DEVELOPMENT OF DYNAMIC BAYESIAN NETWORKS MODELS

Mahdi Shafiee Kamalabad

February 2016

Introduction

0.1 Systems Biology and Postgenomic area of research

Systems Biology aims at an improved understanding of biology and its mechanisms, in particular of the functions and interactions of key elements of living systems (DNA, RNA, proteins, cells). Systems biology is the sum of novel methodological approaches to (a) gather and analyse this knowledge, (b) to translate this knowledge into a thorough understanding that eventually will allow predicting the behavior of the complex and dynamic networks that regulate living specimen. The focus of systems biology is clearly not limited to the description of existing knowledge using a different syntax, but aims at combining available methods synergistically to develop novel approaches for the characterization of the behavior of biological networks. Whereas traditional research in the Life Sciences generally emphasizes the operational necessity of a gene, protein, component for the correct execution of a particular pathway or regulatory mechanism (what is it doing, is it necessary or sufficient?), it is the Systems Biology that expands these questions in an attempt to define the relevance of a particular component or mechanism for fulfilling a biological goal (Why is it performing the way it is to accomplish this or that task? How does the component need to be designed in order to fulfill its function)(Ralf Baumeister(2005-2010)). This field requires the concepts of many other fields such as biology, chemistry, computer science, statistics, physics and engineering. To be more precise, system biology aims to clarify how higher level properties of complex biological systems arise from the interactions among their parts. For example, interactions of cells and also the key elements such as DNA, RNA, proteins in a cell with respect to one another. Thus one can say the objective of systems biology [can be] defined as the understanding of network behavior and in particular their dynamic aspects, which requires the utilization of mathematical modeling tightly linked to experiment. (Cassman, 2005)

The term "postgenomics" refers to functions ascribed to genes and their related proteins that describe the regulatory networks controlling metabolism, protein synthesis and signal transduction pathways (Lange B., Ghassemian M. 2005). In a broad sense postgenomics is transforming our understanding of diseases and health, our environment and the categories of race, class and gender.

0.2 Static Bayesian network and Dynamic Bayesian network in system biology

Inference of gene regulatory networks from high-throughput postgenomic data is one of the major challenging problems in the topical field of system biology. Among the related examples are the reconstruction of transcriptional regulatory networks from gene expression data, the inference of signal transduction pathways from protein concentrations and the identification of neural information

flow operating in the brains of songbirds or ecological networks. Therefore a variety of powerful statistical and computational methods based on graphical models, such as Bayesian networks (BNs) (Friedman et al. 2000), have been proposed and commonly used as a prominent tool for extracting and encoding knowledge from data.

BNs combine principles from graph theory, probability theory, computer science, and statistics. In the same vein, Bayesian network is defined by a graphical structure, a family of (conditional) probability distributions, and their parameters, which together encode the joint probability distribution for a set of variables of interest and represent, therefore, conditional independencies between them.

Recently there has been notable interest in learning and applying dynamic Bayesian networks (DBNs) with a variety of applications in signal processing and computational biology in which a main issue in systems biology nowadays is to understand, analyse and model the dynamic topological and functional properties of biological networks.

DBNs are BNs for dynamic processes. To take the temporal modifications or changes into account, one has to use dynamic Bayesian networks, indicating that the value of the child node depends on the value of the parent at an earlier time, specifying a Bayes net model which has "time delay" links. In fact Bayesian networks allow variables in a Bayesian network to be time dependent, in order to model time series or sequences. This kind of graphical model allows feedback loops and recurrent regulatory structures. This means that graphs do not have to be acyclic, unlike the static Bayesian networks which are directed acyclic graph (DAG). Thus DBNs can have directed cycles, as long as there is a delay link somewhere along each cycle. Delay links can be used to model feedback.

In DBNs, the fundamental assumption is the generation of time-series from a homogeneous Markov process. This presumption is too restrictive to be valid in many real world applications. In particular it cannot deal with heterogeneity and non-stationarity that usually happen in temporal processes. There are many examples which violate this assumption such as changes in response to external stimuli like regulatory interactions and signal transduction processes in the cell which are usually adaptive. Therefore the assumption of stationarity can potentially lead us to inaccurate results. Now the question is how to model and efficiently analyze real world problems (dynamic process) that follow non-homogeneous and non-stationary Markov process.

0.3 literature overview

Recently, a number of papers has been published on the relaxation of the stationarity assumption. First we have seen Talih and Hengartner (2005), Xuan and Murphy (2007), which all treat undirected graphical models. More recently DBN have been introduced for which we refer to Lebre (2007), Grzegorczyk et al (2008), Robinson and Hartemink (2009, 2010), Ahmed and Xing (2009), Kolar et al. (2009), Lebre et al. (2010), Dondelinger et al. (2010, 2012), Husmeier et al. (2010),

Grzegorczyk and Husmeier (2011), Grzegorczyk and Husmeier (2013), Grzegorczyk (2015). Some of these works complemented DBN with a Bayesian multiple changepoint process (Lebre 2007; Robinson and Hartemink 2009, 2010; Lbre et al. 2010; Dondelinger et al. 2010, 2012; Husmeier et al. 2010; Grzegorczyk and Husmeier 2011, Grzegorczyk and Husmeier 2013, Grzegorczyk 2015). We shall discuss some of these articles in a little bit more detail bellow.

Grzegorczyk et al (2008) have proposed a non-linear and non-homogeneous generalization of the BGe score for Bayesian networks using the Bayesian Gaussian Mixture model(BGM). The BGe stands for Bayesian metric for Gussian network having score equivalence. The BGe score was developed as a scoring metric for a Baysian network of continuous variables under the assumption that the data are sampled from a multivariate Gaussian distribution. This score is first derived for a complete Bayesian network where every pair of distinct nodes is connected by a directed edge. It is assumes a prior on parameter to be a normal Wishart distribution so that one can obtain a closed-form marginal distribution(Christine Sinoquet and Raphal Mourad 2014). The discrete alternative is called BDe where the variable have multinomial distribution with unknown parameters. These parameters have Dirichlet distribution as a prior.

A BGM is based on a mixture model. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population and got identified by a sum of weighted distributions over the number of components. The BGM allocates different measurements to different components using latent variables. In a BGM model a uniform prior distribution on graphs for the real gene expression data is assumed. For the number of components, Truncated Poisson distribution is assumed as a prior. This truncated Poisson distribution is known to be suitable for finite mixture models (Nobile, 2005). A further assumption for the probability distribution of the allocation vector conditional on the number of components is given by the product of non-negative mixture weights.

Basically a BGM model using free allocation approach, divides the data into segregated compartments (data subsets). In this model inference of network structures is based on the assumption that network structures are kept fixed for all components. It is important to take into account that the non-varying network allows for some information sharing among compartments. Then each compartment is modelled separately and independently with the Gaussian BGe scoring metric for Bayesian networks.

The Bayesian approch and sampling the graph (for finding the most consistent graph with the given data), the number of components and the assignment of latent variables from the posterior distribution with MCMC, using the allocation sampler of Nobile and Fearnside (2007) as an alternative to RJMCMC introduced by Green(1995), play the main role in BGM approach.

Following that, the authors have evaluated BGM model in three different basis: The first evaluation was in term of network reconstruction accuracy, assuming that the true network structure is given. The reconstruction accuracy can be measured with receiver operator characteristic (ROC)

curves (e.g. Husmeier 2003). There was an improved result for network reconstruction accuracy for both real and synthetic network. Secondly, evaluation was in term of statistical significance where the authors evaluated the improvement on gene expression time series for two different systems (viral challenge of macrophages, and circadian rhythms in plants). In these systems the BGM outperform the classical BGe score. Thirdly they compared the results with biological finding. In this case the results have had high biological validity.

The BGM model can be used for both static and dynamic gene expression data. This amount of flexibility makes its application to time series data suboptimal and substandard to some extent. To improve the performance of the BGM model on time series data Grzegorczyk et.al.(2011) proposed an amended approach by replacing the free allocation of data points (which was used in BGM) with a changepoint process. A changepoint is a data point in which the data points after that start having different trend. Therefore the data points between two consecutive delimiting changepoints seem to have approximately the same nature. The reason for replacing free allocation of data points with changpoint process is to take the temporal structure into consideration. Therefore, each time series are divided into several separated segments and each time series segment belongs to two consecutive demarcating changepoints. In essence in this work the proposed BGMD model divides the time series into homogeneous segments, each corresponding to a component of the mixture model. The rationale behind that is the restriction of the configuration space, from exponential to polynomial complexity which in this case is substantial as well as accessibility of allocation space which is a priori more plausible. Another assumption considering by authors in this work is smoothness, whereby the nature of the regulatory processes at two neighboring time points is a priori more similar. It means that two adjacent time point are allocated to the same compartment of the mixture model than distant ones.

In a BGMD model, Like the BGM, the network does not vary among all segments and remains unchanged, and each segment is modelled independently and separately with the Gaussian BGe model for Bayesian networks. In principle, in this paper the authors generalized the BGe model by latent allocation vector V. This vector plays the role of a filter which divides the data into different compartments. It means that allocating time points t = 2,. . . ,m to different K mixture compartments and then separate independent BGe scores can be computed in closed-form for each of the K compartments given the latent vector V. Here the number of components , K, will be inferred from the data by employing changepoint birth and death moves, with reversible jump Markov chain Monte Carlo sampler (RJMCMC)(Green 1995). In this work the authors took a truncated Poisson distribution for the number of compartment and a uniform distribution over all graph structures subject to a fan-in restriction as prior. A fan-in restriction means restriction in the number of parents nodes. The MCMC sampler, in this paper, does not sample allocation vector directly, just considering local changes of it based on changepoint birth, death and reallocation moves (Green 1995), different from BGM model. In principle the authors decrease the complexity of the allocation

space by using the prior knowledge that adjacent time points belong to the same component(with high probability). Generally the inference here is based on the Bayesian approach, and Markov chain Monte Carlo (MCMC) sampler for sampling the network structure, the number and locations of changepoints from the posterior distribution. One of the empirical results in this paper showed that regarding to time series the proposed BGMD model outperforms the BGM model. To be more specific, the BGMD model approximates non-stationary regulatory relationships more efficiently than the BGM model by avoiding spurious self-loops, and thus conceding more accurate network reconstruction accuracy.

In 2009, Grzegorczyk and Husmeier, have proposed a non-stationary generalization of the BGe model using a mixture of BGe models. This proposed model is continuous-valued non-stationary dynamic Bayesian network. Proposing and evaluating a non-stationary continuous-valued DBN with information sharing among various time series segments via a constrained structure is one of the main goal of this work. Here the authors only allow the parameters to vary with time and the concept of non-stationary is related to the parameters and the network structure is unchanged among all segments. Computing the score in this paper, is based on a non-stationary generalization of the BGe not BDe which, in fact, complements the work of a non-stationary BDe model, avoiding the need for data discretization. Bearing in mind that this node-specific non-stationary patterns provides more flexibility. The allocation matrix defined here divides the data into several separate subsets, and each of them is relevant to a separate BGe model with its own parameters. Note that the allocation matrix is composed of different vectors. The vectors are node-specific, i.e. different nodes can have different breakpoints. According to the above brief discussion, the selected probability model in this paper, is based on a mixture model with local probability distributions. The mixture model can approximate any probability distribution under a free allocation of the latent variables closely to some extent. Assignment of data points to mixture components from a free allocation to a changepoint process is the main feature of this paper. Thus, this will decrease the complexity of the latent variable space and include the prior knowledge that, in a time series, neibouring time points are likely to belong to the same component and have same distribution. The authors have presented a comparative evaluation of the network reconstruction accuracy on synthetic data. Findings confirm that non-stationary BGe model proposed in this paper has an obvious performance improvement comparing with the classical stationary models BDe and BGe as well as the non-linear/non-stationary model that we discuss above. The application of their model has led to a plausible data segmentation, and more agreement with the biological reality.

Previous studies which we discussed some of them above, take no notice of the fact that many systems and processes, e.g., regulatory processes and signalling pathways in the cell, develop and make progress little by little. For example during an organisms development (morphogenesis) or in adaptation to changing environmental conditions. Therefore they just compute the marginal likelihood just from an uninformative parameter prior that is the same for all time series segments. In

other words and in more details each time series segment defined by two delimiting changepoints is associated with separate node-specific DBN parameters, and in this way the conditional probability distributions are allowed to vary from segment to segment. Under certain regularity conditions, specially parameter independence and conjugacy of the prior, the parameters can be marginalized out in closed form in the likelihood. The inference thus reduces to sampling the network structure as well as the number and location of changepoints from the posterior distribution, which can be done by reversible jump Markov chain Monte Carlo (RJMCMC) (Green 1995), e.g., as in Lebre et al. (2010) or Robinson and Hartemink (2010), or with dynamic programming (Fearnhead 2006), as in Grzegorczyk and Husmeier (2011). Considering the real word applications, the assumption of parameter independence will be restricting and therefore doubtful. For example the cellular processes during an organism's development (morphogenesis) or its adaptation to changing environmental conditions lead to over felexibility. Over -flexibility however, is one of the consequences of complete parameter independence which happens by disregarding the fact of evolutionary aspect of adaptation processes, where the majority of segment-specific regulatory relationships among the interdependent quantities tend to undergo minor and gradual adaptations. Hence improvement of inference of interaction strength associated with the given time interval, especially for sparse data by the knowledge behind the interaction at other time interval is one assumption which can be considered. This assumption is more realistic than that of different regulatory circuits from scratch by the reinvention of nature in adjacent time intervals. Thus for solving this problem, there are two approaches for information sharing in time series segmented by multiple changepoints: 1- sequential information coupling, and 2- global information coupling.

In the sequential information coupling, information is shared between neighbor segments. It means that information will be coupled sequentially. In the global information coupling, segments are treated as interchangeable units, and information is shared globally via a certain parameter.

Sequential information coupling is appropriate for a system which is changing gradually in a period of time or in the other words a system in the process of evolution and development, e.g. in morphogenesis, for different stages of insects life cycle. Then one would assume that consecutive or nearby stages , like larvae and embryo, have more commonalities than distant ones, like larvae and adult insect. On the other hand global information coupling is more appropriate when time series segments are related to different experimental scenarios or environmental conditions. Therefore in this case information will be shared globally via one or more particular parameters not sequentially. Verifying a natural order for coupling information in this case is almost impossible like when a yeast strain is open to different carbon sources, such as glucose, galactose, and fructose. In such a case there is no natural order by which information should be shared, and it is better to consider the segments as interchangeable unit. These coupling schemes have been applied till now but not the information coupling with respect to the interaction parameters in the sense discussed above. For

instance regularizing of DBNs with time-varying network structures, which penalize network structure changes sequentially (Dondelinger et al. 2010) and globally (Husmeier et al. 2010; Dondelinger et al. 2012). Both papers presume complete parameter independence, in the same way as Robinson and Hartemink (2009, 2010) and Lbre et al. (2010).

For complementing these studies Grzegorczyk and Husmeier (2012a) proposed a non-homogeneous dynamic Bayesian network with sequentially coupled interaction parameters. The main assumption in this scheme is the similarity of the parameters associated with separate time series segments as a priori. Thus they have introduced a coupling hyperparameter which is shared among the segments sequentially and itself will can be inferred from the data with Bayesian approach. In this approach the prior distribution for next segment's interaction parameter is the posterior probability of current segment but because the first segment does not couple the network to any previous segment (there is no previous segment in this case), the knowledge of previous studies will be used as a prior. Imposing the same prior for all segments and for all intraction parameter vectors is resonable when the processes on all segments are independent. But for seqentially information coupling, it will be assumed that the later process is a modified version of the current process and therefore, the result of the cuurent process will be used for next process(the cellular processes during an organism's development). In this work for the posterior distribution of the regression parameters, they use the approach of standard Bayesian sequential analysis (Carlin and Louis (2009)) with the application of standard Gaussian integrals (Bishop (2006)). More precisely each segment are modelled by linear regression model with different parameter vectors. Using Bayesian sequential analysis approach, for next segment prior they set the product of additive Gaussian iid noise's variance and some matrix C with multiplicative scalar as a variance (for preventing any conflict and paradox) and the posterior mean of previous segment as a mean. The dependence of the covariance matrix on the noise variance leads to a fully conjugate prior in both the regression parameters and the noise variances that allows both parameter groups to be integrated out in the marginal likelihood. Their assessment on simulated data in term of reconstruction accuracy has shown that the proposed model leads to an improvement. Considering robustness their investigation was based on systematic evolution of the interaction parameters (via node-specific vector rotations in parameter space). Also in this paper they have demonstrated on gene expression time series from Drosophila melanogaster and Arabidopsis thaliana that the proposed coupled TV-DBN model (TV refers to time varying) yields more strongly correlated interaction parameters for segmented time series than the conventional uncoupled TV-DBN model. Finally, with altering the (unknown) number of changepoints, they have shown that the proposed model robustness will increase on real data (from RT-PCR experiments in Saccharomyces cerevisiae).

Grzegorczyk and Husmeier (2013) have proposed a model for global information sharing with respect to the interaction parameters. Their model was based on a switching piecewise homogeneous autoregressive process. To improve the result of their study, they complete their previous

conference paper work (Grzegorczyk and Husmeier 2012b) in some new aspects: Firstly, allowance for taking information sharing among nodes into account unlike the previous works. Therefore here, the authors introduced an extra (level-3) layer to the hierarchy of the proposed model. The proposed model hierarchically couples the node-specific noise variances and the node-specific coupling strengths between the segment-specific interaction parameters. In addition, they presented nine different coupling schemes for the noise variance hyperparameters and empirically compared three of them. Secondly, they introduced and applied a novel collapsed Gibbs sampling step, instead of uncollapsed Gibbs sampling step of the original MCMC algorithms which was less efficient. Thirdly the authers introduced blocking techniques and then showed how this collapsed Gibbs sampling step and blocking techniques can develop the original MCMC scheme and introduce a novel advanced MCMC algorithm which performs significantly better than the original MCMC sampling scheme from Grzegorczyk and Husmeier (2012b) in terms of convergence and mixing. The authers also showed that the advanced MCMC sampling scheme reaches a better network reconstruction accuracy. Fourthly, for investigating the robustness of the proposed model, they pay attention to one single hyperparameter setting, (taken from Lebre et al. (2010)) and consider the variation of the fixed (hyper-)hyperparameters. Here they systematically vary the (hyper-)hyperparameters of those (hyper-) priors that are important for the noise variances and coupling strengths among segments and they investigated their influence on the performance. Fifthly, they compared the proposed global information coupling model and the sequential information sharing model (Grzegorczyk and Husmeier 2012a), and showed the potential fundamental improvement achieved with the new approach.

To be more precise in this proposed model, they assume a set of changepoints for each node, that divide the data into K subsets. All subsets are modeled with the separate linear regression model but with different parameter vectors. For information sharing among the segments, it is necessary to introduce a random vectors in which the information coupling will happen via them. In principle these random vectors is the coupling hyperparaeter. Moreover, the authors add an extra layer to the Bayesian hierarchy and consider the mean of regression parameters, let say mg, as a random vector, which is given a conjugate Gaussian prior distribution. Also they assume that the regulatory network structure is kept fixed over time. Note that when the hyperparameters, mean of regression parameters, mg, are fixed, the regression (intraction) parameters, are conditionally independent. Hence, there is no information coupling between them. When the hyperparameters mg are flexible, the so called d-separation is lost, and the intraction parameters, become dependent or coupled, as a consequence of the marginalization over mg. For changpoint they consider 2 scenario 1- fixed change point 2- variable changepoint. later scenario, is based on a point process, where the distribution of the distance between two successive points is a negative binomial distribution. Given the data, the final purpose is to infer the network structure, from the marginal posterior distribution. The other variable quantities are nuisance parameters, which are marginalized over; for example the

changepoints, the interaction parameters, the noise variance hyperparameters, and the signal-to-noise hyperparameters and also the other hyperparameters which help to better inference of the network structure. Sampling from the joint posterior distribution follows a Gibbs sampling sheme, so variables are sampled from their full conditional distributions. Whenever possible, sampling from the closed-form distributions and use collapsing, i.e. marginalizing out (some) variables from the Markov blankets. Where closed form distributions are not available, they employed RJMCMC steps (Green 1995).

0.4 Novel fast Markov chain Monte Carlo (MCMC) sampler

Markov Chain Monte Carlo (MCMC) schemes to construct a Markov chain that has the desired distribution as its equilibrium distribution, will be used for sampling from a probability distribution. The chain's state after a number of steps, burn in phase, is then considered as a sample of the desired distribution. It means that if this Markov chain is simulated for long enough, then it generates a random sample from the desired distribution. The quality of the sample will be improved as the number of steps increases. This approach is called the "MCMC Sampling scheme" or the "MCMC sampler". Therefore this is necessary for the model which we will propose and discuss briefly in next section, a novel fast MCMC algorithms can be employed to search the space of network structures systematically for those that are most consistent with the data. It means that In this thesis we will improve previous MCMC samplers to achieve more accurate results. In this scheme the other related nuisance parameters will also be updated like changrpoint set, noise variance, etc.

0.5 Outline of thesis contribution

In this thesis taking the last research, which we discussed some of them earlier, into consideration, we will develop the dynamic Baysian network (DBN) in two aspects. First we will improve by the sequential coupling with the option to uncouple in between. Second we will combine the sequential and global coupling concept in one network structure to reach better result than before.

Bibliography

Ahmed, A., and Xing, E.(2009). Recovering time-varying networks of dependencies in social and biological studies. *Proceedings of the National Academy of Sciences 106* , 11878–11883.

Baumeister R, Center for Biological Systems Analysis University of Freiburg. *Direktor des ZBSA (2005 - 2010) Mitglied des Direktoriums.*

Dondelinger, F., Lebre, S., and Husmeier, D.(2012). Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure. *Machine Learning 90* , 191–230.

Grzegorczyk M (2015). A non-homogeneous dynamic Bayesian network with a hidden Markov model dependency structure among the temporal data points *Machine Learning* , DOI 10.1007/s10994-015-5503-2.

Grzegorczyk, M., Husmeier, D. (2009). Non-stationary continuous dynamic bayesian networks. *Advances in nural information processing systems 22 conference* , 682–690.

Grzegorczyk M,Husmeier D, EdwardsK,Ghazal P, MillarA (2008). Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler. *Bioinformatics* , 24: 2071–2078.

Grzegorczyk M, Husmeier D (2008). Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Mach Learn* 71: 265–305.

Grzegorczyk, M., Husmeier, D. and Rahnenfuehrer, J.(2011). Modelling non-stationary dynamic gene regulatory processes with the BGM model. *Computational Statistics.* , 26:199–218.

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 82*(4), 711–732.

Grzegorczyk, M., and Husmeier, D. (2012). A Non-Homogeneous Dynamic Bayesian Network with Sequentially Coupled Interaction Parameters for Applications in Systems and Synthetic Biology. *Statistical applications in genetics and molecular biology. 11* (4) [7].

Grzegorczyk, M., and Husmeier, D. (2013). Regularization of non-homogeneous dynamic Bayesian networks with global information-coupling based on hierarchical Bayesian models. *Machine learning 91* (2), 105–154.

Husmeier, D., Dondelinger, F., and Lebre, S. (2010). Inter-time segment information sharing for nonhomogeneous dynamic Bayesian networks. *In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta (Eds.), Proceedings of the 24th annual conference on neural information processing systems (NIPS)* , 901–909 Curran Associates.

Kolar, M., Song, L., and Xing, E. (2009). Sparsistent learning of varying-coefficient models with structural changes. *In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta (Eds.). Advances in neural information processing systems (NIPS) 22*, 1006–1014.

Lange B., Ghassemian M. (2005). Comprehensive post-genomic data analysis approaches integrating biochemical pathway maps. *Phytochemistry 66* (4),413–51.

Lebre, S. (2007). Stochastic process analysis for genomics and dynamic Bayesian networks inference. *PhD thesis, Universit dEvry-Val-dEssonne, France* .

Lebre, S., Becq, J., Devaux, F., Lelandais, G., and Stumpf, M. (2010). Statistical inference of the time-varying structure of gene-regulation networks. *BMC Systems Biology 4* .

Robinson, J., and Hartemink, A. (2009). Non-stationary dynamic Bayesian networks *In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.), Advances in neural information processing systems (NIPS) 21* , 1369–1376, San Francisco: Morgan Kaufmann.

Robinson, J., and Hartemink, A. (2010). Learning non-stationary dynamic Bayesian networks. *Journal of Machine Learning Research 11* , 3647–3680.

Talih, M., and Hengartner, N. (2005). Structural learning with time-varying components: Tracking the crosssection of financial time series. *Journal of the Royal Statistical Society B,* 67, 321–341.

Xuan, X., and Murphy, K. (2007). Modeling changing dependency structure in multivariate time series. *In Z. Ghahramani (Ed.), Proceedings of the 24th annual international conference on machine learning (ICML 2007)* 1055–1062. Omnipress.