed by Marwan et al. [9, 11]. The adjacency matrix of a KNN network is not necessarily symmetric and can be directed, because a node i is among the k-nearest neighbors of a node j does not mean that node j is also among the k-nearest neighbors of node i. However, the adjacency matrix of a recurrence network is symmetric and undirected. In addition, there are other methods to construct a complex network, such as transition networks, correlation networks, and visibility graphs. Note that most of previous network approaches focus on the adjacency matrix and network measures derived from it, as opposed to the self-organizing geometry of network. The network statistics (or graph theoretical properties), derived from the adjacency matrix, describe the characteristics of complex network, but not the topological geometry. Therefore, it is critical to develop an approach to derive the self-organized network topology from an adjacency matrix so that the

network-theoretic measures based on real distance can be obtained. The new network-theoretic measures will potentially discover nonlinear dynamics of the systems.

5.3 Force-directed Recurrence Networks

In this investigation, the spring-electrical model [13] is adopted for the topological self-organization of complex networks. It may be noted that this algorithm was widely used for graph representation in computer sciences [14]. However, few previous works explored the self-organized geometry of a recurrence network with the use of spring-electrical model. As aforementioned, most of previous works focused on extracting network-theoretic measures from the adjacency matrix of recurrence network. The adjacency matrix provides information of vertices and edges in a recurrence network, but its structure can take variable forms. Therefore, we introduce force-directed recurrence networks so as to address the research question, i.e., "what is the self-organizing geometry of a recurrence network?" Also, this work will provide a new way to reproduce the attractor or time series from the recurrence plot.

Let $G = \{V, E\}$ be the undirected and unweighted network, where V is the set of nodes and E is the set of edges. As shown in Fig. 27, the adjacency matrix describes the connectivity between nodes. If two nodes i and j form an edge, there will be a black dot in the plot. The spring-electrical model assigns two forces, i.e., attractive and repulsive forces between nodes. The objective of self-organizing process is to optimally identify the spatial locations of network nodes that minimize the energy within the network. The repulsive force is defined as $f_r(i,j) = -\frac{CK^{1+p}}{|x(i)-x(j)|^p}, i \neq j$, which exists between any two nodes i and j. It may be noted that $f_r(i,j)$ is inversely proportional to the distance between nodes. The force-model parameter p in the denominator is used to control the long-range repulsive force. The long-range repulsive force will be reduced by a bigger p and increased by a smaller one. The attractive force is defined as

$f_a(i,j) = \frac{|x(i)-x(j)|^2}{K}, i \leftrightarrow j$, which exists only between two connected nodes and is proportional to the square of the distance between them. Hence, the combined force on a node i will be

$$f(i, \boldsymbol{x}, K, C) = \sum_{i \neq j} -\frac{CK^{1+p}}{|\boldsymbol{x}(i)-\boldsymbol{x}(j)|^{p+1}}(\boldsymbol{x}(i) - \boldsymbol{x}(j)) + \sum_{i \leftrightarrow j} \frac{|\boldsymbol{x}(i)-\boldsymbol{x}(j)|}{K}(\boldsymbol{x}(i) - \boldsymbol{x}(j)) \quad (4)$$

where K is the natural spring length, C regulates the relative strength of repulsive and attractive forces, and $\vec{x}(i) - \vec{x}(j)$ is the force-directional vector, which is separated from $f_r(i,j)$ and $f_a(i,j)$ to define the direction of combined force $f(i, \vec{x}, K, C)$.

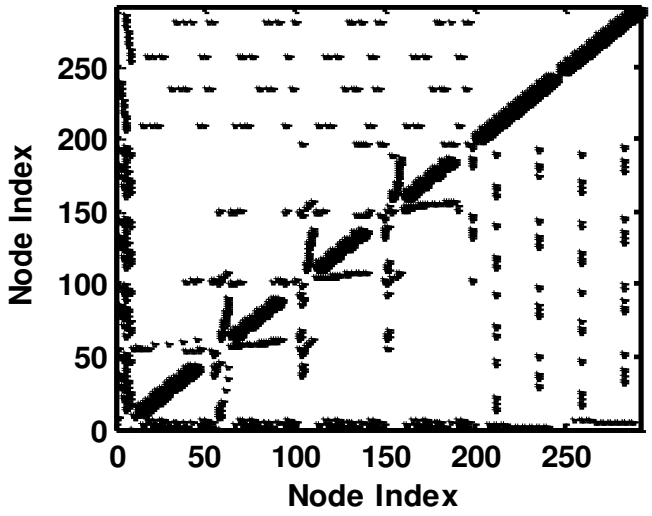


Figure 27 The example of an adjacency matrix in a complex recurrence network.

The objective of self-organizing process is to optimize the spatial locations of nodes that minimize the total energy of the network as: $\mathrm{Min}_x\{E(x, K, C)\} = \mathrm{Min}_x\{\sum_{i \in N} f^2(i, x, K, C)\}$. As a result, the recurrence network is not only steady with the minimal energy but also yields a unique geometry. It may be noted that the variations of parameters K and C will not change the network topology but yield a similar topology in different scales. The system energy is determined by the spatial locations of the nodes. The force-directed algorithm for the self-organizing process is as follows:

1) Network initialization: The nodes are randomly distributed in a multi-dimensional space. Then, the node-to-node connectivity is generated based on the adjacency matrix.

2) Force computation and position update: For the node i, the combined force, $f(i, \boldsymbol{x}, K, C)$, and its direction will be calculated. The position of node i will be moved along the force's direction for a magnitude τ, which is updated according to the following three scenarios: i) if the network energy keeps decreasing for 5 iterations, the magnitude will be increased to $\tau/0.9$. This will accelerate the movement of nodes toward the optimal topology; ii) if the network energy increases in one iteration, the magnitude will be decreased to $0.9 \times \tau$. This indicates that the layout is close to the optimal structure. A smaller magnitude will refine the search and prevent divergence; iii) otherwise, the previous magnitude is maintained.

3) Update the energy of network: Repeat step 2) for all nodes to update the system energy of network.

4) Stopping criteria: The self-organizing process will stop when the decrease of energy becomes small (i.e., $\Delta E < \varepsilon$).

In step 1), the nodes are randomly distributed in a high-dimensional space. It is remarkable that the self-organizing process will optimally choose the right dimension even if a larger dimension is selected. For example, if the network dimension is 2, the self-organizing process will organize the nodes on a plane even if they are randomly initialized in the 3 or higher dimensional space. For the visualization purpose, the dimension is often chosen to be 2 or 3. In step 2), the magnitude τ, for which nodes move along the force's direction, is the learning rate in the self-organization process. The magnitude of τ will drive the learning faster or slower. A bigger magnitude τ will accelerate the self-organizing process when the layout is far from the

optimal topology. A smaller one will slow down the process to refine the search when the layout is approaching the optimal topology.

Fig. 28 shows an example of the self-organizing process of a network with 292 nodes. Their connected edges are represented in the recurrence matrix (see Fig. 27). However, the recurrence matrix does not define the topology of networks. In the self-organizing process, first, 292 nodes are randomly distributed in the space with connected edges (Fig. 28a). The network topology keeps self-organizing iteratively, as shown from Fig. 28b to Fig. 28e. Finally, the optimal layout of network with a minimal energy is achieved (Fig. 28f). The optimal two-wing layout is achieved at 1900 iterations with a minimal energy and a steady topological structure.

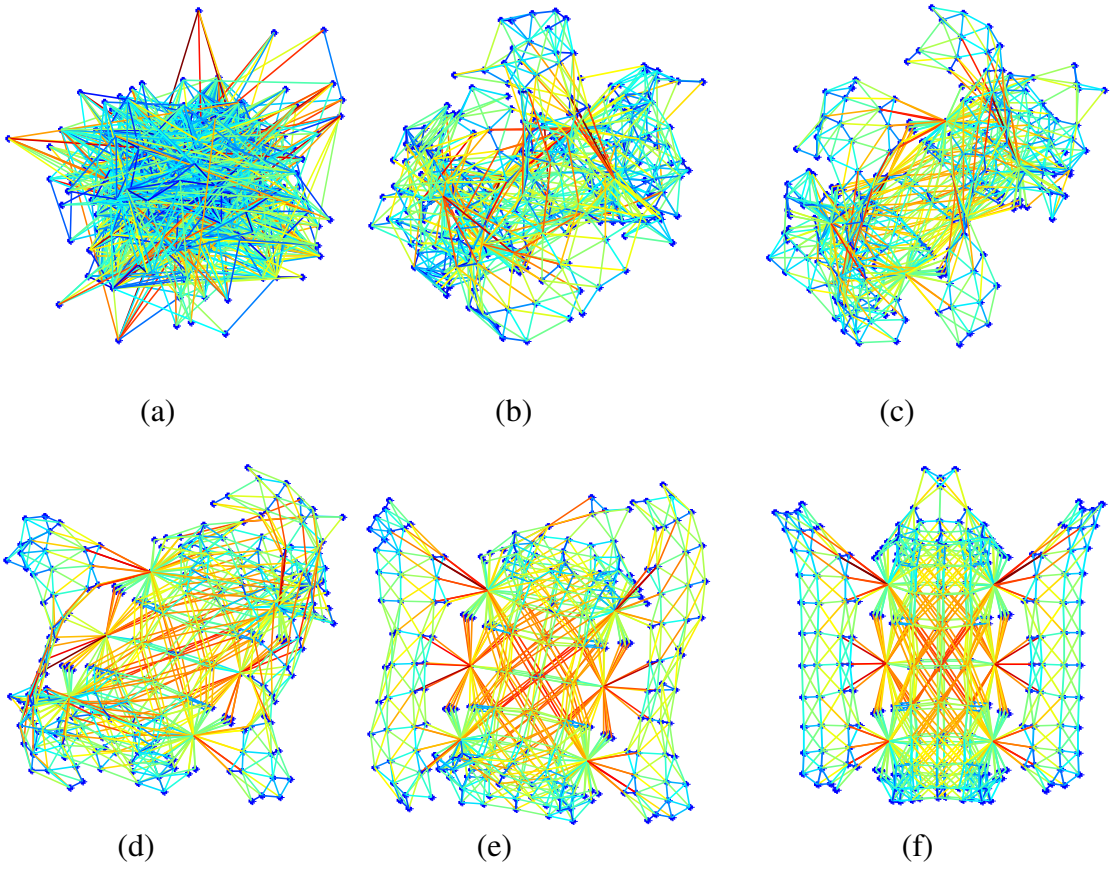


Figure 28 Illustrations of the self-organizing process of complex network based on the nodes and edges in the recurrence matrix.

The self-organizing process simulates forces between nodes and derives a steady geometry for a recurrence-based complex network. However, there are a number of variable factors affecting this self-organizing process, e.g., p in the model of repulsive force, various dynamical systems, and KNN network vs. recurrence network. The design of experiments to investigate these factors is described in the following section IV.

5.4 Experimental Design

A 4-way layout experiment was designed (see Fig. 29) to evaluate the performance variations of self-organizing algorithms due to four factor groups, i.e., dynamical systems, network construction methods, force-model parameter, and nonhomogeneous distribution.

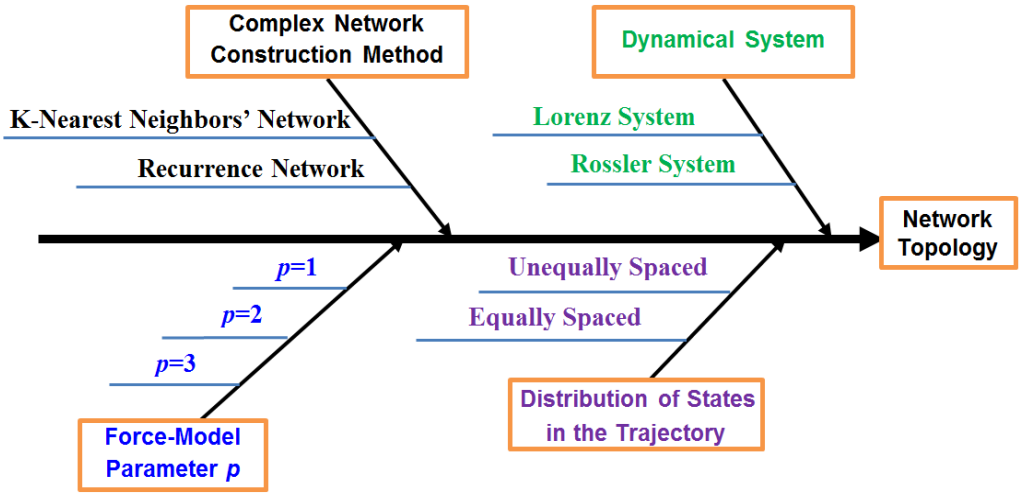


Figure 29 Cause-and-effect diagram for performance evaluation of self-organizing algorithms of recurrence networks.

1) Dynamical systems: We used the well-known dynamical systems (i.e., Lorenz and Rossler systems), and then derive the self-organizing network geometry from the adjacency matrices. Fig.23 (a) shows Lorenz attractor (i.e., unequally spaced) generated from the equations: $x' = 10(y - x)$, $y' = x(28 - z) - y, z' = xy - \frac{8}{3}z$ with the step $\Delta t = 0.01$ and Fig.23 (b) shows the corresponding equally-spaced Lorenz Attractor. Fig.23 (c) shows the original Rossler attractor (i.e., unequally spaced) generated from the

equations: $\mathrm{x}' = -\mathrm{y} - \mathrm{z}, \mathrm{y}' = \mathrm{x} + 0.2\mathrm{y}, \mathrm{z}' = 0.2 + \mathrm{z}(\mathrm{x} - 5.7)$ with the step $\Delta \mathrm{t} = 0.01$, and (d) equally-spaced Rossler attractor. Note that Lorenz attractor is approximately symmetric and balanced, but Rossler attractor is asymmetric and unbalanced. The symmetry will influence the clustering of repulsive and attractive forces.

2) Network construction methods: The adjacency matrix is constructed from the well-known dynamical systems with the use of two different approaches, i.e., *k*-nearest neighbor network and recurrence network. This comparison will show how network construction methods will impact the force-driven self-organizing process of network geometric structure.

3) Force-directed parameter: It may be noted that the force-model parameter p is involved in the model of repulsive force, $f_r(i,j) = -\frac{CK^{1+p}}{|\boldsymbol{x}(i)-\boldsymbol{x}(j)|^p}, i \neq j, i, j \in V, p > 0$. A bigger p will increase the denominator, and thus reduce the long-range repulsive force. We will experimentally vary the parameter p from 1 to 3 and investigate how it will impact the network energy and the final network geometry.

4) Nonhomogeneous distribution: Nonlinear Lorenz and Rossler systems are continuous. The states are unequally spaced along the trajectory due to the numerical integration of nonlinear differential equations with a time step (see Fig. 30a and Fig. 30c). This clustering of states in a local region influences the adjacency matrix, thereby aliasing the network geometry in the self-organizing process. To investigate how nonhomogeneous distribution affects the self-organizing process, we evenly spaced the states along the trajectory of attractors (see Fig. 30b and Fig. 30d) and made comparisons with the nonhomogeneous attractors.

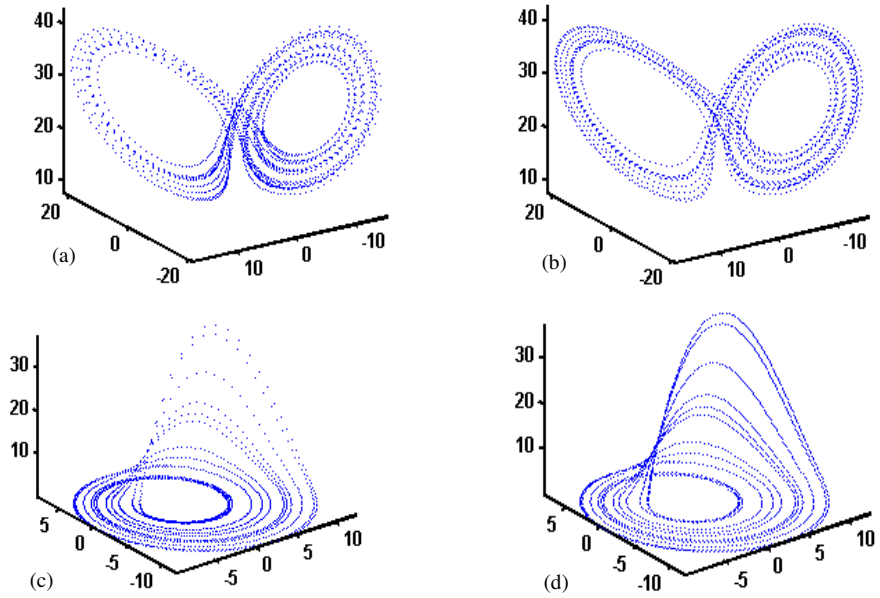


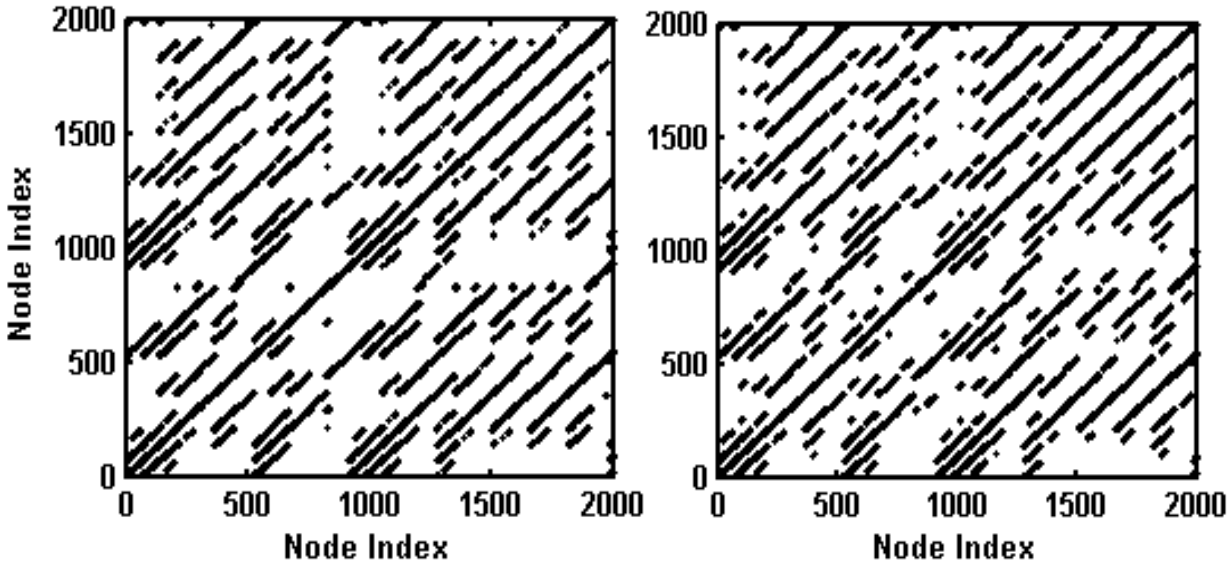
Figure 30 (a) Lorenz attractor, (b) equally-spaced Lorenz attractor, (c) Rossler attractor and (d) equally-spaced Rossler attractor.

5.5 Results

This present investigation made an attempt to self-organize the geometry of a recurrence network from the adjacency matrix. As discussed in the experimental design, we studied how the force-directed self-organizing process is impacted by four factor groups, i.e., dynamical systems, network construction methods, force-model parameter, and nonhomogeneous distribution. The results are as follows.

5.5.1 Effects of Network Construction Methods

As shown in Fig. 31, k-nearest neighbor and recurrence methods are used to derive the adjacency matrices from the original Lorenz attractor. In the KNN approach, k is set to be 51. In the recurrence approach, the recurrence threshold r is 5% of the maximal distance in the attractor. It may be noted that the plots of adjacency matrices (see Fig. 31 a and b) are similar for both methods, and the densities of black dots are approaximately the same, i.e., around 5 percent.



(a) (b)

Figure 31 The adjacency matrices of original Lorenz attractor that are derived with the use of (a) recurrence network and (b) *k*-nearest neighbors network.

Fig. 32 shows the iterative self-organizing process for the recurrence-based adjacency matrix of the Lorenz system. Force-directed algorithms automatically organize the geometric structure of network based on the nodes and edges in the adjacency matrix. Starting from a random layout of 2000 nodes in 3D space (see Fig. 32a), it takes 7000 iterations to reach the stable structure (see Fig. 32e). The topological structures at 1000, 3000, 5000, and 7000 iterations are shown in Fig. 32 (b-e), respectively. It is remarkable that this self-organizing process approximately reconstructed the Lorenz attractor.

It indicates that system dynamics are hidden in the recurrence-based adjacency matrix. Force-directed algorithms are capable of reconstructing the nonlinear system dynamics from the adjacency matrix. The edges between nodes are not drawn in the network, because a large number of edges will cover the trajectory. At the same time, it is seen that the reconstructed trajectory is a little bit different from the original trajectory. It is because the reconstructed trajectory is affected by the selection of recurrent rate and force model parameters.

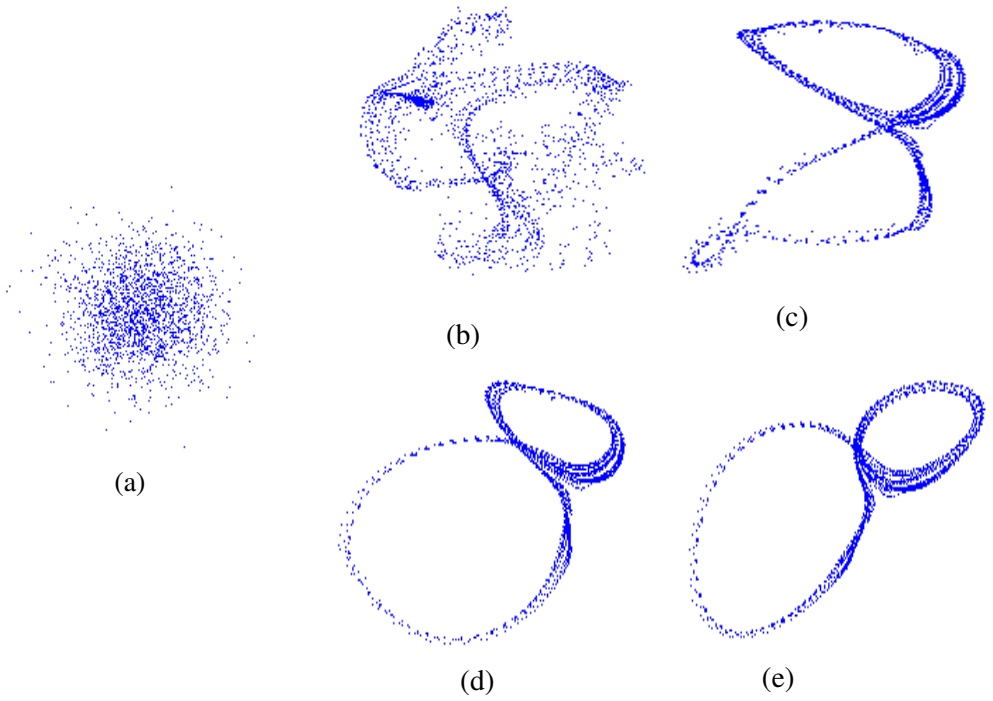


Figure 32 The self-organizing process of complex network based on the recurrence adjacency matrix of Lorenz system.

Fig. 33 shows the iterative self-organizing process for the KNN-based adjacency matrix of the Lorenz system. First, the 2000 nodes are randomly distributed in the 3D space (see Fig. 33a). Force-directed algorithms automatically organize the network structure at 1000, 3000, 5000 and 7000 iterations as shown in Fig. 33 (b-e). A stable network topology is reached after 7000 iterations (see Fig. 33e). Note that the final topology has relatively large differences from the Lorenz attractor. The experimental results show that recurrence networks show better capability to reconstruct the system dynamics in a nonlinear attractor. The reasons are as follows: 1) Recurrence networks reveal the intrinsic structure of nonlinear systems and don't connect two nodes unless they are really close to each other. However, the KNN network considers a fixed number of neighbors that may not be true recurrences. 2) A variable number of edges for each node facilitate the construction of real structures because this is true for the underlying patterns in the original attractor. The KNN network forces each node to have the same number of edges

and thus introduce more artificial effects. As a result, the KNN network leads to a self-organized network topology that has relatively large differences from the original attractor.

Similarly, we have also conducted experiments for the Rossler system. Both the experimental results show that recurrence networks are superior to the KNN network, and have better capability to reconstruct the system dynamics in a nonlinear attractor. It is because the KNN network imposes some artificial links among the nodes even if the distance between two nodes is not actually small. The links are added among the nodes just to ensure that the number of links for each node is the same. That will lead to artificial forces. The recurrence network place the forces just based on the natural distances and the forces are supposed to be existing. Therefore, we will adopt the method of recurrence networks in the following experiments of force-model parameter and nonhomogeneous distribution.

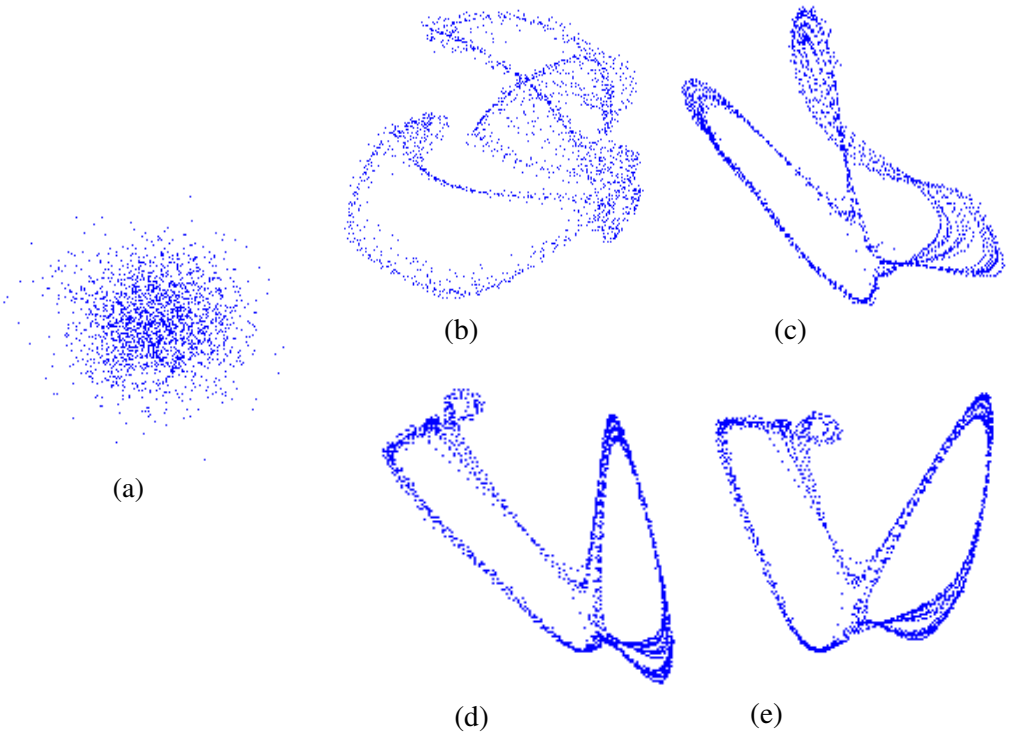


Figure 33 The self-organizing process of complex network based on the KNN adjacency matrix of Lorenz system.

5.5.2 Effects of Force-model Parameter

Besides the network construction methods, the parameter p plays an important role in the model of repulsive forces, thereby affecting the final network topology in the self-organizing process. In this section, the force-model parameter p will be varied from 1 to 2, and 3 for investigating the effects on the self-organized topology of recurrence networks.

Fig. 34 shows the iterative self-organizing process for the recurrence-based adjacency matrix (see Fig. 31a) of the Lorenz system with the force-model parameter p equal to 2. First, 2000 nodes are randomly distributed in the 3D space (see Fig. 34a). Force-directed algorithms automatically organize the network structure at 1000, 3000, 5000 and 7000 iterations as shown in Fig. 34 (b-e). A stable network topology is reached after 7000 iterations (see Fig. 34e). The final topology is an approximate reconstruction of the Lorenz attractor. It may be noted that the peripheral effect is highly reduced when p is set to be 2 in the model of repulsive force. This is shown by the fact that the obits are not tightly close to each other in the attractor. However, the orbits tend to cluster as a single trajectory when the force-model parameter p is 1 (see Fig. 33e). The reasons are as follows: (1) When the parameter p increases, the node-to-node repulsive force becomes smaller. (2) The Lorenz attractor has two circular regions with the nodes (or states) distributed in the boundary. (3) If the node-to-node repulsive force is smaller, the peripheral nodes will receive smaller repulsive forces from all the other nodes in the attractor. (4) If the peripheral nodes get bigger repulsive forces from one direction, they tend to cluster tightly. Otherwise, they will not be close to each other in the peripheral areas. By this analysis, it is suggested that an appropriate model parameter should be selected. To derive such a proper model parameter, we compare the reconstruction performance at different settings of force model parameter *p*.

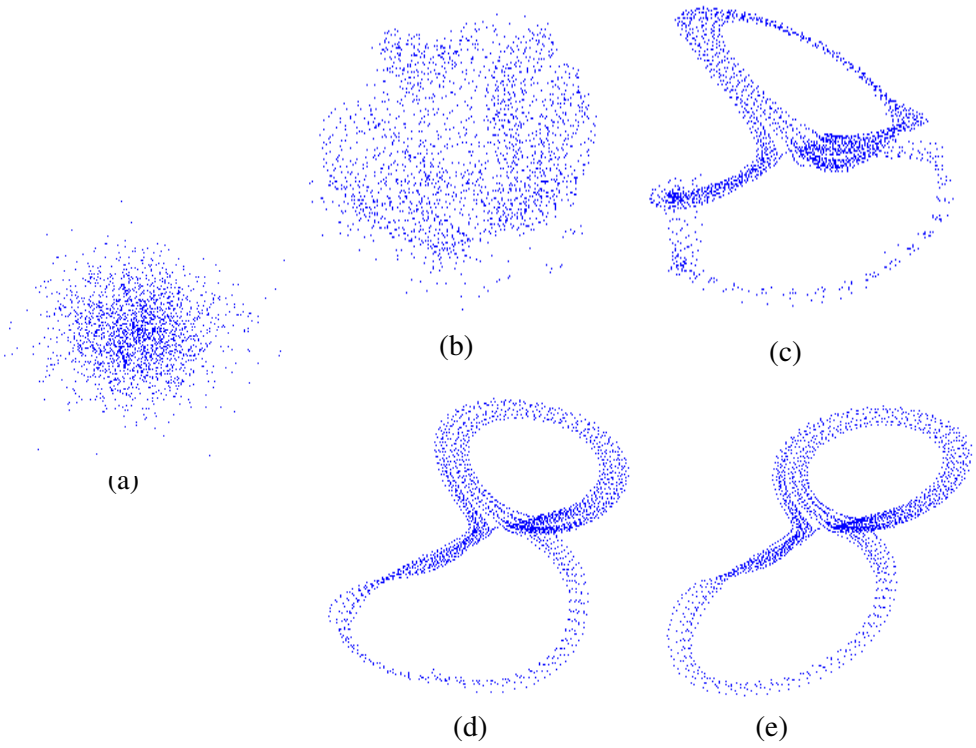


Figure 34 The self-organizing process of recurrence network of Lorenz system with the force-model parameter p equal to 2.

However, when the force-model parameter p is 3, the repulsive force is much smaller. As a result, the orbits have a larger distance between each other (see Fig. 35e). Fig. 35 shows the self-organizing process of recurrence network of Lorenz system with p equal to 3. Similarly, 2000 nodes are randomly distributed in 3D space (see Fig. 35a). Fig. 35 (b-e) show the self-organized topological structures at 1000, 3000, 5000 and 7000. The stable topology is reached at 7000 iterations, which approximately reproduces the Lorenz attractor (see Fig. 35e). It may be noted that the distances between obits are larger and larger when the parameter p increases. However, there are some defects in local regions because of the unbalance between repulsive and attractive forces. Therefore, the force-model parameter p is chosen to be 2 that reduces the peripheral effect and achieves a better performance in reconstructing the system dynamics and its trajectory.

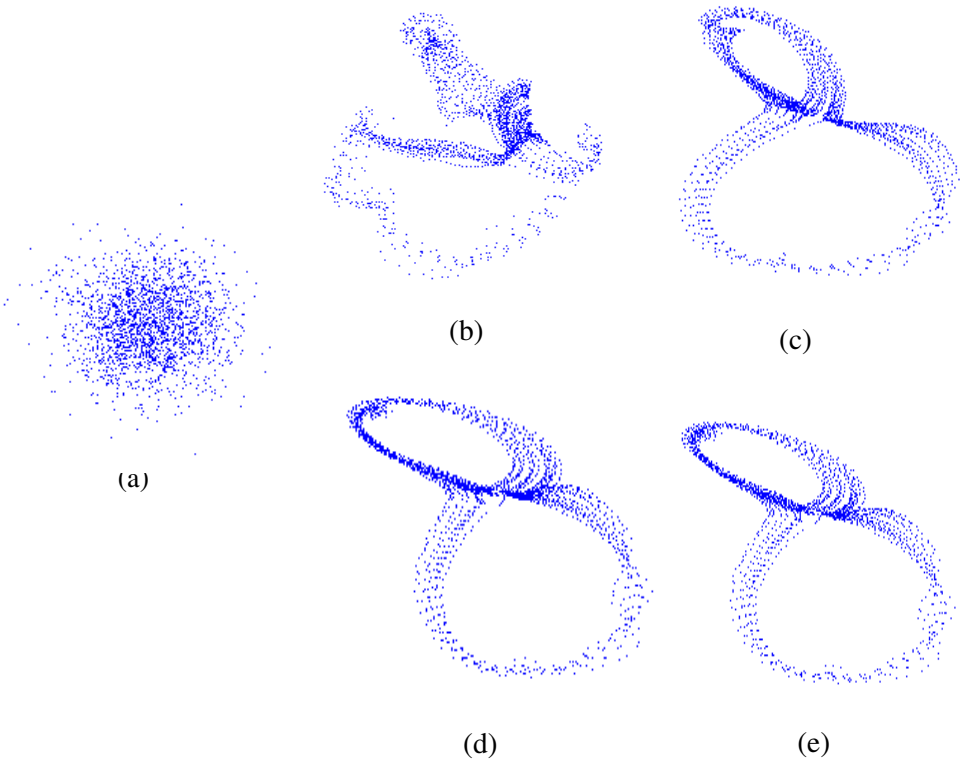


Figure 35 The self-organizing process of recurrence network of Lorenz system with the force-model parameter p equal to 3.

The same conclusion was drawn from the self-organizing process for the recurrence-based adjacency matrix (see Fig. 36a) of the Rossler system. Fig. 36(b-d) show the self-organized stable structure of recurrence-based complex network of Rossler system with the force-model parameter p equal to 1, 2 and 3 respectively. Note that the self-organized topology reconstructs the Rossler attractor when p equal to 2 is better than the case with p equal to 1 and 3. It may be noticed that the final topology has big differences from the Rossler attractor, especially in the Möbius band (i.e., the outer region) for all the three cases. This may be due to the asymmetric shape of Rossler attractor and sparse states discretized in the Möbius band, which leads to our further investigation of nonhomogeneous distribution in the next section. The distribution of the nodes along the trajectory will affect the distribution of assigned forces in the network, i.e., symmetric or asymmetric patterns of forces.

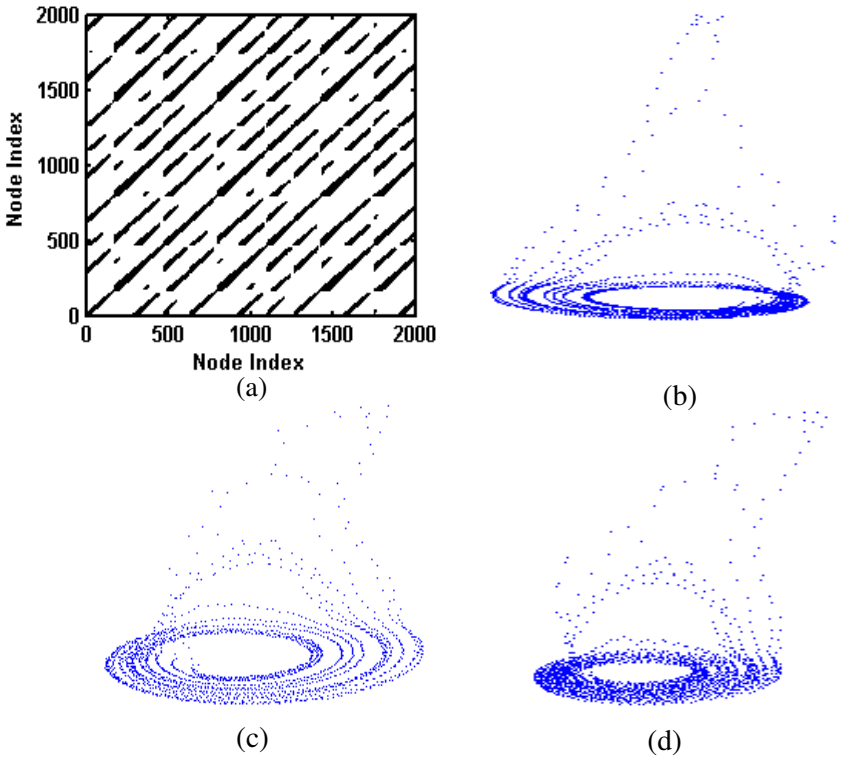


Figure 36 Recurrence adjacency matrix of Rossler system and its self-organized structures with the force-model parameter p equal to 1, 2 and 3.

5.5.3 Effects of Nonhomogeneous Distribution

The experimental results show the self-organized topology approximately reconstructs the attractors of Lorenz and Rossler systems (see Figs 32-35 and Fig 36(b-d)). However, there are aliasing effects in some local regions of the attractors, even if the system energy is getting stable. In particular, we found that there are sparse states distributed in these local regions because of the nonhomogeneous distribution. This factor of nonhomogeneous distribution poses significant difficulties to self-organizing process in the Rossler system. Therefore, we equally spaced the states along the trajectory of Lorenz and Rossler attractors. As shown in Fig. 37a and 37b, the recurrence matrices show different patterns from those in the unequally spaced attractors (i.e., Fig. 31a and Fig. 37a). The recurrent plot of the equally spaced attractors shows a larger recurrent rate.

Fig. 12c shows the self-organized stable structure for the recurrence-based adjacency matrix from the Lorenz attractor with states equally spaced along the trajectory (see Fig. 30b). It may be noted that the orbits are not tightly clustered and reassemble the shape of Lorenz attractor. The equally-spaced nodes along the trajectory achieve a better balance between repulsive and attractive forces in the network. The performance of self-organizing process is significantly improved. Therefore, it is expected that the self-organizing process will achieve a better performance if the interval between equally-spaced nodes is smaller and smaller (i.e., approaching a continuous dynamical system). This is more evident in the case of Rossler system, which has sparser states in the local regions. Fig. 37d shows the self-organized stable structure for the recurrence-based adjacency matrix from the Rossler attractor with states equally spaced along the trajectory (see Fig. 30d). It may be noted that the final topology reconstruct the Rossler attractor better than the unequally-spaced cases. Further, the Möbius band (i.e., the outer region), which was not well-reconstructed in the aforementioned experiments, clearly shows the trajectory of Rossler system. This demonstrates that the recurrence network, i.e., generated from the equally-spaced attractors, is capable of reconstructing the system dynamics through the self-organizing process.

5.6 Conclusions

In summary, simulating the complex network as a physical system, e.g., through the spring-electrical model, yields a self-organized topology. Although many previous methods have been proposed to transform nonlinear time series into complex networks, they have focused on the network representation using the adjacency matrix and the extraction of new network-theoretic measures. It is not clear about “what is the topology of recurrence network?” In other words, the connectivity information of nodes and edges in the adjacency matrix does not

define the geometry of the complex network. To derive the network-theoretic measurements based on the geometric distance, the topological structure of the network is in urgent need.

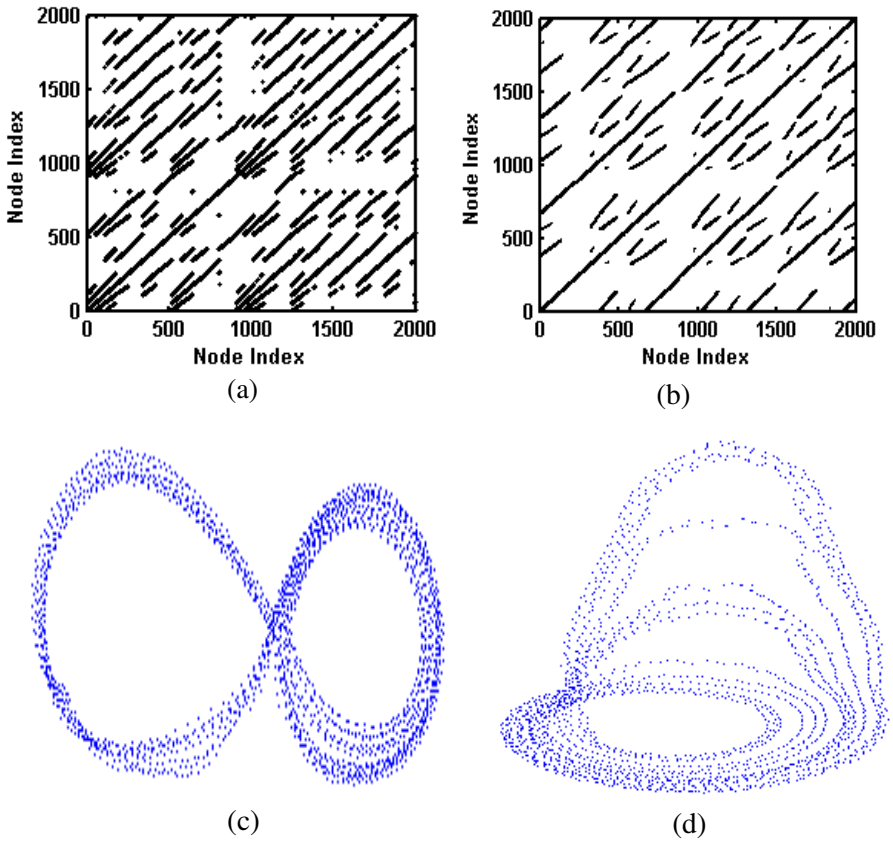


Figure 37 The recurrence adjacency matrices of and self-organized topological structures of equally-spaced (a) Lorenz attractor and (b) Rossler attractor.

This paper presents a self-organizing approach to derive the steady geometric structure of a network from the adjacency matrix. We introduce the spring-electrical model to investigate the hidden topological structure of recurrence-based complex networks. By simulating the edges as springs and the nodes as electrically charged particles, the network topology is self-organized. Interestingly, we found that the self-organizing geometry of a recurrence network recovers the attractor of a dynamical system that produced the recurrence adjacency matrix. This finding not only discloses the geometry of a recurrence network but also provides a new way to reproduce the attractor or time series from the recurrence plot.

Once the self-organized geometry of complex network is available, novel network-theoretic measures (e.g., average path length, diameter, and proximity ratio) can be achieved based on actual node-to-node distances. The paper lays out the solid foundations of most of previous studies on the graph theoretical statistics of recurrence networks. In addition, this paper brings the physical models into the recurrence analysis and discloses the spatial geometry of recurrence networks.

5.6 References

[1] J. P. Eckmann, S. O. Kamphorst and D. Ruelle, "Recurrence plots of dynamical systems," Europhysics Letters, vol. 4, pp. 973-977, 1987.

[2] N. Marwan, M. C. Romano, M. Thiel and J. Kurths, "Recurrence plots for the analysis of complex systems," Physics Reports, vol. 438, pp. 237-329, 2007.

[3] Y. Hirata, K. Judd and D. Kilminster, "Estimating a generating partition from observed time series: Symbolic shadowing," Phys Rev E., vol. 70, pp. 016215, 2004.

[4] J. Zhang and M. Small, "Complex network from pseudoperiodic time series: Topology versus dynamics," Physical Review Letters, vol. 96, pp. 238701, 2006

[5] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura and M. Small, "Characterizing pseudoperiodic time series through the complex network approach," Physica D, vol. 237, pp. 2856-2865, 2008.

[6] Y. Yang and H. Yang, "Complex network-based time series analysis," Physica A, vol. 387, pp. 1381-1386, 2008.

[7] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuno, "From time series to complex networks: The visibility graph," Proceedings of the National Academy of Sciences of the United States of America, vol. 105, pp. 4972-4975, 2008.

[8] X. Xu, J. Zhang and M. Small, "Superfamily phenomena and motifs of networks induced from time series," Proceeding of the National Academy of Sciences, vol. 105, pp. 19601-19605, 2008.

[9] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan and J. Kurths, "Recurrence networks--a novel paradigm for nonlinear time series analysis," New Journal of Physics, vol. 12, pp. 033025, 2010.

[10] L. Cui, S. Kumara and R. Albert, "Complex networks: An engineering view," Circuits and Systems Magazine, IEEE, vol. 10, pp. 10-25, 2010.

[11] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner and J. Kurths, "Complex network approach for recurrence analysis of time series," Physics Letters A, vol. 373, pp. 4246-4254, 2009.

[12] R. V. Donner, J. Heitzig, J. F. Donges, Y. Zou, N. Marwan and J. Kurths, "The Geometry of Chaotic Dynamics - A Complex Network Perspective," Eur. Phys. J. B, vol. 84, pp. 653-672, 2011.

[13] Y. Hu, "Efficient and high quality force-directed graph drawing," The Mathematica Journal, vol. 10, pp. 37-71, 2005.

[14] C. Walshaw, "Multilevel Refinement for Combinatorial Optimization Problems," Annals of Operation Research, vol. 131, pp. 325-372, 2004.

CHAPTER 6: CONCLUSIONS

In order to cope with system complexity and dynamic environments, modern industries are investing in a variety of sensor networks and data acquisition systems to increase information visibility. Multi-sensor systems bring the proliferation of high-dimensional functional Big Data that capture rich information on the evolving dynamics of natural and engineered processes. With spatially and temporally dense data readily available, there is an urgent need to develop advanced methodologies and associated tools that will enable and assist (i) the handling of the big data communicated by the contemporary complex systems, (ii) the extraction and identification of pertinent knowledge about the environmental and operational dynamics driving these systems, and (iii) the exploitation of the acquired knowledge for more enhanced design, analysis, monitoring, diagnostics and control.

My methodological and theoretical research as well as a considerable portion of my applied and collaborative work aims at addressing high-dimensional functional big data communicated by the systems. An innovative contribution of my work is the establishment of a series of systematic methodologies to investigate the complex system informatics including multi-dimensional modeling, feature extraction and selection, model-driven monitoring and root cause diagnostics.

Mathematical modeling of cardiac electrical signals facilitates the simulation of realistic cardiac electrical behaviors, the evaluation of algorithms, and the characterization of underlying space-time patterns. However, there are practical issues pertinent to model efficacy, robustness,

and generality. My first paper presents a multiscale adaptive basis function modeling approach to characterize not only temporal but also spatial behaviors of vectorcardiogram (VCG) signals. Model parameters are adaptively estimated by the "best matching" projections of VCG characteristic waves onto a dictionary of nonlinear basis functions. The comparison of QT intervals shows small relative errors (<5%) between model representations and real-world VCG signals when the model complexity is greater than 10. The proposed model shows great potentials to model space-time cardiac pathological behaviors, and can lead to potential benefits in feature extraction, data compression, algorithm evaluation and disease prognostics.

The mathematical modeling directly extracts the features (e.g. model parameters) from high-dimensional nonlinear functional profiles, which leads to my further investigation on model-based monitoring. Nowadays, the advanced sensor techniques and data acquisition systems provide an unprecedented opportunity for online monitoring of operational quality and integrity of complex systems. However, the classical methodology of statistical process control is not concerned about high-dimensional sensor signals and is limited in the capability to perform multi-sensor fault diagnostics. It is not uncommon that multi-dimensional sensing capabilities are not fully utilized for decision making. My second paper presents a new model-driven parametric monitoring strategy for the detection of dynamic fault patterns in high-dimensional functional profiles that are nonlinear and nonstationary. First, we developed a sparse basis function model of high-dimensional functional profiles, thereby reducing the large amount of data to a parsimonious set of model parameters while preserving the information. Further, we utilized the lasso-penalized logistic regression model to select a low-dimensional set of sensitive predictors for fault diagnostics. Experimental results on real-world data from patient monitoring showed that the proposed methodology outperforms traditional methods and effectively identify a sparse

set of sensitive features from high-dimensional functional data and profiles for process monitoring and fault diagnostics.

The first two projects built the framework for exploring informatics of high-dimensional nonlinear systems including modeling, characterization, pertinent knowledge discovery, diagnostics/prognostics and monitoring. However, currently the Big Data brings significant challenges for data-driven decision making and system optimization. Currently the rapid advancement of sensing and information technology brings the Big Data, which presents a gold mine of the 21st century. In particular, it is not uncommon that a large number of variables (or features) underlie the big data. Some of these variables are strongly correlated among themselves. Complex interdependence structures among variables challenge the traditional framework of predictive modeling. Our third investigation presented a new methodology of self-organizing network for variable clustering and predictive modeling. Experimental results on simulation studies and real-world data demonstrated that the proposed methodology not only outperforms traditional variable clustering algorithms such as hierarchal clustering and oblique principal component analysis, but also effectively identify interdependent structures among variables and further improves the performance of predictive modeling. The proposed new idea of self-organizing network is generally applicable for variable grouping and predictive modeling in many disciplines that involve a large number of variables (or factors, or features) in the big data.

Besides the sensor-based modeling and analysis of complex systems, I also expand my research interests to nonlinear dynamics and the resulting chaotic, recurrence, self-organizing behaviors. The self-organizing algorithm was introduced to learn the spatial topology of the recurrence networks. This investigation innovatively brought the physical models into the recurrence analysis and made me the winner of the student paper competition at 2014 IIE annual

conference. I also have the collaborative research work on the spatiotemporal differentiation of myocardial infarctions and sparse particle filtering for modeling space-time dynamics in distributed sensor networks with CSMMA lab members in University of South Florida.

Being involved in these research projects enabled me gain experience and develop my expertise in spatiotemporal informatics for complex systems monitoring, fault identification and root cause diagnostics especially in a big data environment. Therefore, I plan to continue my research in this realm. I will focus on both solutions to challenging questions rising from cutting edge problems and fundamental methodology development on the related topics in my further career. For example, because simulation has become an integral part of medical education and personnel evaluation and organ modeling is an ongoing research area for simulation purposes, image processing is of high value. As such, by applying image processing techniques, I will conduct research on segmentations and 3D modeling of soft tissues and organs in magnetic resonance imaging (MRI) to provide effective medical simulation of human body. In addition, I aim to apply such image processing techniques to other problems in manufacturing. For example defect on industrial products such as glass, steel welds and metal surfaces can be difficult to detect through visual inspection. Therefore I plan to develop automated inspection models for defect detection through image processing techniques. For another example, the prediction of in-hospital mortality for an intensive care unit (ICU) patient is critical in evaluating the patient's situation and assists the doctors to make the correct decisions. The difficulties lie in the big amount of recordings from up to 50 biomarkers, the variety forms of data stored (e.g., continuous, binary, interval and missing data) and the veracity of the data. I am developing a Gaussian graphical model to learn the root cause of in-hospital death based on the preprocessing of medical recordings. This project will potentially disclose the underlying mechanism that

influences the in-hospital death rate and suggest the most immediate treatment to the doctors. My longer-term aims involve further development of fundamental methodology and computational technology and their translational application to improve human.

APPENDICES

Appendix A Copyright Permissions

Below is permission for the use of material in Chapter 2.

Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

IEEE Requesting permission to reuse content from an IEEE publication

Title: Multiscale Adaptive Basis Function Modeling of Spatiotemporal Vectorcardiogram Signals
Author: Gang Liu; Hui Yang
Publication: Biomedical and Health Informatics, IEEE Journal of
Publisher: IEEE
Date: March 2013
Copyright © 2013, IEEE

LOGIN
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

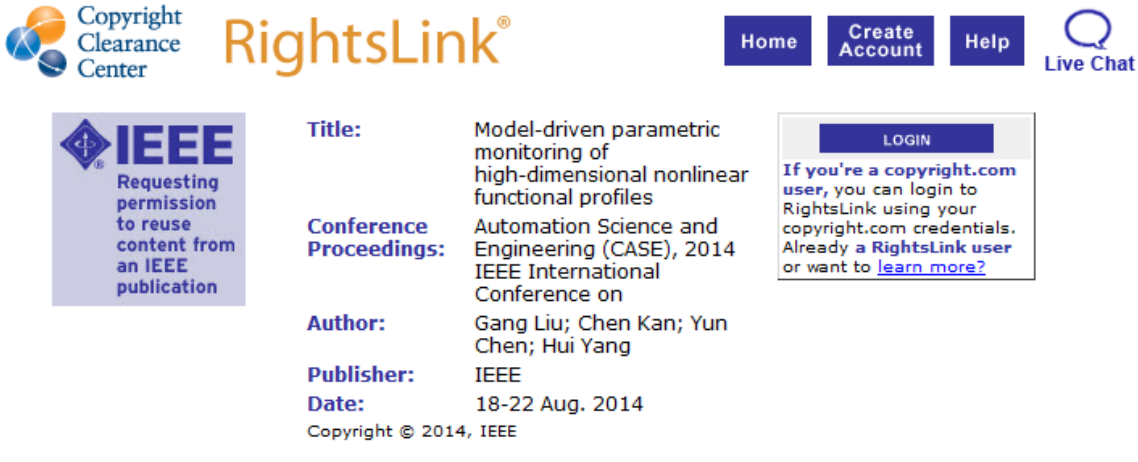
Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Below is permission for the use of material in Chapter 3.



Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

IEEE Requesting permission to reuse content from an IEEE publication

Title:	Model-driven parametric monitoring of high-dimensional nonlinear functional profiles
Conference Proceedings:	Automation Science and Engineering (CASE), 2014 IEEE International Conference on
Author:	Gang Liu; Chen Kan; Yun Chen; Hui Yang
Publisher:	IEEE
Date:	18-22 Aug. 2014

Copyright © 2014, IEEE

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW