



# Effect of environmental stress on regulation of gene expression in the yeast

Eitan Gross

Department of Physics, 226 Physics Building, University of Arkansas, Fayetteville, AR 72701, USA

## HIGHLIGHTS

- Transcriptional responses to external stimuli are modeled using ODE with an unknown forcing term.
- Hyperparameters are estimated using Gaussian Process priors and type II maximum likelihood.
- State-space trajectories representing signature responses to different environmental stimuli are generated.

## ARTICLE INFO

### Article history:

Received 9 September 2014

Received in revised form 18 November 2014

2014

Available online 5 March 2015

### Keywords:

Correlation function

Covariance matrix

Gene expression

Gaussian process

Maximum likelihood

Transcription factor

## ABSTRACT

Several mathematical models have been proposed to predict the activation state of a transcription factor (TF) from the expression levels of its target genes. This inference problem is complicated however due to the fact that different genes may be regulated by different activation schemes (linear, exponential, sigmoidal, etc.). In addition to transcription regulation, the rate of gene expression at any instantaneous point in time is also determined by the independent rates of baseline production and degradation. Consequently, the set of solutions to any model equations describe an infinite number of trajectories in probability space, thus rendering the problem NP-hard. In the current study we used a Gaussian process (GP) approach to address this inverse problem. Experimental gene expression data were modeled by a putative linear activation scheme and discrepancy between theory and experiment was modeled by a GP. Model hyperparameters were calculated using maximum likelihood estimates to generate continuous TF state-space profiles. Identifiability of model parameters was optimized by obtaining TF state-space functions for multiple genes *simultaneously*. We found that model parameters were sensitive to environmental stress conditions, producing different state-space profiles for different stresses.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Transcription factors (TFs) are proteins that regulate the expression of genes in the organism's genome by binding to regulatory binding sites in the promoter region of the gene. Clustering algorithms have been used in the past to identify novel binding sites of TFs from the expression patterns of their target genes [1]. However, unlike the coding regions, the promoter regions of most genes are not well understood and the nucleotide binding motifs are not highly conserved [2].

Furthermore, most TFs themselves need to be activated (i.e. by phosphorylation or dephosphorylation) before they can activate their target genes [3,4]. Several experimental approaches, such as protein binding arrays and chromosome immunoprecipitation (ChIP) have been used to identify the activation state of different TFs. However due to technical limitations and high costs of these experimental procedures they can only give a relatively static view of the system at a limited number of time points along the experiment.

E-mail address: [exg004@uark.edu](mailto:exg004@uark.edu).

<http://dx.doi.org/10.1016/j.physa.2015.02.076>

0378-4371/© 2015 Elsevier B.V. All rights reserved.

The problem can be addressed computationally with mathematical models that extract information on TF activation dynamics from the expression patterns of their target genes. Both linear and non-linear models have been proposed to infer the activation state of TFs and the expression levels of their target genes using this approach [5–9].

The inference problem is further complicated by identifiability of model parameters representing the independent processes of baseline mRNA transcription and degradation. Consequently, the set of solutions to any model describes an infinite number of trajectories in probability space, thus rendering the problem NP-hard.

To address the issue of model uncertainty, we propose here a Bayesian inference approach with a Gaussian process approximation in which the parameters of a putative activation model are estimated using a Gaussian process. Gaussian process (GP) is a non-parametric regression method that places probability distributions over functions. Several non-parametric methods have been used recently in modeling studies of gene expression regulation including Bayesian analysis [10,11], Dirichlet processes [12,13] and the Indian buffet process [14].

GP inference of a linear activation model of gene expression in the fruit fly has recently been used to identify potential target genes of the TFs Twist, and Mef2, under both single-target and multiple-target activation schemes [15]. The two TFs control mesoderm development in the fly. While their model successfully predicted the expression patterns of the target genes under the single-target activation scheme, it fell somewhat short in predicting their expression patterns under the more biologically-realistic multiple-target activation scheme. Here, we addressed this problem by introducing a hidden variable ( $h_i(t)$  in Eq. (1)) into our model which represents the activation level (i.e. concentration of the active form) of the TF. Hidden variables approach has been more recently applied in studying motion capture, computational developmental biology, and geostatistics [16], as well as transcription factor regulation of gene expression [17]. However, in our opinion, the issue of parameter identifiability has not been properly addressed in these studies.

In the current study, we used GP to place a prior on the space of functions of TF expression levels. Since the expression level of the TF in our model is a latent variable, we refer to it as a state-space function. State-space models are based on the notion that there is an unobserved true state of the system, a latent state, evolving over time that can only be observed indirectly. GP has been recently used in various other computational optimization problems including the art gallery problem [18] and the traveling sales person problem [19]. Here, we addressed the identifiability issue by extending the GP model to approximate the responses of *multiple* genes. A similar computational approach has been previously used in school exams score prediction, pollution prediction and gene expression data [20,21].

The use of linear regulation models of gene expression reduces uncertainties about the biological interpretation of model parameters, at the risk of not capturing fine subtleties in the dynamics of these highly complex interacting systems including saturation, repression and combinatorial interactions. To address this dilemma, we used GP to simulate the discrepancy between our linear model predictions and the observed experimental results. In the current study we focused on *Rap1p* and *Abf1p*, two TFs involved in regulation of gene expressions that encode for ribosomal proteins in the yeast *Saccharomyces cerevisiae*, under two related environmental stress conditions: oxidation stress and glucose-limited (starvation) growth. Ribosomes contain over 60 different protein units, most of which are present in a single copy per ribosome [22]. The expression of the multiple ribosomal genes must be tightly coordinated to ensure the production of equimolar amounts of the various ribosomal protein units. Furthermore, the production rate of these ribosomal units must respond to physiological challenges the cell is facing, in order to allow a rapid and accurate adjustment of the rate of ribosome formation to new demands and environmental stresses. Accurate characterization of the activity profiles of TFs that control the expression of various ribosomal proteins may provide important insight into system control of the various signal transduction cascades that may be involved in this tightly coordinated process.

## 2. Model

We assume that the rate of change in mRNA level ( $\dot{x}_i$ ) of target gene  $i$ , can be described by a linear differential equation [6]:

$$\dot{x}_i = \sum_{j=1}^F K_j h_j(t) + b - \mu x_i, \quad i = 1, \dots, N_{\text{exp}} \quad (1)$$

where  $h_i(t)$  is a hidden variable (i.e. state-space function) that represents the activation level (i.e. concentration of the active form) of the  $i$ th TF and  $K_j$  is a coupling constant that represents the coupling strength of mRNA transcription in response to the  $i$ th TF binding to its binding site in the DNA's promoter region. Thus, the first term represents a variable mRNA production (for  $K > 0$ ), or repression (for  $K < 0$ ) term; the second term represents a basal, constant, mRNA production rate and the third term is a degradation term with  $\mu$  being the decay rate of mRNA. Here we assume that genes that belong to the same metabolic pathway (metabolon) have the same basal and activation levels. This approximation is justified by the finding that the mRNA of these genes exhibits similar expression patterns under a given environmental stress condition [23]. Assuming  $x(t=0) = 0$ , for all genes, the solution to the first-order differential equation described by Eq. (1) is given by:

$$x_i = \sum_{j=1}^F \frac{b}{\mu} + K_j e^{-\mu t} \int_0^t h_j(\tau) e^{-\mu \tau} d\tau. \quad (2)$$

Download English Version:

<https://daneshyari.com/en/article/974400>

Download Persian Version:

<https://daneshyari.com/article/974400>

[Daneshyari.com](https://daneshyari.com)