MODELING AND IDENTIFICATION OF GENE REGULATORY NETWORKS: A GRANGER CAUSALITY APPROACH

Z. G. ZHANG[1], Y. S. HUNG[1], S. C. CHAN[1], W. C. XU[1], Y. HU[2]

[1]Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
[2]Department of Orthopaedics and Traumatology, The University of Hong Kong, Duchess of Kent Children's Hospital,
12 Sandy Bay Road, Hong Kong, China
E-MAIL: zgzhang@eee.hku.hk, yshung@eee.hku.hk, scchan@eee.hku.hk, wcxu@eee.hku.hk, yhud@hkusua.hku.hk

Abstract:

It is of increasing interest in systems biology to discover gene regulatory networks (GRNs) from time-series genomic data, i.e., to explore the interactions among a large number of genes and gene products over time. Currently, one common approach is based on Granger causality, which models the time-series genomic data as a vector autoregressive (VAR) process and estimates the GRNs from the VAR coefficient matrix. The main challenge for identification of VAR models is the high dimensionality of genes and limited number of time points, which results in statistically inefficient solution and high computational complexity. Therefore, fast and efficient variable selection techniques are highly desirable. In this paper, an introductory review of identification methods and variable selection techniques for VAR models in learning the GRNs will be presented. Furthermore, a dynamic VAR (DVAR) model, which accounts for dynamic GRNs changing with time during the experimental cycle, and its identification methods are introduced.

Keywords:

Gene regulatory network; Granger causality; Regularization; Time-series genomic data; Variable selection; Vector autoregressive model

1. Introduction

Inference of gene regulatory networks (GRNs) from genomic data is a significant and challenging problem in systems biology, because it not only leads to an insightful understanding of molecular mechanisms, but can also be useful for practical applications, such as cancer prediction and drug discovery [1]. A GRN is a collection of interactions and relationships among a set of genes and gene products (RNA or proteins) in a cell or an organism. Complex biological activities, such as the transcription regulation, genetic pathway, protein-protein interactions, etc., can be revealed by elucidating the structures of GRNs.

Because computational methods for GRN inference are cheaper and quicker than biological experiments, they are attracting the attention of more and more scientists and researchers from statistical and engineering fields [2]-[5]. In last decade, a number of statistical models and system identification methods have been proposed to study the GRNs, such as the Bayesian network, probabilistic Boolean network, graphical Gaussian models, structural equation models, etc. See review papers [2]-[5] and the references therein for details of above methods. These methods represent the GRN on different levels of abstraction and depend on different prior knowledge on reaction mechanisms, and thus they have different performances and superiorities in practical applications. However, all these methods may fail to reveal the causality (i.e., the direction of information flow), which is of special importance in GRNs, among the genes and proteins.

More recently, Granger causality, as a powerful approach to identify the causality between time-series, gains increasing attention in systems biology due to its simplicity and effectiveness. Its successful applications to inferring GRNs of yeast cells, cancer cells, and human blood cells were reported in [6]-[15]. Generally, the Granger causality is tested by means of vector autoregressive (VAR) models. In the case of GRN, time-series genomic data are modelled as a VAR process, with the dimension being the number of genes and proteins in the genomic data. The extent of the Granger causality between two genes is evaluated by the corresponding element in the VAR coefficient matrix. The estimation of the VAR coefficients is conventionally achieved by minimizing the maximum likelihood function, which leads to the ordinary least-squares solution. However, because the number of variables is always much larger than the number of samples, the solution to the VAR model is often statistically inefficient and computationally demanding [16]. Therefore, fast and efficient variable selection techniques are highly desirable for identification of the VAR model.

Prior biological knowledge is helpful for selecting meaningful genes before building the VAR model, but it is not always available for most genomic data. Hence, variable selection is often carried out automatically by regularization techniques during identification of the VAR model [17]-[19].

978-1-4244-6527-9/10/$26.00 ©2010 IEEE

Regularization techniques are extensively studied and widely applied in a variety of fields, because of its effectiveness in reducing the estimation variance and automatic variable selection [20]-[23]. The effect of variable selection can be obtained because some regularization techniques, such as the L_1 regularization [21], the elastic net [22], and the smoothly clipped absolute deviation (SCAD) [23], hold the property of sparsity where small and irrelevant coefficients can be shrunk towards zero. This sparsity property is extremely desirable for inferring GRN, where only a few interactions among a sea of genes exist [24]. In this paper, an introductory review of popular variable selection techniques in identifying VAR models and inferring GRNs will be presented.

Lastly, this paper will give a brief review on inferring of dynamic GRN from time-series genomic data. With the rapid developments of high-throughput technologies and increased availability of time-series microarray data, growing attention has been paid to how the GRNs of a cell or an organism changes with time over a period, say, during different stages of cell cycle or under different experimental conditions. As an extension of VAR models to non-stationary conditions, the dynamic VAR (DVAR) models and its existing identification methods are introduced, while some potential extensions are discussed.

2. Granger causality and VAR models

2.1. Granger causality

Granger causality was originally proposed by the Nobel laureate Clive Granger in 1969 for econometric data [25]. Given two time-series X and Y, if the past values of X can help result in a better prediction of future value of Y than the prediction solely based on past values of Y, then we can say that X "Granger-causes" Y. More precisely, denote the prediction of $Y(n+1)$ based on $X(\tau \mid \tau \le n)$ and $Y(\tau \mid \tau \le n)$ as $\hat{Y}^{(X,Y)}(n+1)$, and denote the prediction of $Y(n+1)$ using only $Y(\tau \mid \tau \le n)$ as $\hat{Y}^{(Y)}(n+1)$. If the prediction error of $\hat{Y}^{(X,Y)}(n+1)$ is smaller than the error of $\hat{Y}^{(Y)}(n+1)$, it is implied that $X(\tau \mid \tau \le n)$ contain useful information for the evolution of Y and thus X "Granger-causes" Y.

Granger causality can be easily extended to K-dimensional ($K > 2$) data $X_1, X_2, \cdots, X_K$. If past values of X_j can yield a more accurate prediction of X_l when past values of other X_k ($k = 1,2,\cdots,K$ and $k \ne j$) are also included in the prediction, then X_j "Granger-causes" X_l. It is worth noting that the Granger causality can be extended to evaluate the spectral causality of two or more time-series in the frequency domain [11].

2.2. VAR models

Granger causality is generally tested by a VAR model, which is actually multiple linear regression accounting for linear relationship. We now consider a set of time-series microarray data containing K genes and measured at N time instants as $\boldsymbol{x}(n) = [x_1(n), x_2(n), \cdots, x_K(n)]^T$ ($n = 1,2,\cdots,N$). A p-order VAR model describes the microarray data as

$$\boldsymbol{x}(n) = \sum_{i=1}^{p} \boldsymbol{A}^{(i)} \boldsymbol{x}(n-i) + \boldsymbol{e}(n), \tag{1}$$

where $\boldsymbol{A}^{(i)} = \begin{bmatrix} a_{1,1}^{(i)} & a_{1,2}^{(i)} & \cdots & a_{1,K}^{(i)} \\ a_{2,1}^{(i)} & a_{2,2}^{(i)} & \cdots & a_{2,K}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K,1}^{(i)} & a_{K,2}^{(i)} & \cdots & a_{K,K}^{(i)} \end{bmatrix}$ ($i = 1,2,\cdots,p$) is a $K \times K$ VAR coefficient matrix, and $\boldsymbol{e}(n) = [e_1(n), e_2(n), \cdots, e_K(n)]^T$ is a K-dimensional error vector, which is often assumed to be a zero-mean multivariate Gaussian process with covariance matrix $\Sigma = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots & \sigma_{1,K}^2 \\ \sigma_{2,1}^2 & \sigma_{2,2}^2 & \cdots & \sigma_{2,K}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K,1}^2 & \sigma_{K,2}^2 & \cdots & \sigma_{K,K}^2 \end{bmatrix}$. The order of the VAR model, p, is commonly selected as 1 in literature because a larger order implies much more VAR coefficients to be estimated and may result in a larger variability for the estimated coefficients [8]. Higher-order ($p > 1$) VAR models were also tested in [13], and the results showed improved performance in some cases.

Granger-causality can be revealed by checking the components of the VAR coefficient matrix $\boldsymbol{A}^{(i)}$. That is to say, an entry $a_{j,l}^{(i)}$, $j = 1,2,\cdots,K$ and $l = 1,2,\cdots,K$, in $\boldsymbol{A}^{(i)}$ denotes whether gene x_l "Granger-causes" gene x_j. If the value of $a_{j,l}^{(i)}$ is significantly deviated from zero, Granger-causality between x_j and x_l can be identified. On the other hand, a zero value of $a_{j,l}^{(i)}$ implies the absence of Granger-causality between x_j and x_l.

2.3. Identification of VAR models

To identify the VAR models, we rewrite (1) in a standard linear regression framework as

$$\boldsymbol{Y}(n) = \boldsymbol{X}(n)\boldsymbol{B} + \boldsymbol{e}(n), \tag{2}$$

where $\boldsymbol{Y}(n) = \boldsymbol{x}^T(n) \in \boldsymbol{R}^{1\times K}$,

$\boldsymbol{X}(n) = [\boldsymbol{x}^T(n-1), \boldsymbol{x}^T(n-2), \cdots, \boldsymbol{x}^T(n-p)] \in \boldsymbol{R}^{1\times pK}$, and

$\boldsymbol{B} = [\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \cdots, \boldsymbol{A}^{(p)}]^T \in \boldsymbol{R}^{pK\times K}$. The VAR coefficient $\boldsymbol{B}$ can be estimated by the maximum likelihood estimation (MLE). Since the additive noise $\boldsymbol{e}(t)$ is assumed to be zero mean and white Gaussian distributed, maximizing the likelihood is equivalent to minimizing the mean squared error (MSE) criterion as follows:

$$\hat{\boldsymbol{B}} = \arg\min_{\boldsymbol{B}} \| \overline{\boldsymbol{Y}} - \overline{\boldsymbol{X}}\boldsymbol{B} \|^2 , \tag{3}$$

where $\overline{\boldsymbol{Y}} = [\boldsymbol{Y}^T(p+1), \boldsymbol{Y}^T(p+2), \cdots, \boldsymbol{Y}^T(N)]^T \in \boldsymbol{R}^{(N-p)\times K}$, and $\overline{\boldsymbol{X}} = [\boldsymbol{X}^T(p+1), \boldsymbol{X}^T(p+2), \cdots, \boldsymbol{X}^T(N)]^T \in \boldsymbol{R}^{(N-p)\times pK}$. Eq. (3) leads to the ordinary least-squares (OLS) solution as

$$\hat{\boldsymbol{B}} = (\overline{\boldsymbol{X}}^T \overline{\boldsymbol{X}})^{-1} \overline{\boldsymbol{X}}\overline{\boldsymbol{Y}} . \tag{4}$$

Note that in (4), the condition $(N-p) \geq pK$ must be satisfied, so that $\overline{\boldsymbol{X}}^T \overline{\boldsymbol{X}}$ is invertible and a unique OLS solution exists. However, time-series gene data typically contain only tens of time points, while they normally contain thousands of genes or more. That is to say, $K >> N$. Therefore, variable (gene) selection is an indispensable part in identification of GRN. In general, there are two kinds of variable selection approaches for inferring GRNs: one is to select a small subset of responsible genes based on prior biological knowledge or from clustered genes and to establish the VAR model using the selected genes [6], [8], [11], [12], and [15]; the other is to build the VAR model using all measured genes and then to automatically select genes during the identification of VAR models [7], [9], [10], [13] and [14]. The latter automatic variable selection approach is of special interest because prior biological knowledge may be absent or vague in most genomic data sets. The automatic variable selection can be achieved by regularization, which will be introduced in the next section.

3. Regularization and variable selection

3.1. Introduction of regularization techniques

In statistics, it is general to impose a regularization term in the OLS estimator to tackle the ill-posed problem. From a Bayesian respective, the regularization is closely related to incorporating prior information on the random variable $\boldsymbol{B}$. With the regularization, the estimator of the regularized cost function is given as

$$\hat{\boldsymbol{\beta}}_k = \arg\min_{\boldsymbol{\beta}} \{ \| \overline{\boldsymbol{y}}_k - \overline{\boldsymbol{X}}\boldsymbol{\beta}_k \|^2 + \sum_{i=1}^{pK} \rho_{\boldsymbol{\theta}}(\beta_{k,i}) \} , \tag{5}$$

where $\boldsymbol{\beta}_k$ ($k = 1, 2, \cdots, K$) is the k-th column of $\boldsymbol{B}$, $\beta_{k,i}$ is the i-th entry of $\boldsymbol{\beta}_k$, $\overline{\boldsymbol{y}}_k$ is the k-th column of $\overline{\boldsymbol{Y}}$, $\rho_{\boldsymbol{\theta}}(\cdot)$ is the regularization or penalty function with one or more regularization parameters $\boldsymbol{\theta}$.

Some commonly-used regularization functions include the L_2 regularization $\rho_\lambda(\beta) = \lambda\beta^2$, which leads to a ridge regression [20], and the L_1 regularization $\rho_\lambda(\beta) = \lambda |\beta|$, which leads to a least absolute shrinkage and selection operator (lasso) [21]. An elastic net regularization, which is a combination of L_1 and L_2 regularization, is also proposed for gene selection [22], and it has the form as, $\rho_{\lambda_1,\lambda_2}(\beta) = \lambda_1 |\beta| + \lambda_2\beta^2$ with two regularization parameters λ_1 and λ_2. The smoothly clipped absolute deviation (SCAD) regularization [23] is another popular technique for gene selection, and it is given as:

$$\rho_{\lambda,a}(\beta) = \begin{cases} \lambda|\beta| & \text{for } |\beta| \leq \lambda, \\ -\dfrac{(|\beta| - a\lambda)^2}{2(a-1)} + \dfrac{(a+1)\lambda^2}{2} & \text{for } \lambda < |\beta| \leq a\lambda, \\ \dfrac{(a+1)\lambda^2}{2} & \text{for } |\beta| > a\lambda, \end{cases} \tag{6}$$

with two regularization parameters $\lambda > 0$ and $a > 2$.

3.2. Desirable properties of regularization techniques

It is claimed in [23] that an appropriate regularization technique should hold three desirable properties: continuity, sparsity, and unbiasedness. All above regularization techniques satisfy the condition of continuity, which means $\rho_{\boldsymbol{\theta}}(\beta)$ are continuous in β to avoid instability. The lasso, elastic net, and SCAD estimators have the property of sparsity, which means that some small coefficients can be automatically set to zero, while the ridge estimator does not have this property. Therefore, the automatic variable selection for inferring GRNs can be achieved by the estimators with the sparsity property. Unbiasedness implies that the modeling bias introduced by the regularization term should be zero when the true coefficients are large enough, which is important for estimating significant VAR coefficients accurately. Among the four regularization techniques, only SCAD has the property of unbiasedness.

In addition, a “grouping effect”, which means that the coefficients of strongly correlated variables tend to be identical, is also considered as a useful property when dealing with gene data [22]. The elastic net is the only regularization technique having the grouping effect. Thus, elastic net can select a group of highly correlated genes once one among them is selected, while lasso and SCAD may only select one of them and discard others.

Table 1 lists the properties of the four commonly-used regularization techniques. The latter three are often used for automatic variable selection in inferring GRNs (lasso in [7], [9], [13], and [14]; elastic net in [10] and [18]; SCAD in [19]). Because no regularization method possesses all the desirable properties and their performance is actually data dependent [17], the most appropriate one should be determined based on known biological knowledge and practical requirements.

TABLE 1. REGULARIZATION TECHNIQUES AND THEIR PROPERTIES FOR ANALYSIS OF GENOMIC DATA

Property / Method	Continuity	Sparsity	Unbiased-ness	Grouping effect
Ridge	√	×	×	×
Lasso	√	√	×	×
Elastic net	√	√	×	√
SCAD	√	√	√	×

3.3. Algorithms and tuning of parameters

If an L_2 regularization is employed, the ridge estimator to the objective function in (5) is given by

$$\hat{\boldsymbol{\beta}}_k = (\bar{\boldsymbol{X}}^T \bar{\boldsymbol{X}} + \lambda \boldsymbol{I})^{-1} \bar{\boldsymbol{X}} \bar{\boldsymbol{y}}_k . \qquad (7)$$

For lasso, elastic net and SCAD, the objective functions are non-differentiable and it is difficult to obtain their solutions in analytic form. An iteratively re-weighted least-squares (IRLS) algorithm for solving this non-differentiable function was proposed in [23]. The IRLS algorithm is based on a Taylor approximation of the regularization term $\rho_{\theta}(\beta)$ as

$$\rho_{\theta}(\beta) \approx \rho_{\theta}(\beta^{(0)}) + \tfrac{1}{2}[\rho_{\theta}{}'(\beta^{(0)}) / |\beta^{(0)}|](\beta^2 - (\beta^{(0)})^2), \qquad (8)$$

where $\beta^{(0)}$ is an expansion point close to β. Then, an approximate closed-form solution to lasso, elastic net, and SCAD estimators can be obtained by rewriting (5) as:

$$\hat{\boldsymbol{\beta}}_k = \arg\min_{\boldsymbol{\beta}} \{ \| \bar{\boldsymbol{y}}_k - \bar{\boldsymbol{X}} \boldsymbol{\beta}_k \|^2 + \sum_{i=1}^{pK} \psi_{\theta}(\beta_{k,i}^{(0)}) \beta_{k,i}^2 \} , \qquad (9)$$

where $\psi_{\theta}(\beta^{(0)}) = \frac{1}{2}\rho_{\theta}{}'(\beta^{(0)}) / |\beta^{(0)}|$. It can be seen that the regularization term in (9) has a similar form to an L_2 norm with a weight $\psi_{\theta}(\beta_{k,i}^{(0)})$. Thus, it was suggested to compute the ridge regression iteratively to yield the IRLS algorithm for the lasso/elastic net/SCAD estimators as:

$$\hat{\boldsymbol{\beta}}_k^{(m+1)} = (\bar{\boldsymbol{X}}^T \bar{\boldsymbol{X}} + \boldsymbol{\Psi}^{(m)})^{-1} \bar{\boldsymbol{X}} \bar{\boldsymbol{y}}_k , \qquad (10)$$

where $m = 0,1,...$ is the number of iteration, and $\boldsymbol{\Psi}^{(i)} = diag\{\psi_{\lambda}(\beta_{k,1}^{(i)}), \cdots, \psi_{\lambda}(\beta_{k,pK}^{(i)})\}$. A good initial value $\hat{\boldsymbol{\beta}}^{(0)}$ for (10) is the ridge estimate. The iteration will stop until a maximum number is reached or the difference between two successive steps is small enough.

In practical implementation, the regularization parameters have considerable impact on the results and need to be tuned as well. Conventionally, it can be chosen as the best one among a series of candidate values by the cross-validation (CV) criterion. For example, in a Q-fold CV for a data set containing D observations, denote the testing set and training set as $D^{(q)}$ ($q = 1,2,\cdots,Q$) and $D - D^q$, respectively. For each testing set $D^{(q)}$, an estimate of $\hat{\boldsymbol{\beta}}_{(\boldsymbol{\theta})}^{(q)}$ is obtained by candidate parameters $\boldsymbol{\theta}$. The optimal $\boldsymbol{\theta}$ will be the one minimizing the CV criterion as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{q=1}^{Q} \sum_{(\bar{\mathbf{y}},\bar{\mathbf{X}})\in D^q} \| \bar{\mathbf{y}} - \bar{\mathbf{X}} \hat{\boldsymbol{\beta}}_{(\boldsymbol{\theta})}^q \|^2 . \qquad (13)$$

The CV procedure is often very time-consuming, so that other criteria, such as the generalized cross-validation (GCV) criterion, [7] and [23], and the Bayesian information criterion (BIC), [13] and [22], have also been adopted for selecting regularization parameters in identification of GRNs.

4. Dynamic Granger causality and DVAR models

4.1. Dynamic VAR models

The VAR model of (1) is only suitable for time-series microarray data with stationary network structure because $\boldsymbol{A}^{(i)}$ is fixed and doesn't vary with time. However, more and more biological experiments have shown that the transcriptional regulatory, protein-protein interactions, and other biological activities, exhibit substantial time-variant patterns throughout the cell cycle or other experimental period [26]-[28]. Therefore, the dynamic changes in GRNs cannot be inferred from the conventional VAR model.

To address the problem, a dynamic VAR (DVAR) model is proposed to describe the time-series genomic data as

$$\boldsymbol{x}(n) = \sum_{i=1}^{p} \boldsymbol{A}^{(i)}(n) \boldsymbol{x}(n-i) + \boldsymbol{e}(n) , \qquad (14)$$

where $\boldsymbol{A}^{(i)}(n) = \begin{bmatrix} a_{1,1}^{(i)}(n) & a_{1,2}^{(i)}(n) & \cdots & a_{1,K}^{(i)}(n) \\ a_{2,1}^{(i)}(n) & a_{2,2}^{(i)}(n) & \cdots & a_{2,K}^{(i)}(n) \\ \vdots & \vdots & \ddots & \vdots \\ a_{K,1}^{(i)}(n) & a_{K,2}^{(i)}(n) & \cdots & a_{K,K}^{(i)}(n) \end{bmatrix}$, and $\boldsymbol{e}(n)$ is the error vector with zero-mean and time-varying covariance matrix

$$\Sigma(n) = \begin{bmatrix} \sigma_{1,1}^2(n) & \sigma_{1,2}^2(n) & \cdots & \sigma_{1,K}^2(n) \\ \sigma_{2,1}^2(n) & \sigma_{2,2}^2(n) & \cdots & \sigma_{2,K}^2(n) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K,1}^2(n) & \sigma_{K,2}^2(n) & \cdots & \sigma_{K,K}^2(n) \end{bmatrix}.$$

Since the DVAR coefficient matrix $\boldsymbol{A}^{(i)}(n)$ is time-dependent, the time-varying Granger causality in the dynamic GRNs can be revealed. More precisely, in $\boldsymbol{A}^{(i)}(n)$, the value of $a_{j,l}^{(i)}(n)$ ($j = 1,2,\cdots,K$ and $l = 1,2,\cdots,K$) implies the Granger-causality between gene x_l and gene x_j at the given time instant n.

4.2. Identification of DVAR models

The DVAR model is actually a time-varying linear regression which takes the form

$$\boldsymbol{Y}(n) = \boldsymbol{X}(n)\boldsymbol{\beta}(t) + \boldsymbol{e}(n), \tag{15}$$

where $\boldsymbol{\beta}(t) = [\boldsymbol{A}^{(1)}(t), \boldsymbol{A}^{(2)}(t), \cdots, \boldsymbol{A}^{(p)}(t)]^T \in \boldsymbol{R}^{pK\times K}$. The identification of DVAR models is more difficult than identification of VAR models, because the DVAR coefficient matrix should be estimated based on data measured at one time point. To achieve an accurate estimation with small variability, several deterministic or stochastic models are proposed to account for the variations of the DVAR coefficient matrix and they lead to three categories of identification methods for DVAR models [29]: 1) basis expansion modeling (BEM), 2) Kalman filtering (KF), and 3) weighted least-squares (WLS).

1). In the BEM method, an explicit deterministic model of the coefficient variations is assumed, and the time-varying coefficients are approximated by a linear combination of known basis functions of time. The performance of the BEM method is greatly dependent on the basis functions and the expansion level, and their optimal selection is not easily accessible. This method has been used for inferring time-varying GRNs in [8]. The authors of [8] adopted a wavelet basis function to model the variation of VAR coefficient $a_{j,l}^{(i)}(n)$ as

$$a_{j,l}^{(i)}(n) = \sum_{\zeta=0}^{Z}\sum_{\gamma=0}^{2^{\zeta}-1} c_{\zeta,\gamma}^{(i)}\psi_{\zeta,\gamma}(n), \tag{16}$$

where $\psi_{\zeta,\gamma}(n) = 2^{\zeta/2}\psi(2^{\zeta}n - \gamma)$ is the wavelet basis function dilated and shifted from a mother wavelet function $\psi(n)$, Z is the expansion level, and $c_{\zeta,\gamma}^{(i)}$ is the time-independent wavelet coefficients. By substituting (16) into (14), the estimation of $a_{j,l}^{(i)}(n)$ is reduced to the estimation of $c_{\zeta,\gamma}^{(i)}$, lowering the number of parameters to be estimated significantly. However, their work did not consider automatic variable selection during the estimation, and only a few variables were selected to build the DVAR model based on prior knowledge.

2). The Kalman filtering method employs a state-space model (SSM) to describe the coefficient variations, where the current coefficients are treated as the system state and are obtained by a linear transformation of the previous coefficients or state plus an innovation variable through the SSM. Given the prior information of the coefficient variations in form of the SSM, the Kalman filtering is an optimal recursive estimator in the minimum mean-square error sense. However, such prior knowledge is often vague in real-world applications, so that the Kalman filtering is often accompanied by an expectation-maximization (EM) algorithm to approximate the SSM parameters [14]. According to what we now so far, the Kalman filtering is still not an option for identifying DVAR models for GRN inference purpose. The authors of [14] proposed an alternative approach to combine the SSM and the VAR models (i.e., the gene expression data were modeled as the sum of a latent variable and a measurement noise while the latent variable was described by a VAR model) and employed the lasso method to obtain sparse solution, but they did not consider the dynamic changes of GRNs.

3). The WLS method is similar to the conventional OLS method, except that kernels or windows are employed to assign larger weights to local data and smaller weights to remote data. The time-varying coefficients are then estimated by minimizing a weighted sum of squared estimation errors. It has been shown in [29] that WLS can be viewed as a special case of a more general local polynomial modeling (LPM) estimator for the time-varying linear regression model. In fact, the WLS estimator is LPM with the polynomial order equal to zero. The selection of the window size or bandwidth is critical to the performance of the WLS/LPM method, and automatic data-driven kernel bandwidth selection for WLS is a difficult problem, which considerably hinders its practical application. The WLS estimator is a common method for addressing the similar VAR model identification problem in other applications. However, the WLS method has not been used to infer GRNs, which may be due to the fact that the number of time points of time-series genomic data is quite limited and the selection of window length is therefore difficult.

In summary, the identification of DVAR models for inferring dynamic GRNs is still an open problem, and few successful applications were reported. The conventional identification methods for time-varying linear regression models face many challenges when handling time-series genomic data, which have high dimensionality and limited time points. Therefore, it is necessary to develop more capable and effective DVAR identification methods, which should possess the ability of automatic variable selection and have reliable performance for short-duration data.

5. Conclusions

Granger causality (via VAR modelling) is an important approach to infer gene regulatory network from time-series genomic data. The discovering of time-invariant Granger causality (i.e., the identification of VAR models) has been extensively studied in systems biology. The least-squares estimator with regularization techniques showed good properties for automatic variable selection for identification of VAR models. However, it is still difficult to identify DVAR models due to the limitations of the identification methods and properties of time-series genomic data. Hence, advanced methods for identification of DVAR models are needed to deal with time-series genomic data.

Acknowledgements

This paper was supported by the University of Hong Kong CRCG Small Project Funding.

References

[1] E. H. Davidson, Genomic regulatory systems, Academic Press, San Diego, CA, USA, 2001.

[2] H. Li, J. Xuan, Y. Wang, and M. Zhan, "Inferring regulatory networks", Front. Biosci., Vol. 1. No. 13, pp. 263-275, Jan 2008.

[3] G. Karlebach and R. Shamir, "Modelling and analysis of gene regulatory networks", Nat. Rev. Mol. Cell Biol., Vol. 9, pp. 770-780, Oct 2008.

[4] C. Sima, J. Hua, and S. Jung, "Inference of gene regulatory networks using time-series data: a survey", Curr. Genomics, Vol. 10, No. 6, pp. 416-429, Sep 2009.

[5] N. Sun and H. Zhao, "Reconstructing transcriptional regulatory networks through genomics data", Stat. Methods Med. Res., Vol. 18, No. 6, pp. 595-617, Dec 2009.

[6] N. D. Mukhopadhyay and S. Chatterjee, "Causality and pathway search in microarray time series experiment", Bioinformatics, Vol. 23, No. 4, pp. 442-449, Feb 2007.

[7] R. Opgen-Rhein and K. Strimmer, "Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process", BMC Bioinformatics, Vol. 8 Suppl 2, No. S3, May 2007.

[8] A. Fujita, *et al.*, "Time-varying modeling of gene expression regulatory networks using the wavelet dynamic vector autoregressive method", Bioinformatics, Vol. 23, No. 13, pp. 1623-1630, Jul 2007.

[9] A. Fujita, *et al.*, "Modeling gene expression regulatory networks with the sparse vector autoregressive model", BMC Syst. Biol., Vol. 1, No. 39, Aug 2007.

[10] T. Shimamura, *et al.*, "Recursive regularization for inferring gene networks from time-course gene expression profiles", BMC Syst. Biol., Vol. 3, No. 41, Apr 2009.

[11] J. F. Feng, D. Yi, R. Krishna, S. Guo, and V. Buchanan-Wollaston, "Listen to genes: dealing with microarray data in the frequency domain", PLoS ONE, Vol. 4, No. 4, Apr 2009.

[12] C. Zou, K. J. Denby, and J. F. Feng, "Granger causality vs. dynamic Bayesian network inference: a comparative study", BMC Bioinformatics, Vol. 10, No. 122, Apr 2009.

[13] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, "Grouped graphical Granger modeling for gene expression regulatory networks discovery", Bioinformatics, Vol. 25, No. 12, pp. i110-i118, Jun 2009.

[14] K. Kojima, *et al.*, "A state space representation of VAR models with sparse learning for dynamic gene networks", Genome Inform., Vol. 22, pp. 56-68, Jan 2010.

[15] J. Zhu, *et al.* "Characterizing dynamic changes in the human blood transcriptional network," PLoS Comput. Biol., Vol. 6, No. 2, Feb 2010.

[16] X. Wang, M. Wu, Z. Li, and C. Chan, "Short time-series microarray analysis: methods and challenges", BMC Syst. Biol., Vol. 2, No. 58, Jul 2008.

[17] S. Ma and J. Huang, "Penalized feature selection and classification in bioinformatics", Brief. Bioinform., Vol. 9, No. 5, pp. 392-403, Sep 2008.

[18] C. Li and H. Li, "Network-constrained regularization and variable selection for analysis of genomic data", Bioinformatics, Vol. 24, No. 9, pp. 1175-1182, May 2008.

[19] L. Wang, G. Chen, and H. Li, "Group SCAD regression analysis for microarray time course gene expression data", Bioinformatics, Vol. 23, No. 12, pp. 1486-1494. Jun 2007.

[20] A. N. Tikhonov and V. Y. Arsenin, Solutions of ill-posed problems, Winston and Sons, Washington DC, USA, 1977.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. Ser. B, vol. 58, no. 1, pp. 267-288, 1996.

[22] H. Zou and T. Hastie "Regularization and variable selection via the elastic net", J. R. Stat. Soc. Ser. B, Vol. 67, pp. 301-320. 2005.

[23] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties", J. Am. Stat. Assoc., vol. 96, no. 456, pp. 1348-1360, Dec 2001.

[24] R. D. Leclerc, "Survival of the sparsest: robust gene networks are parsimonious", Mol. Syst. Biol., Vol. 4, No. 213, Aug 2008

[25] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," Econometrica, Vol. 37, No. 3, pp. 424-438, 1969.

[26] N. M. Luscombe, *et al.*, "Genomic analysis of regulatory network dynamics reveals large topological changes", Nature, Vol. 431, pp. 308-312, Sep 2004.

[27] J. Seok, W. Xiao, L. L. Moldawer, R. W. Davis, and M. W. Covert, "A dynamic network of transcription in LPS-treated human subjects", BMC Syst. Biol., Vol. 3, No. 78, Jul 2009.

[28] R. Jothi, *et al.*, "Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture", Mol. Syst. Biol., Vol. 5, No. 294, Aug 2009.

[29] S. C. Chan and Z. G. Zhang, "Local polynomial modeling and bandwidth selection for time-varying linear models," Proceeding of ICICS2009 Conference, Macau, China, Dec 2009.