Lasso Granger Causal Models: Some Strategies and Their Efficiency for Gene Expression Regulatory Networks

Chapter · February 2015

DOI: 10.1007/978-3-319-15144-1_4

CITATIONS
2

READS
1,156

2 authors, including:

Katerina Hlavackova-Schindler
University of Vienna
78 PUBLICATIONS **318** CITATIONS

SEE PROFILE

Granger Lasso Causal Models in Higher Dimensions - Application to Gene Expression Regulatory Networks

Kateřina Hlaváčková-Schindler and Hamed Bouzari

The Gregor Mendel Institute of Molecular Plant Biology,
Austrian Academy of Sciences, Doktor-Bohr-Gasse 3, 1030 Vienna, Austria
katerina.schindler@gmail.com, hamed.bouzari@oeaw.ac.at

Abstract. Granger causality (GC), based on a vector autoregressive model, is one of the most popular methods in uncovering the temporal dependencies among time series. The original Granger model is able to detect only linear causal dependencies and many approaches were recently developed to extend it to the non-linear modeling. The method Copula-Granger from Bahadori and Liu in 2012 introduces non-linearity into the causality modeling by representing the data distribution by copulas. The detection of causality of gene regulatory networks (GRN) from experimental data, such as gene expression measurements, is a challenging problem, being solved by various computational methods with various success. We applied the Granger Lasso method, the Copula Granger method and the combination of dynamic Bayesian Networks with ordinary differential equation method (ODE-DBN) to cell division cycle gene expression data from the human cancer cell line (HeLa) for a regulatory network of 19 selected genes. We tested the causal detection ability of the methods with respect to the selected benchmark network. We compared the performance of the mentioned methods or various statistical measures. All three methods are scalable and can be easily extended to higher dimensions. The results of both Granger Lasso and Copula Granger outperformed the ODE-DBN both in terms of precision and the computational time. We conclude that the DBN combined with ODE method are not feasible for large GRN because of the computational intensity of the methods and surprisingly low precision. This type of methods is more feasible for modeling of local dynamics within a small genetic regulatory networks, rather than for detection of causal relationships in a large genetic regulatory network. We believe that the assumption of Gaussian processes, on which are DBN based, is in larger genetic regulatory networks violated.

Keywords: Granger causality, graphical Granger Lasso method, Copula Granger method, gene expression data, gene regulatory network.

1 Introduction

Granger causality (GC), based on a vector autoregressive model, is one of the most popular methods in uncovering the temporal dependencies among time se-

The Proceedings of ECML/PKDD 2013 Workshop
Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE)
September 23, 2013, Prague, Czech Republic

ries. The original Granger model is able to detect only linear causal dependencies and many approaches were recently developed to extend it to the non-linear modeling. The method Copula-Granger from Bahadori and Liu in 2012 introduces non-linearity into the causality modeling by representing the data distribution by copulas. It is a computationally fast method with respect to the size of the network.

Transcriptional regulation in a cell is a process with a complex non-linear dynamics. Models of transcriptional regulation are commonly depicted in the form of a network, where directed connections between nodes represent the regulatory interactions. The goal of these models is to infer the structure of gene regulatory networks (GRN) from experimental data. Biological samples are usually profiled using gene expression microarrays (GE) and the measured microRNA (mRNA) levels provide a quantitative information to assess molecular control mechanisms. A gene can be computationally represented by a single data value (row) consisting of d measurements $x^i = (x_1^i, \ldots, x_d^i)$. An experiment (sample) y is a single microarray experiment corresponding to a single column in the GE matrix, $y = (x_j^1, \ldots, x_j^n)^T$ where n is the number of genes in the data set. A gene expression profile from microarrays has typically 5000 to 100000 variables and just $15 - 100$ measurements. The detection of inference (causality, in other words) of GRN from experimental data, such as gene expression measurements, is a challenging problem, being solved by various computational methods with various success. The most applied method to model for causal relationships in gene regulatory networks from experimental data are the so called dynamic Bayesian networks (DBN), for example [23]. The exact models for small regulatory networks are commonly approximated by ODE, which can be obtained as the expectation of the chemical master equation under certain assumptions. A number of different modeling approaches using ODE with Bayesian modeling have been proposed, including, among others, Cao and Zhao, [9], Bansal et al., [6] and Zou and Conzen [25]. In our paper we consider the model from Äijö and Lähdesmäki [1]. The authors have in [1] shown that the combination of the DBN and ODE methods outperforms the causality detection in small gene regulatory networks. The drawback of these two models considered separately as well as of their combination is their exponential computational time with respect to the size of the networks, making the computation costly for large networks. Several other methods modeling causal relationships have been recently proposed and applied to gene expression data, such structural equation models, probabilistic Boolean networks, fuzzy controls and differential equations. These methods are mainly applied to small genetic networks and will not be discussed in this paper.

For these reasons, we focused on the class of GC methods which have shown to have a high precision and fast computation even for causality detection in large networks. We applied the Granger Lasso method, the Copula Granger method and the combination of dynamic Bayesian Networks with ordinary differential equation method (ODE-DBN) to cell division cycle gene expression data from the human cancer cell line (HeLa) for a regulatory network of 19 selected genes. We tested their causal detection ability with respect to a benchmark network.

We compared the performance of the mentioned methods on various statistical measures. The results of both Granger Lasso and Copula Granger Lasso outperformed the ODE-DBN both in terms of precision and the computational time. The computation of both Granger methods was a few seconds on a common PC workstation, while the DBN-ODE method needed for each gene two minutes of real time. Concerning the robustness of the methods with respect to noise, the precision of the results was tested with various levels of noise on data. The precision of the inference results was the best for Copula Granger Lasso method, while the DBN-ODE method issued into the over fitting and spurious results already with low levels of noise on data. Both Copula Granger Lasso and Granger Lasso methods are scalable methods and can be easily extended to higher dimensions. We conclude that the DBN combined with ODE method are not feasible for large GRN because of the computational intensity of the methods and surprisingly low precision. This type of methods is more feasible for modeling of local dynamics within a small genetic regulatory networks, rather than for detection of causal relationships in a large genetic regulatory network. We believe that the assumption of Gaussian processes, on which are DBN based, is in larger genetic regulatory networks violated.

2 Granger Causality

Causality has been in the literature defined in many ways. We consider the concept of causality as a time-dependent relationship among time series, as time is relevant in the biological experiments. The most used conception of time dependent causality is the Granger causality GC [12] which is based on the probabilistic notion of causality, and is defined as follows: An event X is a cause to the event Y if (i) X occurs before Y, (ii) likelihood of X is non zero, and (iii) likelihood of occurring Y given X is more than the likelihood of Y occurring alone.

Granger developed this conception of causality into the mathematical scheme based on a VAR. As Granger put it, a consequence of statements (i) and (ii) is that the causal variable can help to forecast the effect variable after other data has been first used [12]. This restricted sense of causality, referred to as Granger causality, characterizes the extent to which a process X_t is leading another process Y_t, and builds upon the notion of incremental predictability. It is said that the process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather then only past values of Y_t. The standard test of GC developed by Granger is based on a linear vector auto-regressive model (VAR)

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \xi_t, \tag{1}$$

where ξ_t are uncorrelated random variables with zero mean and variance σ^2, L is the specified number of time lags, and time $t = L+1, \ldots, N$. The null hypothesis

that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \ldots, L$, reducing Eq. (1) to

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \tilde{\xi}_t. \quad (2)$$

This model leads to two well-known alternative test statistics, the Granger-Sargent and the Granger-Wald test. The equations 1 and 2 express the bivariate causality among time series X and Y. This linear framework for measuring and testing causality has been widely applied not only in economy and finance, but also in diverse fields of natural sciences such as climatology or neurophysiology. The concept of GC can be extended to more than two time series so that the vector autoregressive model VAR is replaced by a multivariate vector autoregressive model MVAR [13]. These models are called graphical Granger models and will be discussed in the following.

Since the conception of Granger causality can detect only linear causal relationships, various nonlinear extensions of GC were proposed, for example the nonlinear predictors based on so-called radial basis functions [8]. Ancona end Marinazzo applied this idea to GC and introduced the so called kernel Granger methods, [2] and [18]. Other extension of GC are from Chen et al. and can be found in [10]. In this paper we will in the following deal with the extension of Granger method from [4].

3 Dynamic Bayesian Networks and Ordinary Differential Equations

A Bayesian network [14] or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclig graph (DAG). A dynamic Bayesain network is a Bayesian network which relates variables to each other over adjacent time steps. A causal Bayesian network is a network with an explicit requirement of causal relationships. The additional semantics of the causal networks specify that if a node X is actively caused to be in a given state x, then the probability density function changes to the one of the network obtained by cutting the links from $X's$ parents to X, and setting X to the caused value x. Using these semantics, one can predict the impact of external interventions from data obtained prior to intervention.

Zou et al. in [24] recently compared the (multivariate) GC and dynamic Bayesian networks on inference problem for both synthesized and experimental data, including GE microarray data. They concluded, that for a small sample size, the inference of DBN is better than of the GC approach, otherwise the GC performs better (in the sense of common precision measures). The drawback of dynamic Bayesian networks was is their computational intensity for large graphs.

In this paper we consider the method from Äijö and Lähdesmäki [1] applying DBN. For experimental comparison to other methods, we considered the publicly available Matlab implementation of this method by the authors. The method

from [1] is based on ODE method and uses non-parametric modeling of molecular kinetics and Bayesian analysis. The method can use both steady-state and time-series data. The experimental results of this methods demonstrated in [1] that this approach provides more accurate network structure predictions than other commonly used ODE and Bayesian methods. Therefore we prefer to use this method for comparison instead of considering ODE and DBN separately.

The model of Äijö and Lähdesmäki, which we call here ODE-DBN, is based on the commonly used first-order ODE model where the ODE describing the unknown function f responsible for the gene regulation is replaced by a first-order approximation of the rates of gene expression as

$$\frac{dx_i(t_k)}{dt} \approx \Delta x_i(t_k) = \frac{x_i(t_{k+1}) - x_i(t_k)}{t_{k+1} - t_k} \tag{3}$$

for a given set of measurement time points. The scalar $x_i(t)$ denotes the expression of gene i at time t and the vector $\mathbf{x}_i(t)$ denotes the expressions of genes that regulate gene i. The method uses Gaussian processes to learn the unknown regulation function from the data. The values of the unknown function f are modeled by a Gaussian process

$$f(x) \approx GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{4}$$

where GP denotes a Gaussian process, $m(x)$ is a mean function and $k(\mathbf{x}, \mathbf{x}')$ is a covariance function. It is further assumed that the mean function is identically zero. It is further assumed that there is a normal i.i.d. additive noise on the measurements and the predictions of the GP are computed analytically from the marginal likelihood, given the covariance matrices (more details in [1]). The method ODE-DBN has two goals: estimation of the non-parametric kinetic models and inference of the network structure. For a given model structure, the regulatory function can be estimated by means of a Gaussian process with the given covariance matrix. Bayesian model structure selection, where the goal is to choose explanatory variables $\mathbf{x_i}$ for each gene i, can be obtained via the marginal likelihood. The posterior probability of a given model can be obtained by applying Bayes theorem. The actual inference procedure is done separately for each gene in the network. That is, for each gene, the model of the ODE is being fit with different combinations of explanatory variables $\mathbf{x}$ and the posterior probabilities are computed. The posterior probabilities of network models are summarized using a square connection matrix, where the (i, j) element represents the posterior probability that gene j is regulated by gene i. Each element of the connection matrix can be computed by summing posterior probabilities of all networks that contain a directed connection from x_i to x_j. The method has an exponential computational complexity of order $O(n2^n)$ where n is the number of genes. The authors tested the method on small networks with five genes on yeast data with time series measurement with length 20 or 15. Another experiment was done with the network of 100 genes and was computed by means of distributed computing. The method was compared to both single ODE method (TSNI method from Bansal et al., [7]) and to single Bayesian networks (BANJO

method, [23]) as well as to the Bayesian networks from Zou and Conzen [25] with respect to common precision recall statistics. Based on the experiments in [1], the ODE-DBN method outperforms the inference of the TSNI method as well as of the method from Zou and Conzen on the time-series data and dynamic and static Bayesian networks on time-series and steady-state data.

4 Graphical Granger Lasso Models

Microarrays of gene expression data are represented by high-dimensional vectors and have short time series of the observations. The related parameter estimation problems are therefore ill-posed, so the straightforward application of the GC method is unfeasible [18]. As a remedy, the Granger method with a penalization method is applied.

Fujita et al. in [11] in 2007 applied a (multivariate) sparse vector autoregressive model SVAR with lasso regression, called a graphical GC.

Consider a graphical model with n variables (can be the number of genes), observed over T time points, and let d be the order of the VAR model or the effective number of time lags ($d = T - 1$). Let X^t denote the design matrix corresponding to t-th time point, and X_i^t be its i-th column.

The Lasso estimate of the graphical Granger model is found by solving the following estimation problem for $i = 1, \ldots, n$:

$$\arg \min_{\theta^t \in R^n} \|X_i^T - \sum_{t=1}^{d} X^{T-t}\theta^t\|_2^2 + \lambda \sum_{t=1}^{d}\sum_{j=1}^{n} |\theta_j^t| w_j^t. \tag{5}$$

The formula (5) has been studied in many variations, which mainly concern the form of the penalty function. Shojaie and Michalidis combined GC with the so called truncating lasso penalty or with so called adaptive lasso penalty [20] and proved their consistency in [21]. Lozano et al. studied the graphical Granger models with group Lasso penalization function in [17]. The complexity of the Granger Lasso depends on the complexity of the selected optimization method (non-convex optimization). The complexity of the optimization method applied in [20] is quadratical with respect to n, the number of genes. In this paper we will consider the VAR with Lasso regression from [3] and extend it to the multivariate case. The implementation of Lasso Granger from [3] has the computational quadratical complexity with respect to n.

5 Copula Granger method

Bahadori und Liu in [4] proved that Granger causality (i.e without any regularization) cannot be consistent in a high-dimensional regime, where insufficient number of observations is given. Utilizing the high dimensional advantages of Lasso regularization, they introduced the semi-parametric approach Copula Granger and showed its consistency in high dimensions as well as its ability to efficiently capture nonlinearity in the data.

The G-NPN model is defined as follows. One says a set of time series $X = (X_1, \ldots, X_n)$ has G-NPN distribution $G - NPN(X, B, F)$ if there exist functions $\{F_j\}_{j=1}^n$ such that $F_j(X_j)$ for $j = 1, \ldots, n$ are jointly Gaussian and can be factorized according to the VAR model with coefficients $B = \{\beta_{i,j}\}$. More specifically, the joint distribution for the transformed random variables $Z_j \hat{=} F_j(X_j)$ can be factorized as following

$$p_Z(z) = \mathcal{N}(z(1, \ldots, L)) \times \prod_{j=1}^{n} \prod_{t=L+1}^{T} p_{\mathcal{N}}(z_j(t); \sum_{t=1}^{n} \beta_{i,j}^T z_i^{t,Lagged}, \sigma_j)$$

where $p_{\mathcal{N}}(z; \mu, \sigma)$ is the Gaussian density function with mean μ and variance σ^2 and $z_i^{t,Lagged} = [z_i(t-L), \ldots, z_i(t-1)]$ is the history of z_i up to time t, L is the maximal time lag, and $\beta_{\mathbf{i,j}} = [\beta_{\mathbf{i,j}}(\mathbf{1}), \ldots, \beta_{\mathbf{i,j}}(\mathbf{L})]$ is the vector of coefficients modeling the effect of time series z_j on the target time series.

The causality is defined as follows: the time series z_j Granger causes z_i if at least one value in the coefficient vector $\beta_{\mathbf{j}}$ is nonzero by statistical significant sense.

Based on the copula method from [16], the G-NPN model aims to separate the marginal properties of the data from its dependency structure. The marginal distribution of the data can be efficiently estimated using the non-parametric techniques with exponential convergence rate [4].

Learning G-NPN models consists of three steps: (i) Find the empirical marginal distribution for each time series $\hat{F}_i$. (ii) Map the observations into the copula space as $\hat{f}_i(X_i^t) = \hat{\mu}_i + \hat{\sigma}_i \Phi^{-1}(\hat{F}_i(X_i^t))$. (iii) Find the GC among $\hat{f}_i(X_i^t)$. In practice the Winsorized estimator of the distribution function is used, to avoid the large numbers $\Phi^{-1}(0^+)$ and $\Phi^{-1}(1^-)$, [4].

Bahadori and Liu have proved that the convergence rate for Copula-Granger is the same as the one for Lasso. This suggests efficient Granger graph learning in high dimensions via Copula-Granger.

The Copula Granger Lasso method was tested with respect to the Granger method and Granger Lasso method on synthetic and experimetal data (Twitter application) with the best precision for Copula Granger Lasso method [4]. To our knowledge, any comparison of the Copula Granger Lasso method to Granger Lasso together with DBN has not been published yet.

6 Application of Granger Lasso Methods to Gene Regulatory Networks: Experimental Results

We investigated the three above discussed methods on genetic regulatory networks. Our selected data set is from the gene database of the genome-wide expression of cell cycle genes in human cancer cell lines (HeLa) analyzed by Whitfield et al. [22]. We used the preselected 19 genes, whose gene regulatory network was reconstructed based on the biological experiments o Li et al. [15]. This causal network we used as a benchmark structure for comparison of the discussed methods. The 19 genes, which play a substantial role at the human cancer

cell lines, have the following names: PCNA, NPAT, E2F1, CCNE1, CDC25A, CDKN1A, BRCA1, CCNF, CCNA2, CDC20, STK15, BUB1B, CKS2, CDC25C, PLK1, CCNB1, CDC25B, TYMS, DHFR. The causal structure for these genes identified by Li et al. was adopted from the figure from [17] and is in Figure 1.

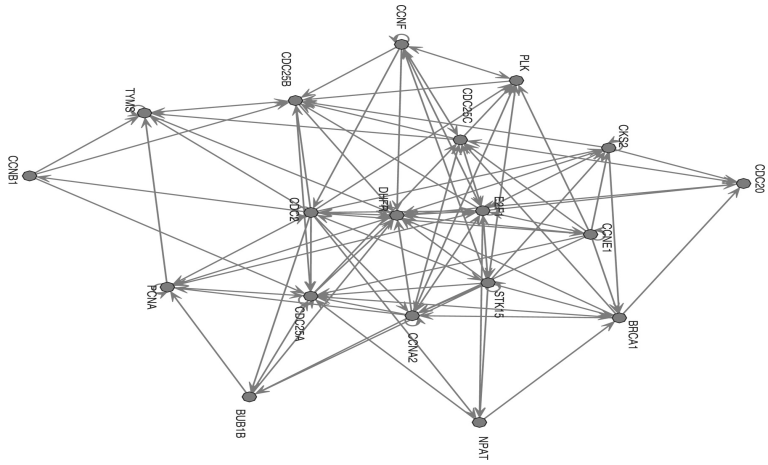

Fig. 1. Causal structure from biological experiments for 19 selected genes (adopted from [17])

The gene expressions in the database from [22] for the 19 genes were given for 48 time observations. In our experiments we used the following Matlab codes: for the inference measured by the combination ODE and Bayesian method (ODE-DBN) we used the code GP4GRN from Tarmo Äijö [1] which we extended with graphical outputs using publicly available Matlab graphical software Graphviz4Matlab Version 2.24. For experiments with Granger Lasso method we used the code from Bahadori [5] and extended it to the multivariate case (i.e from bivariate causal relationships to the multivariate ones). This we extended with the same graphical outputs using Graphviz4Matlab. Similarly we extended the code for Copula Granger Lasso method from the same source.

By TP, TN, FP, FN, respectively we denote the number of true positive outcomes, of true negative outcomes, of false positive outcomes and of false negative outcomes, respectively. We compared these methods for the HeLa data with respect to these classification performance measures: balanced classification rate $BCR = \frac{1}{2}(TP/(TP+FN) + TN/(TN+FP))$, precision $PREC = TP/(TP+FP)$, true positive rate $TPR = TP/(TP+FN)$ and classification accuracy $CA = (TP+TN)/(TP+TN+FP+FN)$. For each of these criteria we ran optimization processes for each of the method.

All the measures had significantly comparable results for the GP4GRN method and Granger Lasso method. It is not surprising, since the number of genes in the

network was small and the strength of the Granger Lasso causality is for large graphs.

The computational time of GP4GRN was very demanding, for each gene ca 2 min of real time at a PC workstation with 64-bit processor, which in our concrete case was 38 minutes. Granger Lasso require only a few seconds run for our simulation. The Copula Granger Lasso, which was running also a few seconds, had the best precision with respect to the above mentioned measures with significant differences to the other methods, see Table 1.

For example, the BCR measure for GP4GRN was 0.5774, for Granger Lasso 0.05789 and for Copula Granger Lasso 0.8006.

Concerning the validation of the methods, it can be done in many ways. The main goal of this paper was to compare the three methods on the well-known HeLa data from [22] which has been already applied for testing other methods [17], [20]. We tested the robustness of the methods with respect to noise with normal distribution having values of order $10^{-6}, 10^{-5}, \ldots, 10^{-1}$ which was added to the original data, so our statements about robustness concern only this type and orders of perturbation. Since the exact distribution of the error function of GE data is unknown, one can do similar tests for other types of noise, for example with Laplace distribution [19].

The precision of the inference results was the best for Copula Granger Lasso method, while the GP4GRN method issued into the overfitting and spurious results (the output GRN was overfitted with causal connections) already with low levels of noise on data. Concretely, the levels of random normal noise were of order $10^{-6}, 10^{-5}, \ldots, 10^{-1}$ on the time series with 48 time measurements of the data from [22]. The performance of the methods measured for BCR and CA measures of the non-perturbed data are summarized in the Table 1.

Table 1. Comparison of the precision of the three methods for various criteria

Method	GP4GRN	Granger Lasso	Copula Granger Lasso
BCR	0.5774	0.5789	0.8006
CA	0.7507	0.5789	0.8006
TP	95 overfitting	38	58

The performance of the methods measured by values for BCR, CA and TP of the perturbed data with levels of normal noise were of order $10^{-6}, 10^{-5}, \ldots, 10^{-1}$ is summarized in the Table 2.

Table 2. Comparison of the precision of the three methods with data perturbed by random noise of order (P. order) 10^{-6} and of 10^{-5} for various criteria; the values in the brackets are the deviations in per cent of the precision error from the precision error for unperturbed method.

P. order 10^{-6}	GP4GRN	Granger Lasso	Copula Granger Lasso
BCR	0.5850 (+1.31%)	0.6011(+3, 8%)	0.7895(−2, 6%)
CA	0.7507 (0%)	0.6011(+3, 8%)	0.7895(−2, 6%)
TP	95 overfitting	38	51
P. order 10^{-5}	GP4GRN	Granger Lasso	Copula Granger Lasso
BCR	0.5850(+1.31%)	0.5850(+0.7%)	0.7313(−8, 6%)
CA	0.7507(0%)	0.7507 (+29, 6%)	0.7313(−8, 6%)
TP	65 overfitting	96 overfitting	49

For the data perturbed at the order of $10^{-4}, \ldots, 10^{-1}$ of noise, both the GP4GRN and Granger Lasso methods exhibited a strong overfitting in the values of TP in the output GRNs, so the results of these methods were spurious. In case of Copula Granger Lasso, this method was robust to the perturbations of the data up to the order 10^{-1} and the precision results remained similar for the unperturbed data, see the Table 3.

Table 3. Comparison of change of the precision of Copula Granger Lasso with respect to BCR, CA and TP depending on the order of random noise on the data.

P. order	10^{-4}	10^{-3}	10^{-2}	10^{-1}
BCR	0.7867(−1.7%)	0.7837(−2.1%)	0.7701(−3.8%)	0.759(−6.4%)
CA	0.7867(−1.7%)	0.7837(−2.1%)	0.7701(−3.8%)	0.759(−6.4%)
TP	51	51	47	40

Figures 2 to 5 show the results of the codes for GP4GRN, Granger Lasso and Copula Granger Lasso respectively, in the gridded form; Figure 5 is the gridded graph corresponding to the graph in Figure 1.

7 Conclusion

We have tested the causality detection of three methods, a DBN-ODE based method, Granger Lasso and Copula Granger Lasso on gene regulatory networks with a genetic data set (HeLa) given by microarrays of gene expression data. The best method with respect to the precision and computational costs was Copula Granger Lasso, which as a non-linear method was able to detect the most causal relationships. Both Copula Granger Lasso and Granger Lasso methods are scalable methods and can be easily expanded to higher dimensions. Because of the low precision of GP4GRN and high computationally costs in our experiments, we conclude that GP4GRN is not feasible for large gene regulatory networks. This method seems to be more appropriate for modeling of local dynamics within a small genetic regulatory network, rather than for detection of general inference relationships in a large genetic regulatory network. We believe that the assumption of Gaussian processes, on which are dynamic Bayesian networks based, is in genetic regulatory networks violated and this violation is more transparent with increasing the size of the network.

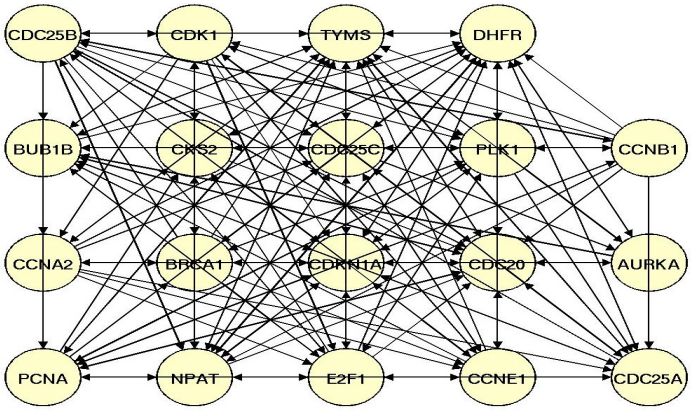

Fig. 2. Causal structure for selected 19 selected genes achieved by GP4GRN: the output graph shows overfitting (too many causal connections)

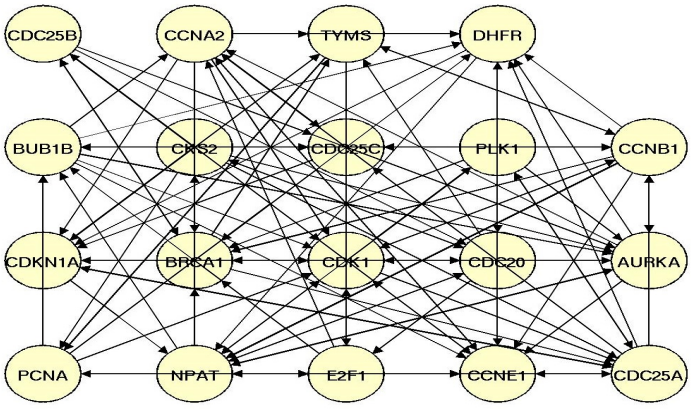

Fig. 3. Causal structure for selected 19 selected genes achieved by Granger Lasso method

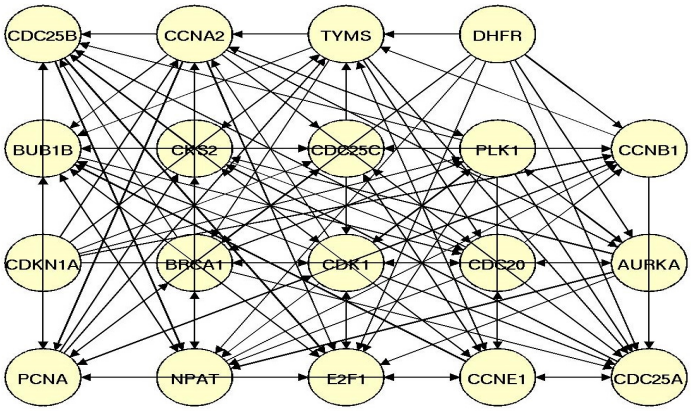

Fig. 4. Causal structure for selected 19 selected genes achieved by Copula Granger Lasso method

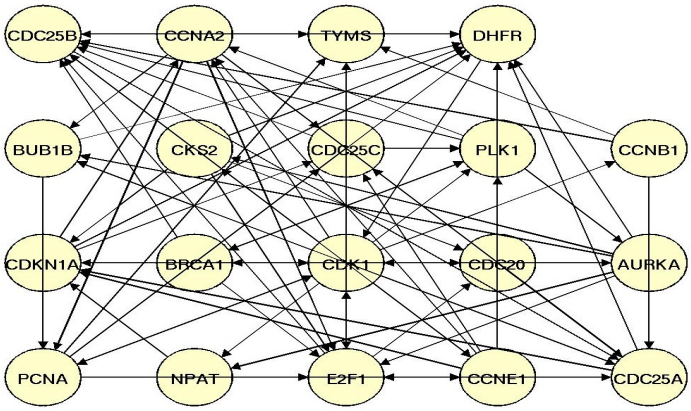

Fig. 5. Benchmark: the gridded graph of the causal structure for selected 19 selected genes from Figure 1

References

1. Äijö, T., Lahdesmäki, H.: Learning Gene Regulatory Networks from Gene Expression Measurements Using Non-Parametric Molecular Kinetics. Bioinformatics, no 22, Vol. 25 (2009)
2. Ancona, N., Marinazzo, D., Stramaglia, S.: Radial Basis Function Approach to Nonlinear Granger Causality of Time Series, Physical Review E 70 (2004)
3. Arnold, A., Liu, Y., Abe, N.:T emporal Causal Modeling with Graphical Granger Methods. In Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD-07) 2007
4. Bahadori, T., Y. Liu, Y.: An Examination of Large-Scale Granger Causality Inference. SIAM Conference on Data Mining (SDM'13) 2013
5. Website, `http// www-scf.usc.edu/∼mohammab/codes/codes.htm`
6. Bansal M., Belcastro, V., Ambesi-Impiombato, A, di Bernardo, D.: How to Infer Gene Networks from Expression Profiles. Mol. Syst. Biol., 3, 78 (2007)
7. Bansal, M., Della Gatta, G., di Bernardo, D. : Inference of Gene Regulatory Networks and Compound Mode of Action from Time Course Gene Expression Profiles. Bioinformatics, 22, 815822 (2006)
8. Broomhead, D.S., Lowe, D.: Multivariate Functional Interpolation and Adaptive Networks, Complex Systems 2, 321 - 355 (1988)
9. Cao, J. and Zhao, H.: Estimating dynamic models for gene regulation networks. Bioinformatics, 24, 16191624, 2008.
10. Chen, Y., Rangarajan, G., Feng, J., Ding, M.: Analyzing Mulitiple Non- Linear Time Series with Extended Granger Causality. Phys. Lett. A 324, 26-35 (2004)
11. Fujita, A., Sato, J.R., Garay-Malpartida, H.M., Yamaguchi, R., Miyano, S., Ferreira, C.E. Modeling Gene Expression Regulatory Networks with the Sparse Vector Autoregressive Model, BMC Systems Biology, 1:37 (2007)
12. Granger, C.W.J.: Investigating Causal Relations by Econometric and Cross- Spectral Methods, Econometrica 37, 424-438 (1969)
13. Kaminski, M., Ding, M.: Truccolo, W.A., Bressler, S.L.: Evaluating Causal Relations in Neural Systems: Granger Causality, Directed Transfer Function and Statistical Assessment of Significance. Biol Cybern, 85, 14557 (2001)
14. Jensen, F.V.: An Introduction to Bayesian Networks, London, UCL Press (1996)
15. Li, X., Rao, S., Jiang, W., Li, C., Xiao, Y., Guo, Z., Zhang, Q., Wang, L., Du, L., Li, J., Li, L., Zhang, T., Wang, Q.K.: Discovery of Time-Delayed Gene Regulatory Networks Based on Temporal Gene Expression Profiling. BMC Bioinformatics, 7, 26 (2006)
16. Liu, H., Lafferty, J.D., Wasserman, T.: The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. Journal of Machine Learning Research, 10, 22952328 (2009)
17. Lozano, A.C., Abe, N., Liu, Y., Rosset, S.: Grouped Graphical Granger Modeling for Gene Expression Regulatory Networks Discovery. Vol. 25 ISMB, pp. i110-i118 (2009)
18. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel-Granger Causality and the Analysis of Dynamic Networks. Physical Review E, 77:056215 (2008)
19. Purdom, E., Holmes, S.P. Error Distribution for Gene Expression Data, Statistical Applications in Genetics and Molecular Biology, Vol. 4, 1, 16 (2005)
20. Shojaie, A. Michalidis, G.: Discovering Graphical Granger Causality Using the Truncating Lasso Penalty, Bioinformatics 26, 18: i517-i523 (2010)

21. Shojaie, A., Basu, S. Michalidis, G.: Adaptive Thresholding for Reconstructing Regulatory Networks from Time Course Gene Expression Data, Manuscript at www.biostat.washington.edu (2011)
22. Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown. P.O., Botstein, D.: Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. Moll.Biol.Cell., 13(6):1977-2000 (2002)
23. Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., Jarvis, E.D.: Advances to Bayesian Network Inference for Generating Causal Networks from Observational Biological Data. Bioinformatics, 20, 35943603 (2004)
24. Zou, C., Feng, J.: Granger Causality vs Dynamic Bayesian Network Inference: A Comparative Study. BMC Bioinformatics, 10:122 (2009)
25. Zou, M., Conzen, S.D.: A New Dynamic Bayesian Network (DBN) Approach for Identifying Gene Regulatory Networks from Time Course Microarray Data. Bioinformatics, 21, 7179 (2005)