

University of Groningen

Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data

Shafiee Kamalabad, Mahdi

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Shafiee Kamalabad, M. (2019). *Advanced non-homogeneous dynamic Bayesian network models for statistical analyses of time series data*. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 1

Introduction

Inferring network topologies of interacting units from temporal data is a statistically challenging task in many scientific disciplines. The goal is to learn the dependencies between the units from the data and to represent them in form of a network. A topical example is the field of computational system biology, where one of the major goals is to learn cellular networks, such as gene regulatory transcription networks (see, e.g., [18]) and protein signaling pathways (see, e.g., [51].) Further examples include neural information flow networks [60] and ecological networks [2].

One class of models that has been widely applied to deal with this challenge, is the class of dynamic Bayesian network (DBN) models. The underlying assumption is that the regulatory processes are homogeneous, so that DBNs assume the network interaction parameters to stay constant in time. For many real-world applications, this homogeneity assumption is too restrictive and can lead to wrong conclusions. To address this shortcoming, non-homogeneous dynamic Bayesian networks (NH-DBNs) have been proposed in the literature. Section 1.3 of this chapter gives an overview to different types of NH-DBNs and also discusses their advantages and disadvantages.

1.1 Static and dynamic Bayesian networks

Dynamic Bayesian networks (DBNs) are a popular class of models for learning the dependencies between random variables from temporal data.¹ Unlike in static Bayesian networks (BNs), a dependency between two random variables X and Y is typically interpreted in terms of a regulatory interaction with a time delay. A directed edge from variable X to variable Y , symbolically $X \rightarrow Y$, indicates that the value of variable Y at any time point t depends on the realisation of X at the previous time point $t - 1$. Therefore, in DBNs, since all interactions are subject to a time lag the network does not have to be acyclic.

¹DBNs extend standard static Bayesian networks (BNs) with the concept of time.

Typically, various variables X_1, \dots, X_k have a regulatory effect on a target Y , and the relationship between X_1, \dots, X_k and Y can be represented by a regression model that takes the time lag into account. E.g., if the time lag is one time point, the regression model takes the form:

$$y_t = \beta_0 + \beta_1 x_{1,t-1} + \dots + \beta_k x_{k,t-1} + u_t \quad (t = 2, \dots, T) \quad (1.1)$$

where T is the number of time points, y_t is the value of Y at time point t , $x_{i,t-1}$ is the value of covariate X_i at time point $t - 1$, β_0, \dots, β_k are regression coefficients, and u_t is the “unexplained” noise at time point t .

1.2 Network inference

In dynamic Bayesian network (DBN) applications there are usually N domain variables Y_1, \dots, Y_N and the goal is to infer the covariates of each variable Y_i . As the covariates can be learned for each Y_i separately, DBN learning can be thought of as learning the covariates for a set of target variables $\{Y_1, \dots, Y_N\}$. There are N regression tasks, and in the i -th regression model, Y_i is the target variable and the remaining $N - 1$ variables take the role of the potential covariates. The goal is to infer a covariate set $\pi_i \subset \{Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_N\}$ for each Y_i . From the covariate sets π_1, \dots, π_N a network can be extracted. The network shows all regulatory interactions among the variables Y_1, \dots, Y_N . An edge $Y_j \rightarrow Y_i$ indicates that Y_j is a covariate of Y_i , i.e. that $Y_j \in \pi_i$. In the terminology of DBNs Y_j is then called a regulator of Y_i . All variables in π_i are regulators of Y_i ($i = 1, \dots, N$).

1.3 Non-homogeneous DBNs (NH-DBNs)

The conventional assumption in dynamic Bayesian network models (DBNs) is that the regulatory relationships are homogeneous, so that the network parameters do not change in time. That is, the regression coefficients β_0, \dots, β_K in Equation (1.1) stay constant across all time points ($t = 2, \dots, T$). Thus DBNs infer the network structure along with one single set of network parameters, and those parameters then apply to the whole time series. This homogeneity assumption is very restrictive and can lead to wrong results and conclusions. Therefore, DBNs cannot deal with non-homogeneous regularity processes, which often arise in systems biology. For example in a cellular network, the strength of the regulatory interactions are often exposed to (unobserved) external factors, such as cellular, environmental and/or experimental conditions (see, e.g., [8]), that influence the interactions. This renders the traditional DBNs inappropriate for most of the applications in systems biology. Therefore non-homogeneous dynamic Bayesian network models (NH-DBNs) have been proposed (see, e.g., [37]). NH-DBNs are a powerful statistical tool and do not make use of the homogeneity assumption.

The concept of non-homogeneity leads to time varying network parameters and/or time varying network structures. Therefore, NH-DBNs can be divided into two conceptual groups: NH-DBNs that allow only the network parameters to vary in time (see, e.g., [23]) and NH-DBNs that also allow the network structure to be time-dependent, see, e.g., [49], [38] or [14]. A statistical problem is that gene expression time series are often short so that NH-DBNs with time-dependent network structures are over-flexible and lead to inflated inference uncertainties. With regard to our biological applications throughout this thesis, we therefore focus on NH-DBNs which only allow the network parameters to change.

NH-DBNs with time-varying network parameters have been implemented with various allocation models to divide the data into disjoint data subsets:

- DBNs have been combined with free mixture models (MIX); see, e.g., [34] or [26].
- DBNs have been combined with hidden Markov models (HMM); see, e.g., [62] or [22].
- DBNs have been combined with multiple changepoint processes (CPS). see, e.g., [38] or [23].

The models infer the data segmentation, the joint network structure and the segment- or component-specific interaction parameters altogether from the data. In this thesis we focus on changepoint-divided (CPS) NH-DBNs, which have become the most widely applied NH-DBNs.

1.3.1 Changepoint-divided NH-DBNs

Changepoint-divided (CPS) non homogeneous dynamic Bayesian networks (NH-DBNs) models infer changepoints, which divide the data into disjunct segments. The data within each segment are modeled with linear regression models. There is a shared network structure among segments, and the segment-specific network parameters are learned for each segment separately. In typical applications in systems biology these NH-DBNs divide a short time series into even shorter segments, containing only a few data points. Learning the network parameters for each segment separately (conventional ‘uncoupled’ NH-DBN models) see, e.g. [38], then inevitably leads to over flexibility and inflated inference uncertainties. Moreover, they do not incorporate the reasonable prior assumption that neighbouring segments are often more likely to have similar network interaction parameters than distant segments.

To address these bottlenecks, more realistic models which allow for gradual adaptations of the network interaction parameters, have been proposed. E.g., the frequentistic models, proposed by [3], [36] and [35]. Those models make use of L1-regularized regression models (‘LASSO’) for the network parameter inference, and they employ a second L1 regularization term to penalize dissimilarities between network parameters of neighbouring segments. In those frequentistic models inference is based on penalized maximum likelihood approaches, and the

fixed regularization parameter has to be optimized by cross-validation or in terms of the Bayesian Information Criterion (“BIC”). Bayesian models with coupling mechanisms between the segment-specific parameters have also been proposed. In [25] it was proposed to globally couple the segment-specific parameters. The key idea is to treat the segments as interchangeable units and to impose a shared hyperprior onto the prior expectations of the segment-specific parameters. In a complementary work ([24]) it was proposed to sequentially couple the parameters. The fully (sequentially) coupled model was developed to keep the network parameters of each segment similar to those of the previous segment. Here the parameters within segment h obtain as prior expectations their posterior expectations from the preceding segment $h - 1$, and the coupling strength i.e., the variance of the network parameter priors (the similarity of the regression coefficients), is regulated by a coupling hyperparameter λ . This model can thus be seen as a Bayesian counterpart of the frequentistic models, mentioned above. The Bayesian models are inferred with Reversible Jump Markov Chain Monte Carlo (RJMCMC) simulations [21], and a comparative evaluation study of network reconstruction methods in [1] showed that the Bayesian models tend to reach higher network reconstruction accuracies than the frequentistic models.

1.3.2 The concept of parameter coupling

Parameter coupling can lead to significantly improved network reconstruction accuracies when the segment-specific parameters are similar, as shown in [24] and [25]. However, recently we have found that coupling can become counter-productive when the segment-specific parameters are dissimilar. The reason for that is that neither the sequential nor the global coupling scheme has an effective mechanism for uncoupling. When the segment-specific parameters are dissimilar, coupled NH-DBNs can only reduce the coupling strengths by making the parameter priors vague. This renders them significantly inferior to NH-DBNs without any coupling mechanism. Moreover, the fully coupled model suffers from another serious bottlenecks: The model couples all neighbouring segments ($h - 1, h$) with the same coupling strength. That is, it possesses only one single coupling hyperparameter λ which is shared among all segments $h > 1$ and all covariates.² To shed more light onto this, we note that both coupling mechanisms have been designed such that if a node A is regulated by a set of other nodes, e.g. $B \rightarrow A \leftarrow C$, then both edges have to be coupled with the same strength across all segments.³ For many real-world applications this is unrealistic. E.g., the regulatory effect of B on A (i.e., the parameter associated with $B \rightarrow A$) can stay similar, while the regulatory effect of C on A can be subject to major changes. To re-use a traffic flow analogy from [48]: The traffic flow on the roads

²The models from [3], [36] and [35] suffer from the same drawbacks. Those models also possess only one single regularization (‘tuning’) parameter which determines the similarity of the network parameters among segments. The coupling strength between segments can neither vary over time nor is there any mechanism for uncoupling segments.

³We will therefore also refer to these models as *fully* coupled NH-DBNs.

is different during rush hours and off-peak times. But rush hours usually do not affect the traffic flow on all roads. Typically there are susceptible roads with tailbacks during rush hours, while the traffic demand on other roads might stay constant.

1.4 Another conceptual problem

In many applications in systems biology, we encounter data that are collected under different experimental conditions. Instead of one single (long) time series, which can be divided into segments with natural temporal order, there are K (short) time series. These individual time series $k = 1, \dots, K$ have no natural order and are exchangeable units. That is, the available data are then automatically divided into K unordered components (=the individual time series), and there is no need for inferring the segmentation. In this situation it is often unclear a priori whether the network parameters are actually component-specific or whether they are constant across components. If the parameters stay constant, all data could be merged and be analyzed altogether with one single homogeneous DBN model. If there are component-specific parameters, then the data should not be merged and it would be better to analyze each time series separately. In the latter case, it can be useful to adapt the global parameter coupling scheme from [25], so as to encourage the network parameters to stay at least similar among components. The bottleneck of both approaches is that either the parameters are assumed to stay constant or that the parameters are assumed to be component-specific. In real-world applications there can be both types of parameters. E.g., if a variable Y is regulated by two other variables, symbolically $X_1 \rightarrow Y \leftarrow X_2$, then the regulatory interactions $X_1 \rightarrow Y$ might not be affected by the experimental conditions, while the regulatory $X_2 \rightarrow Y$ might be influenced by the condition, e.g. for $K = 2$ in terms of a linear regression model, one might have:

$$E[Y|X_1 = x_1, X_2 = x_2] = \begin{cases} \alpha x_1 + \beta x_2 & \text{if } k = 1 \\ \alpha x_1 + \gamma x_2 & \text{if } k = 2 \end{cases} \quad (1.2)$$

A homogeneous model is then inappropriate, since it would ignore that the regression coefficients β and γ are different. A non-homogeneous model comes with the drawback that the same regression coefficient α has to be learned two times separately. This is disadvantageous when the data within each component ($k = 1, 2$) are sparse and uninformative.

1.5 The aim of this thesis

To summarize what has been discussed in the previous sections, Figure 1.1 shows a graphical overview of the various NH-DBN models. In this thesis, we put our

focus on the sequential and global coupling scheme and show how the coupled models can be improved, so as to address the above-mentioned drawbacks. We propose four novel non-homogeneous dynamic Bayesian network (DBN) models, which are more flexible and thus have the potential to capture the underlying interactions more accurately than the earlier proposed models.

1.6 Outline of thesis contribution

This thesis is organized as follows:

In **chapter 2**, we propose two new NH-DBN models to fix the deficits of the fully (sequentially) coupled NH-DBN model from [24]. The partially segment-wise coupled model can be seen as a consensus model between the uncoupled model and the fully coupled model. It has the uncoupled and a sequentially coupled NH-DBN models as limiting cases: If it couples all segments, it effectively becomes the fully coupled model. If it uncouples all segments, it effectively becomes the conventional uncoupled model. Moreover, we propose the generalized coupled model, which is a generalization of the fully sequentially coupled model. Like the fully sequentially coupled model, the new model does not have any option to uncouple, but it possesses segment-specific coupling parameters and allows for different coupling strengths between segments. We will demonstrate that the partially segment-wise coupled model can lead to significantly improved network reconstruction accuracies, while we do not see any significant improvements for the generalized coupled model. In chapter 3 we therefore have a closer look the generalized coupled model and refine it.

In **chapter 3**, we refine the generalized fully coupled model. In particular, we impose a hyperprior onto the second hyperparameter of the coupling parameter prior to allow for more information-exchange among the segment-specific coupling strengths.

In **chapter 4**, we present a novel partially edge-wise coupled model. Unlike the partially coupled model from chapter 2, this model infers for each individual edge whether the associated parameters should be coupled or stay uncoupled across the segments.

In **chapter 5**, we introduce another consensus model, which we refer to as a partially NH-DBN model. This model has been developed for the situation described in Section 1.4. The new model aims to infer the best trade-off between a homogeneous model (with constant parameters) and a non-homogeneous model (with component-specific parameters). In this chapter we also propose a Gaussian process based approach to deal with non-equidistant measurements. The (non-homogeneous) dynamic Bayesian network models assume that the domain variables have been measured at equidistant time points. For applications where this assumption is violated, we propose to employ a Gaussian process to predict the values at equidistant data points.

Chapter 6 presents a study, which is independent to those presented in the previous chapters. In chapter 6 we perform a comparative evaluation study

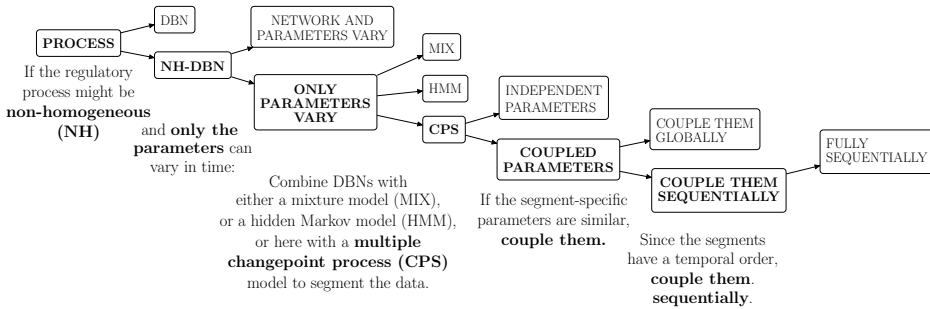


Figure 1.1: Overview of non-homogeneous dynamic Bayesian networks (NH-DBNs). We consider NH-DBNs whose parameters vary in time, and we use a multiple changepoint process (CPS) to segment the data into segments.

on popular non-homogeneous Poisson models for count data. For this study the standard homogeneous Poisson model (HOM) and three non-homogeneous variants, namely a Poisson changepoint model (CPS), a Poisson free mixture model (MIX), and a Poisson hidden Markov model (HMM) are implemented in both conceptual frameworks: a frequentist and a Bayesian framework. This yields 8 models in total, and the goal of this chapter is to shed some light onto their relative merits and shortcomings. The first major objective is to cross-compare the performances of the four models (HOM, CPS, MIX and HMM) independently for both modelling frameworks (Bayesian and frequentist). Subsequently, a pairwise comparison between the four Bayesian and the four frequentist models is performed to elucidate to which extent the results of the two paradigms (‘Bayesian versus frequentist’) differ. The evaluation study is performed on various synthetic Poisson data sets as well as on real-world taxi pick-up counts, extracted from the recently published New York City Taxi (NYCT) database.

Several parts of this thesis have previously been published in form of two journal articles, one in press, and four conference papers. One more paper has been submitted. The references are:

- Shafiee Kamalabad, M., Heberle A.M., Thedieck K. and Grzegorzcyk, M. (2018) (accepted and in press): Partially non-homogeneous dynamic bayesian networks based on Bayesian regression models with partitioned design matrices. *Bioinformatics*, <http://dx.doi.org/10.1093/bioinformatics/bty917>, (chapter 5, see [59]).
- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2018): Hierarchical Bayesian piecewise regression model with partially edge-wise coupled parameters. Submitted to *Journal of Computational and Graphical Statistics* (chapter 4).
- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2018): Improving nonhomogeneous dynamic Bayesian networks with sequentially

coupled parameters. *Statistica Neerlandica*, 72 (3), 281-305 (chapter 3, see [58]).

- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2018): Non-homogeneous dynamic Bayesian networks with edge-wise coupled parameters. *Proceedings of the International Workshop on Statistical Modelling*, vol. 1, 270-275, Bristol, England (chapter 4, see [57]).
- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2018): A new partially Coupled Piece-Wise linear Regression Model for statistical network Structure Inference. *Proceedings of the International Computational Intelligence methods for Bioinformatics and Biostatistics*, page 30, Caparica, Portugal (chapter 2, see [56]).
- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2017): A sequentially coupled non-homogeneous dynamic Bayesian network model with segment-specific coupling strengths. *Proceedings of the International Workshop on Statistical Modelling*, vol. 1, 173-178, Groningen, Netherlands (chapter 3, see [55]).
- Shafiee Kamalabad, M. and Grzegorzcyk, M. (2016): A non-homogeneous dynamic Bayesian network model with partially sequentially coupled network parameters. *Proceedings of the International Workshop on Statistical Modelling*, vol. 1, 139-144, Rennes, France (chapter 2, see [54]).
- Grzegorzcyk, M. and Shafiee Kamalabad, M. (2016): Comparative evaluation of various frequentist and Bayesian non-homogeneous Poisson counting models. *Computational Statistics*, 32 (1), 1-33. (chapter 6, see [28]).