Time-dependent Gene Network Modelling by Sequential Monte Carlo

Sergiy Ancherbak, Ercan E. Kuruoğlu, *Senior Member, IEEE* and Martin Vingron

Abstract—Most existing methods used for gene regulatory network modeling are dedicated to inference of steady state networks, which are prevalent over all time instants. However, gene interactions evolve over time. Information about the gene interactions in different stages of the life cycle of a cell or an organism is of high importance for biology. In the statistical graphical models literature one can find a number of methods for network modelling while the study of time varying networks is rather recent. Using synthetic time series dataset for a gene network, we show that a sequential Monte Carlo method, namely Particle Filtering method is capable of tracking gene expression data and infer time-varying networks online.

Index Terms—Bayesian network, gene expression data, gene network, particle filter, sequencial Monte Carlo, time series

1 INTRODUCTION

CHANGES in the gene expression in time and under different external/internal stimulus play an important role in protein production. It is well known that expression level of one gene can influence on that of another gene. Usually, genes with similar expression profiles are more likely to encode interacting proteins [1], [2]. It was also shown that the genes of experimentally derived protein complexes are often co-expressed [3]. Such variations in expression levels can have significant effect on the functions of the genes in a single cell or in a complex organism.

Thus, genes should be studied in a group, as a number of objects, which form a network, interacting with each other, and not as isolated entities. Interactions between genes have long been studied in model organisms in order to identify functional relationships among genes or their corresponding gene products [4], [5], [6]. Examples of model organisms studied in literature include *yeast* [6], [7] and *Drosophila Melanogaster* [8].

The interactions among genes in a network are not stationary during the life cycle of an organism. These relations evolve over time as shown experimentally in [7], [9]. Such relations/interactions can change over time depending on life period and/or various external factors [7], [8], [9], [10], [11]. Information about dynamically changing gene network in different stages of a life cycle is of high importance for biology. It plays an important role in understanding of human disease [12], [13], [14] and in designing personalized treatment plans.

Usually to model relations between genes in a network gene expression data are used. A large amount of gene expression data measured at a single time instant can be found in the literature [15], [16], [17], [18], [19], [20]. These data are obtained via microarray experiments which can measure thousands of genes of an organism, providing a "genomic" viewpoint on gene expression. To understand the temporal variations in relations among genes, we are interested in time series data, i.e. gene expression levels measured with time during the life cycle of an organism. Some sources present experimental data on the evolution of temporal sequence datasets for gene expression during the *yeast* cell cycle and the life cycle of *Drosophila Melanogaster* [7], [8], [21], [22], [23] and also studies on circadian rhythms in mammals [10], [11]. However, for most of them only one to ten different time series are available for each gene. Moreover, all experimental data are measured for a quite short time length. This lack of experimental information significantly limits the success of inference on network topology.

To simplify the analysis, many authors discretize measured gene expression values into two (expressed (1), not expressed (-1)) or three levels (under-expressed (-1), normal (0), and over-expressed (1), depending on whether the gene expression value is significantly lower than, similar to, or greater than certain threshold value, respectively) [24]. It is obvious that discretising the measured expression data important information can be lost. However, such simplification makes possible to specify with a certain probability at which discrete expression level a gene can be. Then the relations between genes in a network can be represented by a conditional probability table. This is one of the most commonly used representations which can describe any discrete conditional distribution.

Another popular representation is to use continuous variables of gene expression data. Unlike the case of discrete variables there is no way in the representation of all possible densities. An obvious choice for multivariate continuous variables is to use certain distributions, i.e. Gaussian one [25].

Due to lack of experimental data, most existing methods used for gene regulatory network reconstruction are concentrated on the inference of steady state networks

- *S. Ancherbak and E.E. Kuruoglu are with the Istituto di Scienza e Tecnologie dell'Informazione, "A. Faedo", CNR (Institute of the National Research Council of Italy), via G. Moruzzi 1, 56124, Pisa, Italy. E-mails: ancherbaks@gmail.com, ercan.kuruoglu@isti.cnr.it.*
- *M. Vingron is with the Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195, Berlin, Germany.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

[24], [26], [27], [28]. Others divide whole time-series se-sequences into a number of homogeneous subsequencies in such case that one averaged network is estimated for each of subintervals [29], [30]. In such cases, the dynamics in the gene network during the life cycle is represented by a number of frames averaged for the subintervals. This procedure again does not take into account information between adjacent time intervals.

In the statistical graphical models literature, one can find a number of methods for studying the network structure [31], [32], [33], [34], [35] while the study of time-varying networks is rather recent. In this work, we propose a linear vector regression model for time changes in a gene interaction network which we learn via a state of the art sequential Monte Carlo method, namely Particle Filtering. Such networks describe the conditional dependence structure between multiple interacting quantities, in our case expression values of different genes, which are changing in time. Moreover, our approach is capable of handling noise which is always present in experimental data. Therefore, the interactions, the signal of which in the data is strong, can be detected and represented in the network with certain probability. The Particle Filtering method was proposed recently for time-varying network modelling in [36] and its success was demonstrated on computer vision data. Differing from [36], the concentration of which is the tracking of the topology of the network directly, we track the changes in the dependence of gene expression values of various genes over time, from which we infer also the network topology.

Thus, the motivation of this work is to model varying gene interactions over the life time of an organism or a biological process such as circadian cycles.

2 METHODS

2.1 Time Series Data Modelling

The great interest in signal processing, computational biology, finance, automation, etc is to model and analyze the evolution of temporal sequences. Bayesian Networks are established tools for efficiently modelling multivariate data. A DBN is an extention of Bayesian network to temporal domain. In time domain conditional dependencies can be modelled between random variables within as well as across time epochs. The conditional distributions in DBN are assumed to be homogeneous, i.e. the structure and parameters of a network are maintained constant throughout the time. Taking this into account, a DBN is simply constructed by unwrapping a Bayesian network in time domain that causes significant simplification to the model learning procedure. At the same time, this assumption constrains the strenght of DBN in modelling non-stationary sequences, where intrinsic relationships between different variables change in time [37]. These non-stationary sequences are present everywhere in common life, for example the gene interactions in different stages of a life cycle. Obviously, usage of a stationary statistical model is insufficient for modeling gene expression data sequences at all time instances.

Learning a time varying network is not a trivial problem. One may try naively to learn a dynamically changing network independently for each time epoch. However, this is a complex task as there are very little available observations at one time epoch for most applications in real life. One way to overcome the problem of data scarcity is to divide temporal sequences into segments - stationary epochs, with an assumption that in each epoch data are generated from the same probability distribution. However, the lack of knowledge about models in each segment makes the problem more complex. Moreover, the solution space grows exponentially with the length of time sequence [36]. From another point of view, since observations are most often distorted by noise, statistics can be recovered from these perturbations.

To overcome all the difficulties mentioned above, one of the propositions is to expand DBN to nonstationary scenarios by introducing various additional conditions on the type of a network and how the network can change in time. The works done before have been mainly concentrated on nonstationary models with static structure. One of the most popular is the time-varying autoregression model (TVAR) [38]. TVAR is able to describe nonstationary linear dynamic systems, coefficients and noise variances of which continuously change with time. In order to estimate recursively the regression parameters, normalized least squares algorithm can be used. And an error of estimation is shown to be bounded when the model parameters change smoothly [39]. TVAR modelling is widely used in the works related to gene expression data [40]. Extented TVAR models have also been developed for other time-varying processes, e.g. Poisson counting process [41] and non-Gaussian autoregression [42].

One of the tutorials on Bayesian network application to inference the interactions between genes is Nir Friedman *et al.* paper [24]. The authors proposed a framework built on the use of Bayesian networks for describing statistical dependences between variables. The method was applied to the time series gene expression data of the *S. cerevisiae* cell-cycle measurements of Spellman et al. [7]. The proposed approach is quite different from the clustering approach used by [7], [43], [44], [45], in that it attempted to learn a more rich structure from the data, even without use of any prior biological knowledge. However, in learning from time series data [7], the authors treated each measurement as an independent sample from a distribution, and did not assume the temporal aspect of the measurement. Thus, the complex network structure inferred from the experimental data and causal relationships, interactions between genes were “frozen”, not varying in time.

2.2 Observation of Time Varying Networks

J. Khan *et al.* [30] for the inference of time-varying genetic networks from a limited number of noisy observations derived the LASSO-Kalman smoother, which recursively computes the minimum mean-square sparse estimate of the network connectivity at each time point. To overcome

the limited number of observations with respect to the size of domain (small n large p problem) authors performed target tracking in a compressed domain. Chopping the time-series sequence with gene expression values (experimental data were taken from [8]) into homogeneous subsequences they estimated 21 dynamic gene networks, one per 3 time points, during the life cycle of *Drosophila Melanogaster*.

Another application of LASSO-based least squares regression operation for description of the regulatory network together with a particle filter-based state estimation algorithm, which was used to capture the nonlinearity of the system, was proposed in an article of Noor *et al.* [26]. The parameters which characterize the regulatory relations between different genes were estimated online using Kalman filter. The parameter vector was expected to be sparse "since a particular gene interacts with a few other genes only, and as such, many of the system parameters modelling "weak" relationships are irrelevant" [26]. The authors used both synthetic and real biological data of *Drosophila Melanogaster* for learning steady state gene regulatory networks.

W.C. Young *et al.* [27] introduced a Bayesian Model Averaging method (ScanBMA) which is able to infer gene regulatory networks using time series data. Time series generated data from the DREAM competition [46] as well as experimental time series *S. cerevisiae* data were used for estimation of static network structures. Authors state that their method allows inference for networks of thousands of genes to be completed much faster respect to other competing methods.

3 MODEL

A network is assumed to be consisting of I genes. Genes are represented by nodes in the studied network, whereas the relationships among interacting genes are modelled by edges which connect the related nodes. Our aim is to model how the gene expression values evolve over time. We propose a multivariate linear regression model relating the expression value of each gene at a given time to the gene expression values of the previous time instant. Hence, the observation equation describes the gene expression values at a particular time epoch t:

$$x_{i,t} = a_{i1,t} \cdot x_{1,t-1} + a_{i2,t} \cdot x_{2,t-1} + \cdots + a_{iI,t} \cdot x_{I,t-1} + \eta_t, \quad i = 1, \dots, I; \qquad I - total\ number\ of\ genes, \tag{1}$$

where $x_{i,t}$ denote set of observations for all genes in a network for each time epoch t; $a_{ij,t}$ – coefficients of regression equation which model regulatory relations between gene i and gene j in adjacent time epochs.

Unlike classical regression, this multivariate regression is a time-varying regression. The regression coefficients $a_{ij,t}$ are not constant and can be changing at each instant.

We also propose a parametric model to model the changes in the process coefficients. For simplicity, we assume a linear model for the time being, but it can be extended to nonlinear models:

$$a_{ij,t} = a_{ij,t-1} + v_t. \tag{2}$$

More explicitly, equation (1) can be written in a vector form as:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \dots \\ x_{I,t} \end{bmatrix} = \begin{bmatrix} a_{11,t} & a_{12,t} & \cdots & a_{1I,t} \\ a_{21,t} & a_{22,t} & \cdots & a_{2I,t} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1,t} & a_{I2,t} & \cdots & a_{II,t} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \dots \\ x_{I,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \dots \\ \eta_{I,t} \end{bmatrix} \tag{3}$$

In both equations (1) and (2) noise terms η_t and v_t can be of any distribution which can be decided depending on the nature of the data. In the simplest case, they can be assumed to be *i.i.d.* Gaussian such that $\eta_t^{(n)} \sim \mathcal{N}(0, \sigma_\eta^2)$ and $v_t^{(n)} \sim \mathcal{N}(0, \sigma_v^2)$. However, data with outliers or impulsive noise might require a model with heavy-tailed distributions. In such cases, models such as Cauchy or alpha-stable [47] distribution are preferable as alternatives to Gaussian distribution over other heavy-tailed distributions due to satisfying a generalized version of the central limit theorem.

This is clearly a linear regression model and in the context of Bayesian networks reminds one of how partial correlations are calculated to estimate dependencies between nodes [48]. Although, similar to the partial correlations, one tries to recover dependencies all on variables, there is a fundamental difference in this case, where the regression is made over gene expressions of the previous time instant and we are trying to estimate a network which is not describing a given instant but how the gene expressions evolve overtime.

Hence, such a model (1) describes an inter-slice network structure and shows which relations exist between all genes in the network at the previous time epoch, $\{x_{1,t-1}, x_{2,t-1}, \dots, x_{I,t-1}\}$, and the gene under consideration at current time, $\{x_{i,t}\}$. How a certain gene at the current time epoch is influenced by the expression level of other genes at the previous time epoch. The coefficients $a_{ij,t}$ approximate linearly the conditional dependence of the expression of the ith gene at time t on the expression of the jth gene at time t-1. That is, $a_{ij,t} \approx f(x_{i,t}|x_{j,t-1})$. An exact expression would involve also nonlinear dependence terms and dependences on further past.

4 PARTICLE FILTER OR SEQUENTIAL MONTE CARLO

To construct the gene interaction network, we need to devise an iterative algorithm to solve for $a_{ij,t}$. This is a classical problem in signal processing: to recover hidden (possibly time varying) model parameters from observations. Classical approaches to solve the problem include *Kalman filtering* which is a generalization of Wiener filtering, that is finding the optimal linear to recover original signal from noisy observations, to non-stationary data. We would like to address the problem in a Bayesian setup which would allow us also to utilize any prior information we might have and also let us reason in terms of probability distributions rather than single point

estimates. To solve Eqs. (1) and (2) and estimate unknown time varying network parameters $a_{ij,t}$, we propose to utilize an influential *Sequential Monte Carlo* method namely Particle Filtering.

Particle Filtering method was applied recently to gene regulatory network inference by Noor et. al. [26], who used this method for estimation of the states recursively using nonlinear state equations for the model; however, in their formulation the hidden states to be estimated are the gene expression values. The parameters $a_{ij,t}$ characterizing the regulatory relationships between genes were estimated applying Kalman filter on gene expression values. Then they go to reduce solution space with LASSO-regularized least squares estimation to obtain the steady state network structure. Whereas in our paper state equation describes the evolution of the parameters which characterize the regulatory relations among genes in a network and we used Particle Filtering method for direct inference of these parameters at each time epoch.

For many applications, it is very important to know the observation order of the data [49], [50]. In case of the nonstationary ones, the probability density function (pdf) of the related parameter is obviously changing with time. Thus, in nonstationary cases, the expectations and the pdf should be sequentially updated as the new data become available. With this aim, the dynamic systems usually can be described in terms of state-space equations:

$$\boldsymbol{a}_t = f_t(\boldsymbol{a}_{t-1}, \nu_t), \tag{4}$$

$$\boldsymbol{x}_t = h_t(\boldsymbol{a}_t, \eta_t), \tag{5}$$

where $\boldsymbol{a}_t$ and $\boldsymbol{x}_t$ are related to the hidden variables and the observation vectors at actual time t, respectively. In this paper, we assume a linear regression model; however, we would like to note that Particle Filtering method can also accurately model any kinds of nonlinearities present in the state and observation equations. ν_t and η_t are the process and observation noise terms, respectively, which are modelled as Gaussian processes in this paper and their statistics are assumed to be known. The aim is to estimate sequentially the posterior distribution of unknown variables, the hidden state parameters $\boldsymbol{a}_t$, as a set of observations become available online, i.e. $p(\boldsymbol{a}_{0:t}|\boldsymbol{x}_{1:t})$, where $\boldsymbol{a}_{0:t} = \{\boldsymbol{a}_0, \boldsymbol{a}_1, \dots, \boldsymbol{a}_t\}$ and $\boldsymbol{x}_{1:t} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t\}$.

The posterior distribution of parameters $\boldsymbol{a_t}$ is estimated at each time instant in two steps: first is prediction and second - update. In the first step, the value of the hidden parameter, $\boldsymbol{a_t}$, at time t is predicted from the previous time instant *t-1* according to the first order Markov process (4) $p(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})$ [51]. At time instant t, the predicted parameter is updated via Bayes rule as an observation $\boldsymbol{x}_t$ becomes available as follows:

$$p(\boldsymbol{a}_t|\boldsymbol{x}_{1:t}) = \frac{p(\boldsymbol{x}_t|\boldsymbol{a}_t)p(\boldsymbol{a}_t|\boldsymbol{x}_{1:t-1})}{p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1})}, \tag{6}$$

where assuming a Gaussian observation noise, $p(\boldsymbol{x}_t|\boldsymbol{a}_t)$is given by

$$p(x_t|a_t) = \frac{1}{\left(2\pi\sigma_\eta^2\right)^{1/2}} \exp\left(-\frac{(x_t^{obs} - x_t)^2}{2\sigma_\eta^2}\right). \tag{7}$$

The relation (6) gives the optimal Bayesian solution. In general this recursive solution cannot be determined analytically [49]. Thus, the idea of sequential Monte Carlo method is to form a finite set of N weighted state samples, called particles, which approximates the posterior distribution [49]:

$$p(\boldsymbol{a}_t|\boldsymbol{x}_{1:t}) \approx \sum_{n=1}^{N} w_t^{(n)} \delta\left(\boldsymbol{a}_t - \boldsymbol{a}_t^{(n)}\right), \tag{8}$$

where δ denotes the delta-Dirac function. When number of particles, N, goes to infinity, the approximation approaches the exact solution, true distribution.

Briefly, at each time epoch t the filtering starts with the sample set $\{\boldsymbol{a}_{t-1}^{(n)}, w_{t-1}^{(n)}\}_{n=1\dots N}$ related to the previous time epoch. New samples for time t are drawn from a proposal distribution:

$$\boldsymbol{a}_t^{(n)} \sim q\left(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}^{(n)}, \boldsymbol{x}_t\right). \tag{9}$$

Sampling from $q\left(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}^{(n)}, \boldsymbol{x}_t\right)$ to approximate posterior $p(\boldsymbol{a}_{0:t}|\boldsymbol{x}_{1:t})$ as shown above, is known as the "sequential importance sampling" method [49].

When new samples $\boldsymbol{a}_t^{(n)}$ are obtained, we estimate the likelihood of each of the particle using the observation model. After that their weights are updated by

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p\left(\boldsymbol{x}_t|\boldsymbol{a}_t^{(n)}\right)p\left(\boldsymbol{a}_t^{(n)}|\boldsymbol{a}_{t-1}^{(n)}\right)}{q\left(\boldsymbol{a}_t^{(n)}|\boldsymbol{a}_{t-1}^{(n)}, \boldsymbol{x}_t\right)}. \tag{10}$$

Choosing a good proposal density $q(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}, \boldsymbol{x}_t)$ is crucial for the efficiency of particle filter. State transition distribution $p(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})$ is a convenient choice for proposal distribution and reduces weight update in Eq. (10) into trivial accumulation of likelihoods as shown below:

$$w_t^{(n)} \propto w_{t-1}^{(n)} \cdot p\left(\boldsymbol{x}_t|\boldsymbol{a}_t^{(n)}\right). \tag{11}$$

After this step, the importance weights are normalized to form a probability distribution [49]:

$$w_t^{(n)} = \frac{w_t^{(n)}}{\sum_{n=1}^{N} w_t^{(n)}}. \tag{12}$$

The resultant sample set $\{\boldsymbol{a}_t^{(n)}, w_t^{(n)}\}_{n=1,\dots N}$ can give estimations for the posterior distribution of current data and information about network structure.

However, a peculiar problem known as "degeneracy" rises when one uses the sequential Monte-Carlo

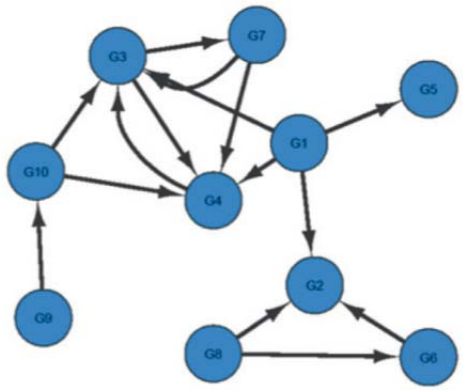

Fig. 1. A regulatory network of *S. cerevisiae* used for synthetic data generation in this work (reproduced from [27]).

algorithm. It occurs when the dimension of the state space is high: after few iterations most of the samples drawn from the state transition distribution may have very small weights. In such situation sampling efficiency should be increased by performing resampling [49], [50]. The number of effective particles N_{eff} is a measure of the degeneracy and can be calculated by

$$N_{eff} = \frac{1}{w_t^{(n)} \cdot w_t'^{(n)}}. \quad (13)$$

When N_{eff} is below a certain threshold value, the resampling should be performed: samples with low importance weights are eliminated, and those with high importance weights are allocated to obtain n equally weighted samples of $\boldsymbol{a}_t^{(n)}$.

5 Results and Discussion

5.1 Synthetic Data

The proposed algorithm was tested on time-series simulated data from the DREAM competition [46]. Data were generated from known transcriptional regulatory networks of *S. cerevisiae* using GeneNetWeaver 3.1.1 Beta (GNW) software [52]. Its dynamics was simulated using a detailed kinetic model of gene regulation. Both independent and synergistic gene regulation occur in the networks. Both transcription and translation are modelled. Simulated data correspond to noisy measurements of mRNA levels, which have been normalized. To model internal noise in the dynamics of the networks the simulations are based on stochastic differential equations (Langevin equations). In addition, measurement noise was added to the generated gene expression datasets. Existing model of noise observed in microarrays, which is very similar to a mix of normal and lognormal noise, was also used [46], [53].

Each time series has 201 time points. The network structure used for synthetic simuation is shown in Fig. 1.

Fig. 2 shows generated time series for each gene (201 time points) and how expression levels respond to a perturbation. Perturbation was simulated by slightly increasing or decreasing the basal activation of all genes of the network simultaneously by different random amounts [46]. It was applied to the data at t=0 and removed at t=100. The first half of the time series (until t=100) shows the response of the network to the perturbation. The second half of the time series (until t=201) shows how the gene expression levels go back from the perturbed to the steady-state levels.

As discussed before, the idea is to dynamically estimate unknown regulatory coefficients $\boldsymbol{a}_{ij,t}$ using particle filtering method. Careful estimation of $\boldsymbol{a}_{ij,t}$ allows us to highlight the main gene interactions in the network. The number of particles used is N = 300. The variance of the noise in the state update equation (1), v_t, is taken to be 10^{-2} and the variance of measurement noise is assumed to be known and constant, $\sigma_\eta^2 = 10^{-3}$.

It is important to specify a certain threshold in order to estimate the strongest relations between genes. Threshold is estimated as dispersion from the average of $\boldsymbol{a}_{ij,t}$ (mean) at particular time epoch, i.e. at the final time epoch both mean and standard deviation (std) of estimated $\boldsymbol{a}_{ij,T}$ are calculated. Thus, values of $\boldsymbol{a}_{ij,T}$ above a threshold (mean + std) indicate that gene $\boldsymbol{i}$ activates gene $\boldsymbol{j}$, and values below the threshold (mean - std) are considered as repression activity of gene $\boldsymbol{j}$ in relation to gene $\boldsymbol{i}$. The parameters $\boldsymbol{a}_{ij,T}$ whose values are between two limit thresholds indicate an absence of relations or weak

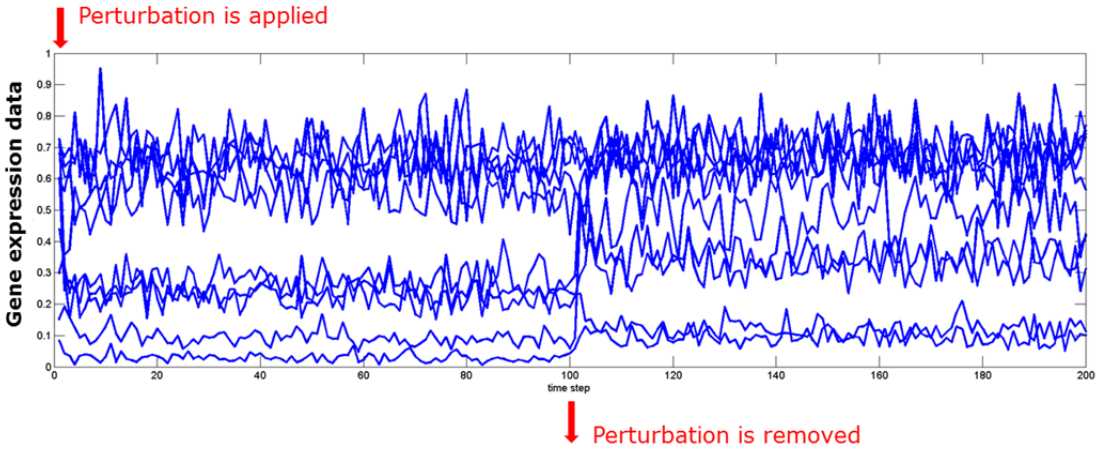

Fig. 2. Time evolution of gene expression data for 10 genes undergone a perturbation at t<100.

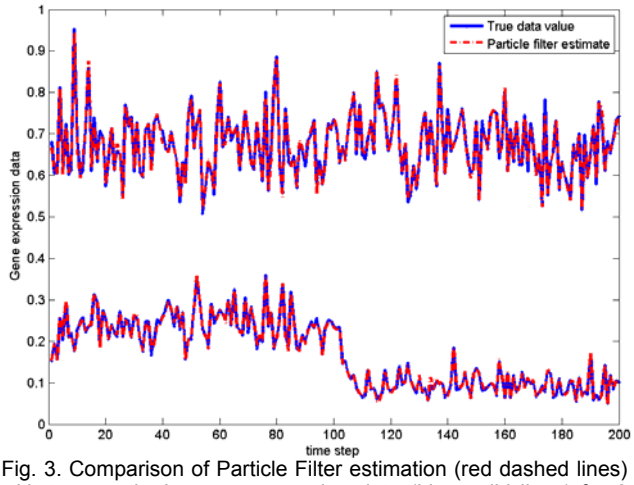

Fig. 3. Comparison of Particle Filter estimation (red dashed lines) with true synthetic gene expression data (blue solid lines) for 2 genes.

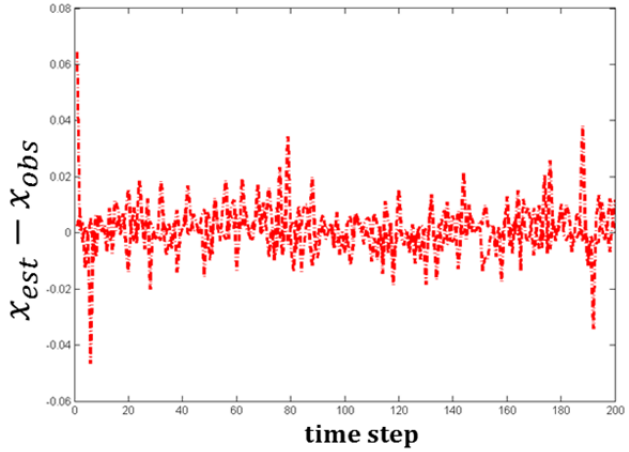

Fig. 4. Residuals between synthetic gene expression data, $\boldsymbol{x_{obs}}$, and estimated by Particle Filter, $\boldsymbol{x_{est}}$, for two genes.

relations among the genes, and they are not considered in further analysis.

The algorithm used for inferring the network parameters from Eqs. (1) and (2) is displayed below.

5.2 Algorithm. Time-varying Gene Network Learning

1. Input gene expression time series data $\boldsymbol{x}_{i,1:t}^{obs}$.
2. Initialization of all parameters.
3. **for** n=1,…,N do
4. Generating particles for $\boldsymbol{a}_{ij,t=0}^{(n)}$ from $p(\boldsymbol{a}_{ij,t=0}|\boldsymbol{x}_{t=0})$.
5. Assigning equal weights to the particles.
6. **end for**
7. **for** t=1,…,T
8. **for** n=1,…,N
9. Prediction of $\boldsymbol{a}_{ij,t}^{(n)}$ from $p(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}^{(n)}, \boldsymbol{x}_t)$.
10. Prediction of $\boldsymbol{x}_{i,t}$ using (10), $\boldsymbol{a}_{ij,t}^{(n)}$ and observations of $\boldsymbol{x}_{i,t-1}^{obs}$.
11. Update the weights of each particle using (11).
12. **end for**
13. Normalization of the weights, (12).
14. Resampling, if necessary, (13).
15. Set of $\{\boldsymbol{a}_{ij,t}^{(n)}, w_t^{(n)}\}_{n=1...N}$ is used for the next time epoch, step 8.
16. **end for**.

5.3 Results of the Modeling

Results for gene expression estimation by Particle filter are shown in Fig. 3 for a two genes network. It is clearly observed that the proposed model, Eqs. (1) and (2), makes a good enough prediction of the gene expression data. Moreover, the use of particle filter allows us to follow with high accuracy all changes in gene expression data undergone due to different external perturbations. Residuals of the estimated gene expression values with respect to the synthetic ones are shown in Fig. 4.

As it was discussed above the idea is not only to predict and follow expression levels changes in a gene network, but also to estimate the posterior distributions of $\boldsymbol{a}_{ij,t}$ coefficients which model the relations between genes in time. The relations in a network consisting of 2 genes are described by 4 coefficients - $\{\boldsymbol{a}_{11}, \boldsymbol{a}_{12}, \boldsymbol{a}_{21}, \boldsymbol{a}_{22}\}$. Coefficients $\boldsymbol{a}_{11}$ and $\boldsymbol{a}_{22}$, autocorrelation terms, are related to auto-regulation of a gene in adjacent time epochs; $\boldsymbol{a}_{12}$ and $\boldsymbol{a}_{21}$ – cross-correlation terms, which show how expression level of one gene at the time epoch (t-1) effects on expression level of other gene at time t.

Since the results obtained are stochastic and can fluctuate slightly from realisation to realization we launched the algorithm for a number of runs and calculated an average values of coefficients $\boldsymbol{a}_{ij,T}$.

Histograms for the distributions of the estimated network coefficients $\boldsymbol{a}_{ij,T}$ at the final time epoch (t=201) are shown in Fig. 5. After running the algorithm 100 times the average values of $\boldsymbol{a}_{ij,T}$ at the final time epoch were calculated. Running the algorithm even 1000 times does not change significantly the final estimates: histograms become more refined, meanwhile the average values and variance remain almost the same (see Fig. 6 and Table 1).

TABLE 1
AVERAGE VALUES OF $\boldsymbol{a}_{ij,T}$ COEFFICIENTS AND THEIR VARIANCE ESTIMATED FOR N RUNS

N of runs	$\boldsymbol{a}_{11}$	$\boldsymbol{a}_{12}$	$\boldsymbol{a}_{21}$	$\boldsymbol{a}_{22}$	Standard Deviation
100	0.620	1.629	0.152	-0.334	0.837
1000	0.632	1.538	0.152	-0.333	0.798

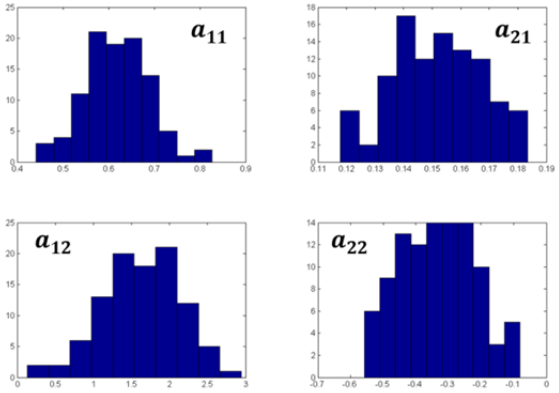

Fig. 5. Histograms for distributions of the inferred $\boldsymbol{a}_{ij,T}$ coefficients for two genes network at the final time epoch averaged using 100 runs. Vertical axis – number of particles with certain value, horizontal axis – values of $\boldsymbol{a}_{ij,T}$.

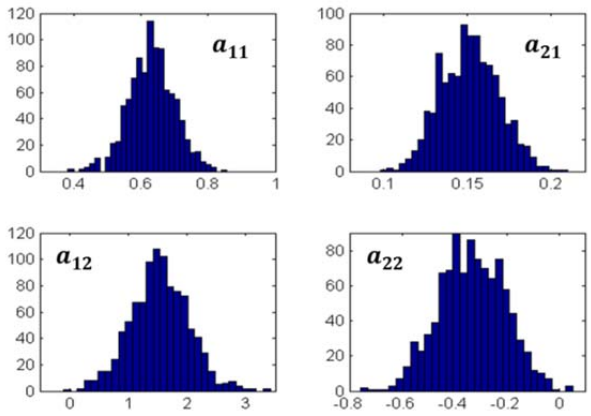

Fig. 6. Histograms for distributions of the inferred $\boldsymbol{a}_{ij,T}$ coefficients for two genes network at the final time epoch averaged using 1000 runs. Vertical axis – number of particles with certain value, horizontal axis – values of $\boldsymbol{a}_{ij,T}$.

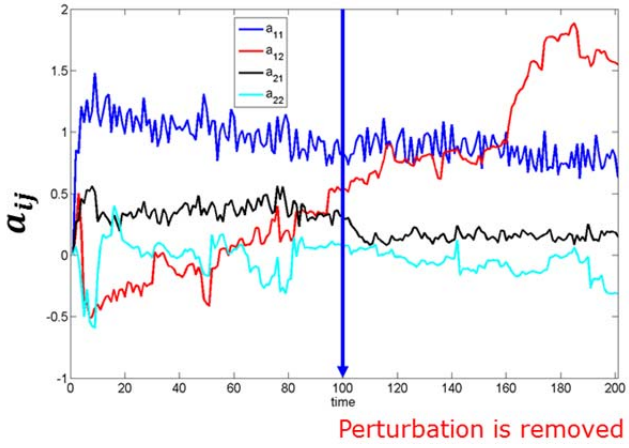

Fig. 7. Time dependence of coefficients $\boldsymbol{a}_{ij,t}$ averaged using 100 runs for 2 genes.

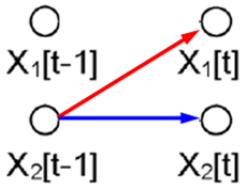

Fig. 8. Possible time varying network structure with two genes. Snapshot done for t=201. Red arrow indicates expression activity at adjacent time epochs and blue arrow shows repression activity.

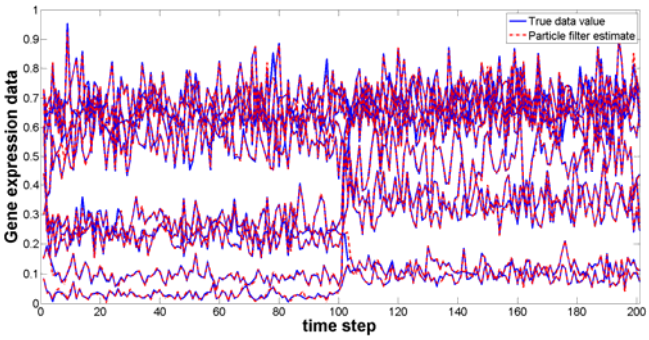

Fig. 9. Comparison of Particle Filter estimation (red dashed lines) with synthetic gene expression data (blue solid lines) for 10 genes undergone a perturbation at t<100.

Time dependence of the estimated coefficients $\boldsymbol{a}_{ij,t}$ is shown in Fig. 7. It can be seen that at the beginning, inferred values are scattered and after some time (t≈10) they become more stable smoothly changing with time with some insignificant fluctuations. However the time dependence of the coefficient $\boldsymbol{a}_{12,t}$ constitutes an exception: it monotonically increases during the whole learning time period changing its values from negative to positive ones. It could mean that the bonding between those two genes becomes stronger with time.

After successful gene expression data tracking and estimation of coefficients $\boldsymbol{a}_{ij,t}$, we would like to understand how they reflect the relations between genes. Unfortunately, there is still no certain interpretation for the coefficients which model the relationships in gene

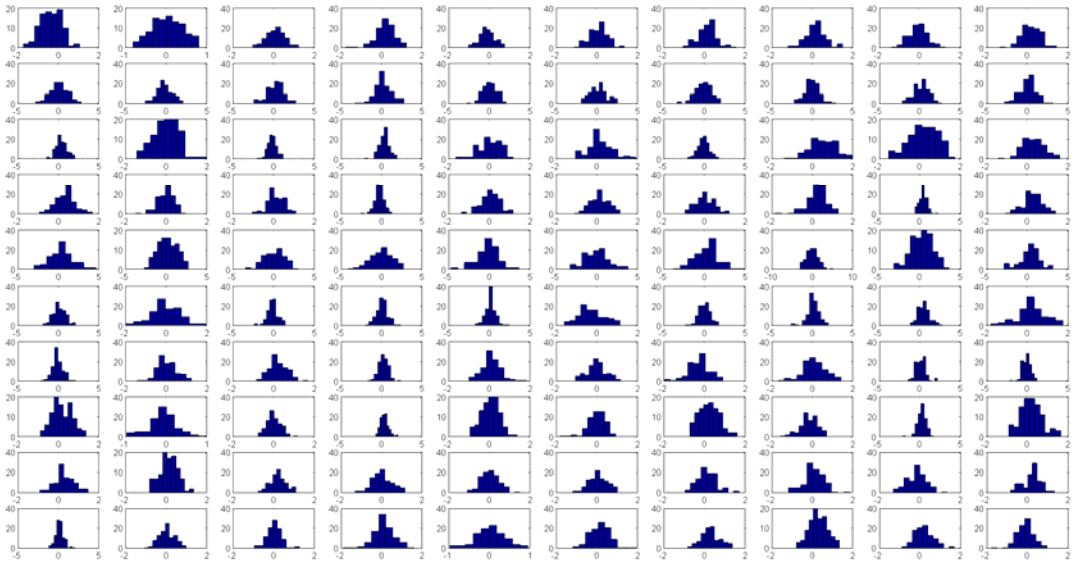

Fig. 10. Histograms for distributions of the inferred $\boldsymbol{a}_{ij,T}$ coefficients for 10 genes network at the final time epoch averaged using 100 runs. Vertical axis – number of particles with certain value, horizontal axis – values of $\boldsymbol{a}_{ij,T}$.

TABLE 2
MEAN VALUES OF $\boldsymbol{a}_{ij,T}$ COEFFICIENTS FOR TEN GENES NETWORK AT THE FINAL TIME EPOCH AVERAGED USING 100 RUNS TOGETHER WITH STANDARD DEVIATION

Genes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	-0.369	0.104	0.363	0.330	0.439	0.087	-0.068	0.181	0.290	0.178
2	0.027	-0.159	0.006	-0.014	0.140	-0.052	0.030	-0.076	0.129	0.035
3	0.092	0.188	-0.188	0.092	0.047	-0.157	0.213	0.007	0.204	0.018
4	0.182	0.208	0.354	-0.267	-0.039	0.072	0.227	0.357	0.037	0.115
5	-0.113	0.043	0.012	0.054	-0.164	0.134	0.098	0.045	0.055	0.025
6	0.075	0.183	0.163	0.119	-0.030	-0.267	-0.018	-0.026	0.057	0.145
7	0.126	-0.164	-0.118	0.024	0.332	-0.022	-0.274	0.227	0.115	0.287
8	0.174	-0.030	0.503	0.235	0.030	0.143	0.096	-0.170	0.088	0.298
9	-0.095	0.176	0.167	0.373	0.210	0.440	0.075	0.233	-0.180	0.189
10	0.172	0.068	0.277	0.260	0.257	0.192	-0.127	0.184	0.215	-0.128
									Standard Deviation	0.169

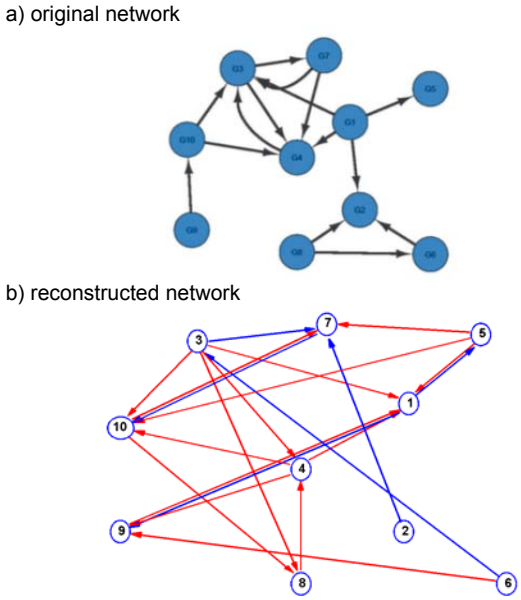

Fig. 11. Comparison of original "static" with a snapshot of time-varying network structure with 10 genes. The network in b) is reconstructed for t=201 after 100 runs. Red and blue arrows indicate expression and repression activities at adjacent time epochs, respectively; red, blue and black colors of nodes indicate self-expression, self-repression and no expression of a gene in adjacent time epochs (t=200 and t=201), respectively.

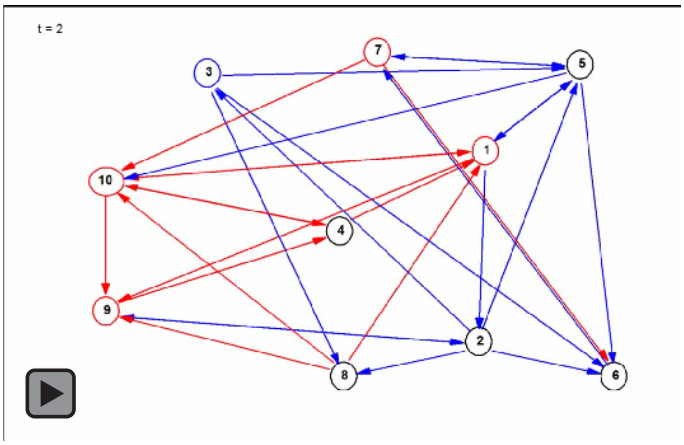

Fig. 12. Possible time-varying network structure with 10 genes. Red and blue arrows indicate expression and repression activities at adjacent time epochs, respectively; red, blue and black colors of nodes indicate self-expression, self-repression and no expression of a gene in adjacent time epochs, respectively.

regulatory networks. Some authors [26] suggest that posi-positive value of $\boldsymbol{a}_{ij,t}$ above certain threshold indicates that gene i activates gene j, whereas negative values show repression activity. However, these authors use different models and methods from ours for inferring gene networks. The statistical interpretation we give is that they approximate the conditional probabilities: $a_{ij,t} \approx f(x_{i,t}|x_{j,t-1})$.

In our case for two-gene network, average coefficients at the last time epoch (t=201) are $\{\boldsymbol{a}_{11} = 0.63; \boldsymbol{a}_{12} = 1.56; \boldsymbol{a}_{21} = 0.15; \boldsymbol{a}_{22} = -0.32\}$. If we assume that the threshold would be the mean value of all four coefficients (≈0.5) plus/minus standard deviation (≈0.8) we get upper threshold value≈1.3 and lower≈–0.3, and take into account the values above and below that threshold, only $\boldsymbol{a}_{12}$ and $\boldsymbol{a}_{22}$ will remain. It could suggest that in our two gene network gene N.2 at time epoch t=200 expresses gene N.1 at t=201, and gene N.2 at time epoch t=200 represses itself at t=201. Graphically it can be shown in Fig. 8.

Extending our analysis to a gene network consisting of 10 genes as it was aimed above, we can again observe very good results of PF data tracking: excellent agreement between synthetic gene expression data and PF estimates (see Fig. 9).

The time dependence of learned coefficients $\boldsymbol{a}_{ij,t}$ for 10 genes network is not shown here due to overlapping of 100 coefficients on each other that does not give clear picture of what is happening with time. However, histograms for the distributions of the estimated network coefficients $\boldsymbol{a}_{ij,T}$ at the final time epoch and their averaged values for 100 runs of the algorithm are shown in Fig. 10 and Table 2, respectively, together with the standard deviation. Possible gene network results of averaging of $\boldsymbol{a}_{ij,T}$ for the final time epoch are shown in Fig. 11. An animated tyme-varying network structure is

displayed in Fig. 12. Fig. 11 demonstrates the comparison of original network with reconstructed one using 100 runs for averaging. However, care must be taken while comparing the two networks since the original network shows static structure for the whole time period while the second inferred network indicates a dynamic case: relations between genes at two adjacent time epochs. Red arrows indicate the expression activities of a gene where an arrow starts at time epoch t=200 in relation to another gene where an arrow points at time epoch t=201. On the other hand blue arrows show the repression activities of a gene where an arrow starts at time epoch t=200 in relation to another gene where an arrow points at time epoch t=201. Red, blue and black colors of the nodes indicate self-expression, self-repression and no expression of a gene in adjacent time epochs (t=200 and t=201), respectively.
Both networks in Figure 11 a) and b) are compatible except changes of several connections from/to genes N.2, N.5, N.7, and N.9.

6 Conclusion

Application of particle filtering method to synthetic gene expression time series dataset for a network of genes helped us to follow with high accuracy the changes in gene expression data undergone due to different external perturbations. Moreover, gene expression temporal sequence data were utilized for online learning of time varying gene network structure. The proposed model is capable of discovering causal relationships, interactions between genes that vary in time.

Acknowledgment

This work was funded by the CNR Flagship project InterOmics. Ercan E. Kuruoglu gratefully acknowledges partial support from Alexander von Humboldt Foundation in the form of an Experienced Research Fellowship which made his stay in the Max Planck Institute possible.

References

[1] E. Segal, H. Wang, and D. Koller, "Discovering molecular pathways from protein interaction and gene expression data," *Bioinformatics*, vol. 19, no. 1, pp. i264-i272, 2003.

[2] H. Ge, Z. Liu, G.M. Church, and M. Vidal, "Correlation between transcriptome and interactome mapping data from saccharomyces cerevisiae," *Nature Genetics*, vol. 29, no. 4, pp. 482-486, 2001.

[3] R. Jansen, D. Greenbaum, and M. Gerstein, "Relating Whole-Genome Expression Data with Protein-Protein Interactions," *Genome Research*, vol. 12, no. 1, pp. 37-46, 2002.

[4] P.C. Phillips, S.P. Otto, and M.C. Whitlock, "Beyond the average. The evolutionary importance of gene interactions and variability of epistatic effects," in *Epistasis and the Evolutionary Process*, J. B. Wolf, E. D. Brodie III, and M. J. Wade, eds. Oxford University Press, pp. 20-38, 2000.

[5] J.L. Hartman IV, B. Garvik, and L. Hartwell, "Cell biology: Principles for the buffering of genetic variation," *Science*, vol. 291, no. 5506, pp. 1001-1004, 2001.

[6] C. Boone, H. Bussey, and B.J. Andrews, "Exploring genetic interactions and networks with yeast," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 437-449, 2007.

[7] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher, "Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," *Molecular Biology of the Cell*, vol. 9, no. 12, pp. 3273-3297, 1998.

[8] M.N. Arbeitman, E.E.M. Furlong, F. Imam, E. Johnson, B.H. Null, B.S. Baker, M.A. Krasnow, M.P. Scott, R.W. Davis, and K.P. White, "Gene expression during the life cycle of drosophila melanogaster," *Science*, vol. 297, no. 5590, pp. 2270-2275, 2002.

[9] V.R. Iyer, M.B. Eisen, D.T. Ross, G. Schuler, T. Moore, J.C.F. Lee, J.M. Trent, L.M. Staudt, J. Hudson Jr., M.S. Boguski, D. Lashkari, D. Shalon, D. Botstein, and P.O. Brown, "The transcriptional program in the response of human fibroblasts to serum," *Science*, vol. 283, no. 5398, pp. 83-87, 1999.

[10] M.E. Hughes, L. DiTacchio, K.R. Hayes, C. Vollmers, S. Pulivarthy, J.E. Baggs, S. Panda, and J.B. Hogenesch, "Harmonics of circadian gene transcription in mammals," *PLoS Genetics*, vol. 5, no. 4, p. e1000442, 2009.

[11] A. Korenčič, G. Bordyugov, R. Košir, D. Rozman, M. Goličnik, and H. Herzel, "The interplay of cis-regulatory elements rules circadian rhythms in mouse liver," *PLoS ONE*, vol. 7, no. 11, p. e46835, 2012.

[12] S. Wiltshire, J.T. Bell, C.J. Groves, C. Dina, A.T. Hattersley, T.M. Frayling, M. Walker, G.A. Hitman, M. Vaxillaire, M. Farrall, P. Froguel, and M.I. Mccarthy, "Epistasis between type 2 diabetes susceptibility loci on chromosomes 1q21-25 and 10q23-26 in northern Europeans," *Annals of Human Genetics*, vol. 70, no. 6, pp. 726-737, 2006.

[13] O. Combarros, M. Cortina-Borja, A.D. Smith, and D.J. Lehmann, "Epistasis in sporadic Alzheimer's disease," *Neurobiology of Aging*, vol. 30, no. 9, pp. 1333-1349, 2009.

[14] D.M. Evans, C.C.A. Spencer, J.J. Pointon, Z. Su, D. Harvey, G. Kochan, U. Opperman, A. Dilthey, M. Pirinen, M.A. Stone, L. Appleton, L. Moutsianis, S. Leslie, T. Wordsworth, T.J. Kenna, T. Karaderi, G.P. Thomas, M.M. Ward, M.H. Weisman, C. Farrar, L.A. Bradbury, P. Danoy, R.D. Inman, W. Maksymowych, D. Gladman, P. Rahman, A. Morgan, H. Marzo-Ortega, P. Bowness, K. Gaffney, J.S.H. Gaston, M. Smith, J. Bruges-Armas, A. Couto, R. Sorrentino, F. Paladini, M.A. Ferreira, H. Xu, Y. Liu, L. Jiang, C. Lopez-Larrea, R. Díaz-Peña, A. Lóepez-Vázquez, T. Zayats, G. Band, C. Bellenguez, H. Blackburn, J.M. Blackwell, E. Bramon, S.J. Bumpstead, J.P. Casas, A. Corvin, N. Craddock, P. Deloukas, S. Dronov, A. Duncanson, S. Edkins, C. Freeman, M. Gillman, E. Gray, R. Gwilliam, N. Hammond, S.E. Hunt, J. Jankowski, A. Jayakumar, C. Langford, J. Liddle, H.S. Markus, C.G. Mathew, O.T. McCann, M.I. McCarthy, C.N.A. Palmer, L. Peltonen, R. Plomin, S.C. Potter, A. Rautanen, R. Ravindrarajah, M. Ricketts, N. Samani, S.J. Sawcer, A. Strange, R.C. Trembath, A.C. Viswanathan, M. Waller, P. Weston, P. Whittaker, S. Widaa, N.W. Wood, G. McVean, J.D. Reveille, B.P. Wordsworth, M.A. Brown, and P. Donnelly, "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility," *Nature Genetics*, vol. 43, no. 8, pp. 761-767, 2011.

[15] J.L. DeRisi, V.R. Iyer, and P.O. Brown, "Exploring the metabolic and genetic control of gene expression on a genomic scale," *Science*, vol. 278, no. 5338, pp. 680-686, 1997.

[16] D.J. Lockhart, H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, and E.L. Brown, "Expression monitoring by hybridization to high-density oligonucleotide arrays," *Nature Biotechnology*, vol. 14, no. 13, pp. 1675-1680, 1996.

[17] X. Wen, S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker, and

R. Somogyi, "Large-scale temporal gene expression mapping of central nervous system development," *Proc. National Academy of Science USA*, pp. 334-339, 1998.

[18] L. Li, B.A. Edgar, and S.S. Grewal, "Nutritional control of gene expression in Drosophila larvae via TOR, Myc and a novel cis-regulatory element," *BMC Cell Biology*, vol. 11, no. 7, pp. 1-13, 2010.

[19] J. Liang, L. Zhang, Z. Xiang, and N. He, "Expression profile of cuticular genes of silkworm: Bombyx mori," *BMC Genomics*, vol. 11, no. 173, pp. 1-13, 2010.

[20] F. Pilot, J.-M. Philippe, C. Lemmers, J.-P. Chauvin, and T. Lecuit, "Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of Drosophila cellularisation," *Development*, vol. 133, pp. 711-723, 2006.

[21] T. Pramila, W. Wu, S. Miles, W.S. Noble, and L.L. Breeden, "The forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle," *Genes and Development*, vol. 20, no. 16, pp. 2266-2278, 2006.

[22] D.F. Simola, C. Francis, P.D. Sniegowski, and J. Kim, "Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes," *Genome Biology*, vol. 11, no. 10, pp. R105-R121, 2010.

[23] S. Wichert, K. Fonkianos, and K. Strimmer, "Identifying periodically expressed trascripts in microarry time series data," *Bioinformatics*, vol. 20, no. 1, pp. 5-20, 2004.

[24] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian Network to Analyze Expression Data," *J. Computational Biology*, vol. 7, pp. 7,601-620, 2000.

[25] S.L. Lauritzen and N. Wermuth, "Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative," *The Annals of Statistics*, vol. 17, no. 1, pp. 31-57, 1989.

[26] A. Noor, E. Serpedin, M. Nounou, and H. Nounou, "Inferring gene regulatory networks via nonlinear state-space models and exploiting sparsity," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 4, pp. 1203-1211, 2012.

[27] W.C. Young, A.E. Raftery, and K.Y. Yeung, "Fast Bayesian inference for gene regulatory networks using ScanBMA," *BMC Systems Biology*, vol. 8, no. 1, pp. 47-57, 2014.

[28] D.A.K. Maduranga, J. Zheng, P.A. Mundra, and J.C. Rajapakse, "Inferring gene regulatory networks from time-series expressions using random forests ensemble," *Proc. of the 8th IAPR International Conference, (PRIB 2013)*, pp. 13-22, 2013.

[29] M. Masnadi-Shirazi, M.R. Maurya, and S. Subramaniam, "Time-varying causal inference from phosphoproteomic measurements in macrophage cells," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 1, pp. 74-86, Feb 2014, doi:10.1109/TBCAS.2013.2288035.

[30] J. Khan, N. Bouaynaya, and H.M. Fathallah-Shaykh, "Tracking of time-varying genomic regulatory networks with a LASSO-Kalman smoother," *EURASIP J. Bioinformatics and Systems Biology*, vol. 2014, no. 1, pp. 3-17, 2014.

[31] K. Murphy and S. Mian, "Modelling Gene Expression Data Using Dynamic Bayesian Networks," Technical Report, Computer Science Division, University of California, Berkeley, 1999.

[32] Y. Zhang, Z. Deng, H. Jiang, and P. Jia, "Inferring gene regulatory networks from multiple data sources via a dynamic Bayesian network with structural em," *Proc. of the 4th international conference on Data integration in the life sciences*, pp. 204-214, 2007.

[33] M. Quach, N. Brunel, and F. D'alché-Buc, "Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference," *Bioinformatics*, vol. 23, no. 23, pp. 3209-3216, 2007.

[34] F.X. Wu, W.J. Zhang, and A.J. Kusalik, "Modeling gene expression from microarray expression data with state-space equations," *Proc. of the Computational Systems Bioinformatics Conference*, pp. 581-592, 2004.

[35] Y. Huang, I.M. Tienda-Luna, and Y. Wang, "Reverse engineering gene regulatory networks, A survey of statistical models," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 76-97, 2009.

[36] Z. Wang, E.E. Kuruoğlu, X. Yang, Y. Xu, and T.S. Huang, "Time varying dynamic bayesian network for nonstationary events modeling and online inference," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1553-1568, Apr 2011, doi:10.1109/TSP.2010.2103071.

[37] K. Murphy, "Dynamic Bayesian networks, Representation, inference and learning," PhD thesis, Computer Science Division, University of California, Berkeley, 2002.

[38] R. Prado, G. Huerta, and M. West, "Bayesian time-varying autoregressions: Theory, methods and applications," *Special issue on Time Series and Related Topics of "Resenhas", J. Institute of Mathematics and Statistics of the University of Sao Paolo*, pp. 405-422, 2000.

[39] E. Moulines, P. Priouret, and F. Roueff, "On recursive estimation for time varying autoregressive processes," *Annals of Statistics*, vol. 33, no. 6, pp. 2610-2654, 2005.

[40] A. Rao, A. Hero III, D.J. States, and J.D. Engel, "Inferring time-varying network topologies from gene expression data," *EURASIP J. Bioinformatics and System Biology - Special Issue on Gene Networks*, p. 51947, 2007.

[41] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," *Proc. of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 207-216, 2006.

[42] D. Gençağa, E.E. Kuruoğlu, and A. Ertüzün, "Modeling non-Gaussian time-varying vector autoregressive processes by particle filtering," *Multidimensional Systems and Signal Processing*, vol. 21, no. 1, pp. 73-85, 2010.

[43] U. Alon, N. Barka, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, and A.J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. National Academy of Science USA*, pp. 6745-6750, 1999.

[44] A. Ben-Dor, R. Shamir, and Z. Yakhini, "Clustering gene expression patterns," *J. Computational Biology*, vol. 6, no. 3-4, pp. 281-297, 1999.

[45] G.S. Michaels, D.B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi, "Cluster analysis and data visualization of large-scale gene expression data," *Pacific Symposium on Biocomputing*, pp. 42-53, 1998.

[46] DREAM4 In Silico Network Challenge, http://wiki.c2b2.columbia.edu/dream/index.php?title=D4c2.

[47] N. Misra and E.E. Kuruoğlu, "Stable graphical models," http://arxiv.org/abs/1404.4351, submitted for publication.

[48] A. De la Fuente, N. Bing, I. Hoeschele, and P. Mendes, "Discovery of meaningful associations in genomic data using partial correlation coefficients," *Bioinformatics*, vol. 18, pp. 3565-3574, 2004.

[49] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197-208, 2000.

[50] P.M. Djurić, J.H. Kotecha, F. Esteve, and E. Perret, "Sequential parameter estimation of time-varying non-gaussian autoregressive processes," *EURASIP J. Applied Signal Processing*, vol. 2002, no. 8, pp. 865-875, 2002.

[51] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, Feb 2002, doi: 10.1109/78.978374.

[52] T. Schaffter, D. Marbach, and D. Floreano, "GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods," *Bioinformatics*, vol. 27, no. 16, pp. 2263-2270, 2011.

[53] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano, "Generating realistic in silico gene networks for performance assessment of reverse engineering methods," *J. Computational Biology*, vol. 16, no. 2, pp. 229-

239, 2009.

Sergiy Ancherbak received the BS and MS degrees in thermal physics from Institute of Refrigeration, cryotechnology and eco-energy, Odessa, Ukraine in 2002 and 2003, respectively, and the PhD degree in applied physics from the University of Pisa, Pisa, Italy in July 2011. He is currently a postdoc at the ISTI-CNR, Pisa, Italy. His research interests include multivariate statistical analysis with applications in bioinformatics.

Ercan E. Kuruoğlu (M'98–SM'06) received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Bilkent University, Turkey, in 1991 and 1993, and the M.Phil. and Ph.D. degrees in information engineering from the University of Cambridge, Cambridge, U.K., in 1995 and 1998, respectively. In 1998, he joined the Xerox Research Center Europe, Cambridge. He was an ERCIM Fellow in 2000 with INRIA-Sophia Antipolis, France. In January 2002, he joined ISTI-CNR, Pisa, Italy. He was a visiting professor with Georgia Tech-Shanghai in autumn terms 2007 and 2011. He is currently a Senior Researcher at ISTI-CNR. He was an Associate Editor for the IEEE Transactions on Signal Processing and IEEE Transactions on Image Processing. He is currently the Editor in Chief of Digital Signal Processing: A Review Journal. He acted as a Technical co-Chair for EUSIPCO 2006 and a Tutorials co-Chair of ICASSP 2014. He was a member of the IEEE Technical Committee on Signal Processing Theory and Methods. He was a Chinese Government 111 Project Foreign Expert 2007-2011. He was an Alexander von Humboldt Experienced Research Fellow in Max Planck Institute for Molecular Genetics in 2013-2014. His research interests are in the areas of statistical signal and image processing and information and coding theory with applications in computational biology, astrophysics and telecommunications. He is a Senior Member of IEEE since 2006.

Martin Vingron is a mathematician by education who has done his PhD in computational biology at EMBL in 1991. At the time and for a number of years of postdoctoral training his research has focused on the analysis of protein sequences, sequence analysis, sequence comparison, and molecular evolution. Methods of discrete optimisation were used for the design of comparison algorithms and probability theory was applied to answer questions of significance of computational results. Later, as a department head at the German Cancer Research Center and from 2000 on as a director at the Max Planck Institute for Molecular Genetics in Berlin, his research focus shifted towards the processing and mathematical analysis of functional genomics data. He has contributed to analysis methods for DNA microarrays and in recent years has been working on transcriptional regulation. Vingron received the Max Planck Research Prize in 2004, is an elected member of the German Academy of Natural Scientists Leopoldina, and also holds a secondary appointment at the CAS-MPG Partner Institute for Computational Biology in Shanghai. He is in the editorial board of Journal of Computational Biology and in the steering committee of RECOMB.

All biographies should be limited to one paragraph consisting of the following: sequentially ordered list of degrees, including years achieved; sequentially ordered places of employ concluding with current employment; association with any official journals or conferences; major professional and/or academic achievements, i.e., best paper awards, research grants, etc.; any publication information (number of papers and titles of books published); current research interests; association with any professional associations. Author membership information, e.g., is a member of the IEEE and the IEEE Computer Society, if applicable, is noted at the end of the biography