Modelling Transcriptional Regulation using Switched Dynamical Latent Force Models

Andrés Felipe López-Lopera

Supervisor: Mauricio Alexánder Álvarez López

Universidad Tecnológica de Pereira
Engineering Faculty - Electrical Engineering Program
Master in Electrical Engineering
Research Group in Automática
Pereira, Risaralda
2015

Content

List of Figures

List of Tables

1 Acknowledgements

I want to thank all my family for the continuous support and for teaching me the importance of education. I am grateful with Ph.D. Mauricio Álvarez for guiding me during my masters degree, and for his opportune supervisions in both my education as master student and my profile as research scientist. I am also grateful with him for giving me the opportunity to work in different research projects related to my masters degree. Finally, I want to thank to the research group in Automática and the master program in electrical engineering of the Universidad Tecnológica de Pereira due academic orientation.

This work was funded by Colciencias and the Universidad Tecnológica de Pereira, under the agreement 645 of Jóvenes Investigadores under the project "Sparse latent force models for reverse engineering of multiple transcription factor". This project has also been possible thanks to the project "Human-motion Synthesis through Physically-inspired Machine Learning Models" with code 6-15-3, funded by Universidad Tecnológica de Pereira, and the project "Probabilistic spatio-temporal models based on partial differential equations for the description of the regulatory dynamics for the Bicoid protein in the Drosophila melanogaster body segmentation" funded by Colciencias (Colombian Government) and ECOS-NORD (French Government).

2 Abstract

Transcriptional regulation (TR) is a biological process in all living organisms by which cells respond to external stimuli, regulating the conversion of DNA to mRNA. The stimulus is driven by special transcription factor (TF) proteins. To control gene activities, TF can transit from inactive to active state, and bind to specific genes to activate or inhibit their transcription. This process finely tunes the amount of mRNA which is being produced through a variety of regulatory mechanisms, which is then translated into protein to act on the cellular environment. This makes key the understanding of regulatory mechanism for biological processes. However, while some gene-specific constants are relatively easy to quantify, experimental techniques to measure proteins concentrations, or to evaluate their effect on genes, are difficult and time consuming. In order to develop statistical approaches for transcription networks, statistical community has proposed several methods to infer activity levels of proteins, from time-series measurements of targets' expression levels. Current simplified and approximated methods based on ordinary differential equation (ODE) are among the approaches most over-represented for modelling TR (e.g. single input motif, SIM, simplification). However, the implementation of SIM-based models for quick time-varying behaviour of protein stimulus may not be appropriate if TR models involve only a handful of genes and one TF. To deal with this drawback, a few number of approaches have been proposed in order to outperform the representation of fast switching time instants, but computational overheads are significant due to complex inference algorithms (e.g. MCMC methods). Using the theory related to latent force models (LFM), the development of this project provide a switched dynamical hybrid model based on Gaussian processes (GPs). To deal with discontinuities in the dynamical systems (or the latent driving force), an extension of the single input motif approach is introduced, that switches between different protein concentrations, and different dynamical systems. This creates a versatile representation for transcription networks that can capture discrete changes and non-linearities in the dynamics. The proposed method is evaluated on both simulated data and real data (e.g. Escherichia coli and Yeast datasets), concluding that our framework provides a computationally efficient statistical inference module of continuous-time concentration profiles, and allows an easy estimation of the associated model parameters.

3 Introduction

Transcriptional regulation (TR) is a biological process in all living organisms by which cells respond to external stimuli, regulating the conversion of DNA to mRNA (transcription). The stimulus is driven by special proteins and transcription factor (TF) proteins. To control the gene activity, TF can transit from inactive to active state, and bind to specific genes to activate or inhibit their transcription [1]. This process finely tunes the amount of mRNA which is being produced through a variety of regulatory mechanisms [1]. The understanding of this mechanism is key for cellular processes such as biomedical and bioengineering applications. Drug design, bacterial transcription, and genetic engineering of drought-resistant crops, are some cases under study in which TR holds important biological information [1, 2].

Nowadays, micro-array technology allows the measurement of RNA abundance on a genome-wide scale. Techniques such as chromatin immunoprecipitation (ChIP) have largely unveiled the wiring of the cellular transcriptional regulatory network, identifying which genes are bound by which genes [3]. However, experimental techniques to measure active protein concentrations or to evaluate their effect on genes are difficult and time consuming [1, 2]. The knowledge of key biological constants also are required to obtain a full quantitative description of transcription networks. While some gene-specific constants are relatively easy to measure (e.g. messenger RNA decay rates and its abundance levels), it is still very hard to measure the active concentration levels of the proteins which drive the process and the sensitivity of target genes to these concentrations [3]. This makes that TR is far from being wholly understood [1, 4].

In order to develop probabilistic approaches for transcription networks, statistical community has proposed several methods to infer the activity levels of proteins from time-series measurements of the targets' expression levels [1]. There are two common schemes for inferring the protein activity. First, models which attempt to capture the simultaneous behaviour of all the proteins (also TFs), and all genes in an organism [1]. Second, models of small sub networks based on ordinary differential equations (ODEs), involving only a handful of genes and one protein (e.g. Hill function, and single input motifs - SIM) [3–5]. However, current studies suggest that SIM approaches are amongst the most over-represented for transcription networks [1, 3, 4, 6]. Barenco et al. [4] presented a parametric approach based on a time-dependent linear ODE to describe the production rate of certain genes as a function of protein concentration. The authors performed a Markov Chain Monte Carlo (MCMC) algorithm, requiring substantial computational resources to carry out the inference of concentrations [3,4]. Also, this approach limits the inference to discrete time-points where the data were collected. Later, Lawrence et al. showed that Gaussian processes (GP) provide a simple and computationally efficient method for statistical inference of continuous-time protein concentration profiles and associated model parameters [3]. Using the continuous ODE version proposed in [4], Lawrence et al. inferred the gene activity profile placing a Gaussian process (GP) prior distribution over the protein concentration. In [3], the authors concluded that GPs greatly outperform the MCMC techniques in terms of computational efficiency. Further, the proposal in [3] was extended in [6] for inferring latent chemical species in biochemical networks, evidencing promising results.

The proposals in [3, 4, 6] can be used effectively for modelling the dynamics of simpler regulatory networks, where the gene expression level is driven by only one protein. However, external stimuli (e.g. sudden oxygen starvation, or damage to the cell) can cause the quick activation of TFs, driving sudden changes in protein concentration profiles. Given the difficulty to measure experimentally the TF activities, and the necessity to capture discrete changes in dynamics, the quick adaptation of statistical models to the discontinuities in the dynamical systems (or the latent driving force) is

often key to represent the protein activities [1]. In order to cope with this type of drawbacks, TFs are often post-transcriptionally regulated via fast processes such as dimerization and phosphorylation, so TFs can be turned on or off as soon as the signal is received [5]. In this case, this is clearly a problem in discrete-time models as the piecewise constant assumption cannot be justified [1,7]. In [1], Sanguinetti et al. presented a method for continuous-time TF activity inference which specifically models quick response to stress signals. The authors modelled the latent process proposed in [4] as a Markovian stochastic dynamical process, improving the representation of the transition between two TF states. Recently, a multi-switch approach was proposed in [7] to handle an arbitrary number of transitions where the transcription rate is changed. The framework proposed by Jenkins et al. consists in a piecewise linear ODE model of RNA dynamics, which can be fitted efficiently with a reversible jump MCMC sampler for estimating the gene-specific parameters [7]. However, the computational overheads from [1, 7] are significant due to complex inference algorithms.

It is evident that modelling transcription networks requires the study of a continuous dynamical problem in which the quick time-varying behaviour of the input stimulus can be taken into account. The continuous-time method in [3] using GPs is more appealing, simpler and computationally efficient approach for inference purposes. The mistakes in protein inference due to fast switching time instants can be avoided by using a non-stationary covariance, but the computational overheads would be significant [1,8]. In 2009, Álvarez et al. introduced a novel hybrid methodology using GPs and differential equations, known as latent force models (LFM). LFM combine data driven modelling with physical models of systems [9, 10]. Here, the statistical inference of continuous quantities is possible without a discretization process, and it still being computationally efficient. Also, Álvarez et al. [11] proposed a switched dynamical LFM version for the problem of determining robot motor primitives using a second-order ODE, obtaining versatile representation for robot movements that can capture discrete changes and non-linearities in the dynamics.

The development of this project presents a switched dynamical latent force model when modelling transcriptional regulation of gene expression. To deal with discontinuities in the dynamical systems (or the latent driving force), we introduce an extension of the single input motif approach, that switches between different protein concentrations, and different dynamical systems. This creates a versatile representation for transcription networks that can capture discrete changes and non-linearities in the dynamics. We evaluate the method on both simulated data and real data (e.g. p53, Escherichia coli and Yeast datasets), concluding that our framework provides a computationally efficient statistical inference module of continuous-time concentration profiles, and allows an easy estimation of the associated model parameters.

This project is organized as follows. Section 4 shows the general objective and the specific objectives of the project. Section 5 reviews the fundamental of transcription networks as well as the mathematical framework of the proposed methodology when modelling transcriptional regulation of gene expression. In section 6, we present the materials and methods used for the development of the project, as well as the formulation of the proposed switched dynamical latent force model. Section 7 shows and discusses the results of the proposed framework with artificial data and real biological data. First, we present the results obtained in different toy experiments using the proposed switched dynamical latent force model for the first-order ODE. Next, we present the results of the model when modelling transcriptional regulation of gene expression. Finally, in section 8, we summarize the conclusions of this project, and present the possible future works.

4 Objectives

4.1 General objective

To develop a switched dynamical latent force model using Gaussian processes for modelling transcriptional regulation.

4.2 Specific objectives

1. To establish the mathematical formulation of a switched dynamical latent force model based on Gaussian processes, that allows the prediction of the solution for the first-order non-homogeneous ordinary differential equation.

2. To develop an statistical inference procedure for computing the posterior probability distribution over the Gaussian process variables, and for hyper-parameter estimation of the covariance functions.

3. To validate the performance of the proposed framework when modelling transcriptional regulation of gene expression.

5 Background

In this section we describe the fundamentals of transcription networks as well as the mathematical framework of the proposed methodology when modelling transcriptional regulation of gene expression. First, we present a brief introduction about transcription network. This part is completely based on the theory exposed by Alon in [5]. Next, we provide a detailed description of the dynamics and response time of simple gene regulation, and how to derive the single input motif (SIM) approach proposed by Lawrence et al. in [3]. This explanation involves the use of ordinary differential equations (ODE), and some biological assumptions as starting point. Finally, we present the theory of probability that supports the use of latent force models (LFM), as well as the solution to the first-order non-homogeneous ODE for R independent latent functions.

5.1 Transcription networks

5.1.1 Definition

Cells are integrated devices made of thousands types of interacting proteins, where each protein is responsible to carry out a specific task with precision. Different environmental conditions (internal or external), make a cell requires different proteins. For example, when a cell senses the presence of sugar, special proteins are produced to transport the sugar into the cell. On the other hand, when cell is damaged, repair proteins are produced. The cells therefore continuously monitor its environment, and calculate the proper amount at which each type of protein is needed.

To represent the environmental conditions, cells use special proteins, namely, transcription factors (TFs). The TFs are usually designed to transit rapidly between active and inactive molecular states, as a rate that is modulated by a specific environmental signal (input). The proteins can bind the deoxyribonucleic acid (DNA) to regulate the rate at which specific target genes are read (transcribed). The genes are transcribed into the messenger ribonucleic acid (mRNA), and then they are translated into proteins in order to act on the environment. Several different environmental states are summarized by a particular TF activity, concluding that these activities in a cell may be considered as an internal representation of the environment.

Transcription networks describe the interaction between proteins and genes, where each gene is a stretch of DNA whose sequence encodes the information needed for production of a protein. On the other hand, TF proteins are themselves encoded by genes, which are regulated by other TF, which in turn may be regulated by yet other TF, and so on. This set of interactions forms a transcription network. The transcription network describes all of the known regulatory transcription iterations in a cell [5].

Figure 1 shows the mapping between environmental signals, TFs inside the cell, and the genes that they regulate [5]. Here, the environmental signals (inputs) activate specific TF proteins, and then the activated TFs bind the DNA to change the transcription rate of specific target genes (rate at which mRNA is produced). Next, the mRNA is translated into proteins. Finally, these proteins affect the internal environment. In the transcription network, the nodes correspond to genes, and edges represent transcriptional regulation of one gene by the protein product of another gene.

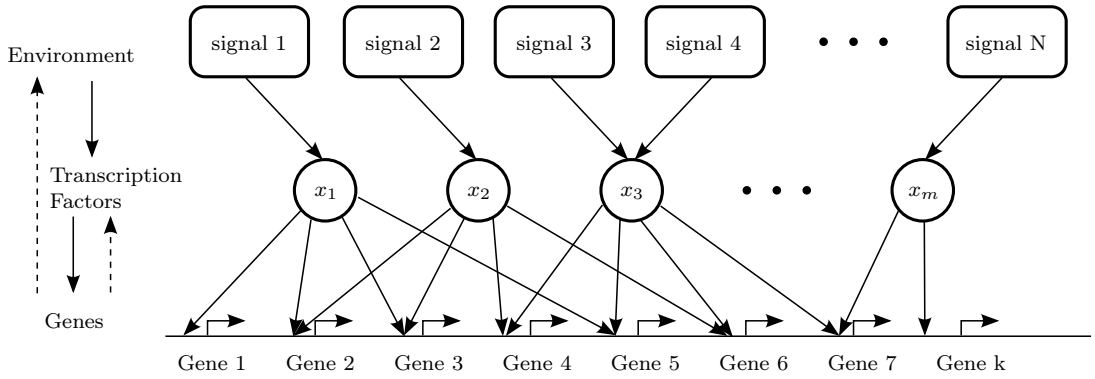


Figure 1: Transcription network scheme. Mapping between environmental signals, transcription factors inside the cell, and the genes that they regulate. This representation was proposed by Alon in [5].

5.1.2 Dynamics and response time of simple gene regulation

Previously, we presented a brief introduction about transcription network. We concluded that each edge in a transcription network corresponds to an interaction in which a TF directly controls the transcription rate of a gene. Now, let us focus on the dynamics of a single edge in the network. Consider a gene that is regulated by single regulator, with no additional inputs. This transcription interaction is described in the network by $x \rightarrow y$, which reads "transcription factor x regulates gene y". In the absence of input signal, x is inactive, and y is nor produced. When the signal S_x appears, x rapidly transits to its active form x^* and binds the promoter of the gene y. Once x becomes activated by an environmental signal S_x, the concentration y begins to change. The gene y begins to be transcribed, and the mRNA is translated, resulting in accumulation of protein y. Here, the cell begins to produce protein y at a constant rate B (units of concentration per unit time). The production of y is balanced by two processes: protein degradation and dilution. The first process specifies the destruction by specialized proteins in the cell. The second process describes the reduction in concentration due to the increase of cell volume during growth. The total degradation/dilution rate, also know as the decay rate, will be denoted as γ (units of 1/time). Finally, the change in the concentration of y is due to the difference between its production and degradation/dilution. This phenomenon is described by the following equation

$$\frac{dy(t)}{dt} = B - \gamma y(t). \tag{1}$$

At steady state, $y(t)$ reaches a constant concentration y_{st}. The steady-state concentration can be found by solving the condition $dy(t)/dt = 0$. This shows that the steady-state concentration is given by $y_{st} = B/\gamma$. From Equation (1), we see that for a higher production rate B, higher will be the protein concentration reached, y_{st}. On the other hand, a higher degradation/dilution rate γ produce a lower y_{st}. More details can be founded in [5].

5.1.3 Statistical modelling for transcriptional regulation

As we saw at the beginning of the section, transcriptional regulation (TR) is the means by which a cell regulates the conversion of DNA to mRNA (transcription), thereby orchestrating gene activity. This process is vital in all of the known regulatory transcription iterations in a cell. The regulation of the transcription rate is driven by transcription factors (TFs), and other special proteins, working together in concert to finely tune the amount of mRNA. We also show that the knowledge of a number of key biological quantities are required to model the dynamics of cellular transcriptional processes (e.g. the mRNA decay rates and the concentration levels of the TF proteins which drive the process) [3]. However, the direct prediction of active concentration levels of proteins, and the sensitivity of target genes to these concentrations, is a difficult task due the ignorance of many biochemical parameters [2].

In order to develop probabilistic approaches for transcription networks, statistical community has proposed several methods to infer the activity levels of proteins from time-series measurements of the targets' expression levels [1, 12]. Recent studies have showed that there is an underlying simplification for modelling transcriptional regulation of gene expression in which concentrations of reactants evolve continuously and differentially, making sense the use of differential equations [13]. The complexity of mechanistic models used in approaching TR and gene expression prediction, goes from a linear ordinary differential equation (ODE), to a non-linear partial differential equation (PDE) [5, 12]. The ODE given in the Equation (1) has been used since the early days of molecular biology [5], evidencing good agreement with high-resolution dynamics experiments done under conditions of protein activation during exponential growth of bacteria [5, 14]. Moreover, Barenco et al. [4] have proposed a more generalized approach based on a first-order non-homogeneous ODE to model transcription networks. Here, the transcription of a number of genes is driven by the concentration of a single protein. This framework is known as single input motif (SIM). In this approach, Barenco et al. limit the inference process to discrete time-points where the data were collected. Later, Lawrence et al. [3] adapted the approach proposed by Barenco et al. for modelling multi-ouput transcription networks using a set of D output functions $\{y_d(t)\}_{d=1}^{D}$ of a series of D coupled first-order ODEs. The multi-output SIM approach proposed in [3] can also be extended to R protein concentrations and continuous genes profiles [3, 9], obtaining the following expression

$$\frac{dy_d(t)}{dt} + \gamma_d y_d(t) = B_d + \sum_{r=1}^{R} S_{d,r} u_r(t), \tag{2}$$

where $u_r(t)$ represents the concentration of the r-th protein (which is difficult to measure directly), $y_d(t)$ are the mRNA abundance levels for different genes, $B_d \in \mathbb{R}^+$ and $\gamma_d \in \mathbb{R}^+$ are the basal transcription and the decay rates of d-th gene, respectively. The terms $S_{d,r}$ represent the sensitivity of the gene d to the protein concentration r. The Equation (2) assumes that the transcript is degraded proportionally to its concentration, with the degradation rate γ_d. The production term $B_d + \sum_{r=1}^{R} S_{d,r} u_r(t)$ comprises a basal transcription rate B_d, which may be increased proportionally by the protein activities $\{u_r(t)\}_{r=1}^{R}$ [4]. The SIM networks can be seen as an overly simple system, but according to the state-of-art, they are amongst the frameworks most over-represented in biological and bacterial transcription networks [1, 3–6].

5.2 Gaussian processes

A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP of a real function $f(\mathbf{x})$ is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$, which are given by [15]

$$\begin{aligned} m(\mathbf{x}) =& \mathbb{E}\{f(\mathbf{x})\} \\ k(\mathbf{x}, \mathbf{x}') =& \mathbb{E}\{[f(\mathbf{x}) - m(\mathbf{x})][f(\mathbf{x}') - m(\mathbf{x}')]\}, \end{aligned}$$

where $f(\mathbf{x})$ is a random process evaluated in $\mathbf{x}$. Here, the operator $\mathbb{E}\{\cdot\}$ defines the expected value of a random variable, which is intuitively the long-run average value of repetitions of the experiment it represents. We can write the GP over the function $f(\mathbf{x})$ as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')).$$

The definition "A Gaussian process is defined as a collection of random variables", automatically implies the marginalization property. This property means that if the GP specifies $(y_1, y_2) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,[1] then it must also specify $y_1 \sim \mathcal{N}(\mu_1, \Sigma_{1,1})$ where $\Sigma_{1,1}$ is the relevant submatrix of $\boldsymbol{\Sigma}$. In other words, examination of a larger set of variables does not change the distribution of the smaller set [15].

5.3 Latent force models using Gaussian processes

Latent force models (LFM) based on Gaussian process (GP) were first introduced by Álvarez et al. in [9]. The general framework of LFM is to combine a mechanistic model with a probabilistic prior over some latent function [9]. In [10], a set of coupled first and second order ODEs were introduced for transcriptional regulation and motion capture data, respectively. Also, LFM can also be applied in the context of partial differential equations (PDE) in order to recover multi-output Gaussian processes [9]. Diffusion in heavy metal pollutants [10], gap-gene network of Drosophila Melanogaster [16], and electric propagation in deep brain stimulation are some examples where LFM is applied for PDE [17].

Considering the mechanistic model described in Equation (2) with $D = 1$ and $R = 1$, the LFM assume that the input driving force $u(t)$ is an unknown function (later named as latent function) which follows a GP prior, with zero mean and covariance function $k_{u,u}(t, t')$, given by

$$u(t) \sim \mathcal{GP}(0, k_{u,u}(t, t')). \tag{3}$$

Assuming linearity in the proposed mechanistic model, the solution $y(t)$ also corresponds to a GP prior with mean function $m(t)$, and covariance function $k_{y,y}(t, t')$, this is

$$y(t) \sim \mathcal{GP}(m(t), k_{y,y}(t, t')). \tag{4}$$

We will see later in the following section how to get the expressions for $m(t)$ and $k_{y,y}(t, t')$. Furthermore, a covariance function $k_{y,u}(t, t')$ between $y(t)$ and $u(t)$ can also be computed. Also,

[1] Here, the variables $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspond to the mean vector and the covariance matrix of the Gaussian distribution, respectively.

we can assume that the latent force $u(t)$ together with the solution $y(t)$ have a joint multivariate Gaussian distribution given by

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{y}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \boldsymbol{m}\end{bmatrix}, \begin{bmatrix}K_{u,u} & K_{y,u}^\top\\ K_{y,u} & K_{y,y}\end{bmatrix}\right),$$

where the vectors $\mathbf{u}$, $\mathbf{y}$ and $\boldsymbol{m}$ represent the functions $u(t)$, $y(t)$ and $m(t)$ at particular time instants, respectively. The covariance matrices $K_{u,u}$, $K_{y,u}$ and $K_{y,y}$ are covariance matrices computed from functions $k_{u,u}(t,t')$, $k_{y,u}(t,t')$, and $k_{y,y}(t,t')$ at particular time instants t and t'. The covariance matrices $K_{u,u}$ and $K_{y,y}$ are obtained with the respective covariance functions mentioned previously in Equation (3) and Equation (4), respectively. The matrix covariance $K_{y,u}$ is obtained using the cross-covariance kernel between the output $y(t)$ and the latent function $u(t)$. We are interested in getting the posterior distribution over the solution function $y(t)$, given a specific latent force $u(t)$. An useful property of the multivariate Gaussian distribution is that if two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is also Gaussian distribution [15, 18]. Using this property we can get the posterior distribution [19]

$$y(t)|u(t) \sim \mathcal{N}\left(m(t) + K_{y,u}K_{u,u}^{-1}u(t), K_{y,y} - K_{y,u}K_{u,u}^{-1}K_{y,u}^\top\right). \tag{5}$$

The conditional Gaussian distribution properties also allow to compute the inverse problem, i.e. to get the posterior distribution over the latent force $u(t)$, given an specific solution $y(t)$,

$$u(t)|y(t) \sim \mathcal{N}\left(K_{y,u}^\top K_{y,y}^{-1}[y(t) - m(t)], K_{u,u} - K_{y,u}^\top K_{y,y}^{-1}K_{y,u}\right). \tag{6}$$

5.4 Latent force model approach for SIM networks

As described in section 5.3, the single input motif (SIM) approximation proposed by [4] can be seen as an LFM where the mechanistic model is governed by a first-order non-homogeneous ODE given in Equation (2) with $R = 1$. For transcription networks, the mRNA abundance level $y_d(t)$ (outputs) are driven by a single protein concentration $u(t)$ (latent forces). In [3], Lawrence et al. introduce an approach based on GP for a SIM network. They assume a GP prior over the protein concentrations $u(t) \sim \mathcal{GP}(0, k_{u,u}(t,t'))$ with covariance function $k_{u,u}(t,t')$, and then they compute the covariance functions described in section 5.3. In order to compute the covariance between the mRNA abundance levels $y_d(t)$, we need to compute the solution of the ODE described in Equation (2) for $R = 1$. According to the appendix A, the solution of the first-order ODE follows

$$y_d(t) = [1 - c_d(t)]\frac{B_d}{\gamma_d} + c_d(t)y_d(0) + f_d(t,u), \tag{7}$$

with

$$c_d(t) = \exp\{-\gamma_d t\}, \qquad f_d(t,u) = S_d c_d(t)\int_0^t u(\tau)\exp\{\gamma_d\tau\}\,d\tau.$$

Here, S_d represents the sensitivity of the gene d to the protein concentration $u(t)$. We can note that $f_d(t,u)$ has an implicit dependence on the latent force $u(t)$. The uncertainty in this model is due to the fact that the latent force $u(t)$ and the initial condition $y_d(0)$ are not known. Here, the GP for the

output is given by $y_d(t) \sim \mathcal{GP}(m_d(t), k_{y_d,y_{d'}}(t,t'))$, with the mean function $m_d(t)$ and the covariance function $k_{y_d,y_{d'}}(t,t')$. We can also assume that initial conditions, $\mathbf{y}_{IC} = [y_1(0), y_2(0), \cdots, y_D(0)]^\top$, are independent of $u(t)$ and distributed as a zero mean Gaussian with covariance K_{IC}. Because the latent function $u(t)$ is a zero mean GP prior, and the initial conditions are distributed as a zero mean Gaussian, the mean function $m_d(t)$ is given by the constant term $[1 - c_d(t)]B_d/\gamma_d$. Finally, the covariance function between any two output functions, d and d' at any two times, t and t', is given by

$$k_{y_d,y_{d'}}(t,t') = c_d(t)c_{d'}(t')\sigma_{y_d,y_{d'}} + k_{f_d,f_{d'}}(t,t'), \tag{8}$$

where $\sigma_{y_d,y_{d'}}$ are entries of the covariance matrix K_{IC}, and

$$k_{f_d,f_{d'}}(t,t') = S_d S_{d'} c_d(t) c_{d'}(t') \int_0^t \exp\{\gamma_d \tau\} \int_0^{t'} \exp\left\{\gamma_{d'}\tau'\right\} k_{u,u}(\tau,\tau') d\tau' d\tau.$$

The covariance function $k_{f_d,f_{d'}}(t,t')$ depends on the covariance function of the latent force $k_{u,u}(t,t')$. If we assume that $k_{u,u}(t,t')$ follows a radial basis function (RBF) covariance given by

$$k_{u,u}(t,t') = \exp\left\{ -\frac{(t-t')^2}{\ell^2} \right\},$$

where ℓ controls the width of the basis functions, then $k_{f_d,f_{d'}}(t,t')$ can be computed analytically [3,9]. The resulting covariance $k_{f_d,f_{d'}}(t,t')$ follows

$$k_{f_d,f_{d'}}(t,t') = \frac{S_d S_{d'} \ell \sqrt{\pi}}{2} k^{(1)}_{f_d,f_{d'}}(t,t'),$$

where

$$k^{(1)}_{f_d,f_{d'}}(t,t') = \hat{h}(\gamma_{d'},\gamma_d,t,t') + \hat{h}(\gamma_d,\gamma_{d'},t',t),$$

$$\hat{h}(\gamma_{d'},\gamma_d,t,t') = \frac{1}{\gamma_d + \gamma_{d'}} \Big[\Upsilon(\gamma_{d'},t',t) - \exp\{-\gamma_d t\}\, \Upsilon(\gamma_{d'},t',0) \Big],$$

and

$$\Upsilon(\gamma_{d'},t',t) = \exp\{\nu_{d'}^2\} \exp\{-\gamma_{d'}(t'-t)\} \left[\operatorname{erf}\left\{ \frac{t'-t}{\ell} - \nu_{d'} \right\} + \operatorname{erf}\left\{ \frac{t}{\ell} + \nu_{d'} \right\} \right],$$

with $\nu_{d'} = \ell\gamma_{d'}/2$. On the other hand, the cross-covariance between the d-th output and latent force at any two times, t and t', is given by

$$k_{y_d,u}(t,t') = S_d c_d(t) \int_0^t \exp\{\gamma_d \tau\}\, k_{u,u}(\tau,t') d\tau.$$

The resulting cross-covariance can also be computed analytically, obtaining

$$k_{y_d,u}(t,t') = \frac{S_d \ell \sqrt{\pi}}{2} \Upsilon(\gamma_d,t,t').$$

Finally, we can build the GP for the outputs, $y_d(t) \sim \mathcal{GP}(m_d(t), k_{y_d,y_{d'}}(t,t'))$ with the mean function $m_d(t) = [1 - c_d(t)]B_d/\gamma_d$, and covariance function given in the Equation (8). Appendix B describes all the procedure for computing the covariances of the SIM framework.

5.5 Latent force model approach for SIM networks with R independent protein concentrations

The SIM approximation proposed in [3] can be generalized for R independent protein concentrations according to the methodology proposed by Álvarez et al. in [9]. Details related to the general solution of the mechanistic model proposed in Equation (2) is described in the appendix A. For a set of R independent latent forces $\{u_r(t)\}_{r=1}^{R}$, the set of D outputs $\{y_d(t)\}_{d=1}^{D}$ is represented by a Gaussian process with covariance function given by

$$k_{y_d,y_{d'}}(t,t') = c_d(t)c_{d'}(t')\sigma_{y_d,y_{d'}} + \sum_{r=1}^{R} k_{f_{d,r},f_{d',r}}(t,t'), \tag{9}$$

with

$$k_{f_{d,r},f_{d',r}} = \frac{S_{d,r}S_{d',r}\ell_r\sqrt{\pi}}{2} k^{(r)}_{f_d,f_{d'}}(t,t').$$

Here, $S_{d,r}$ represents the sensitivity of gene d to r-th concentration, and ℓ_r controls the width of the RBF covariance which describes the GP for latent force r. The covariance $k^{(r)}_{f_d,f_{d'}}(t,t')$ describes the covariance between the functions $f_d(t)$ and $f_{d'}(t')$ under the effect of the r-th latent force, and it follows the same structure obtained in section 5.4 for the SIM with $R = 1$. The resulting cross-covariance between outputs and latent forces also can be computed analytically, obtaining

$$k_{y_d,u}(t,t') = \sum_{r=1}^{R} k_{y_d,u_r}(t,t') = \sum_{r=1}^{R} \frac{S_{d,r}\ell_r\sqrt{\pi}}{2}\Upsilon^r(\gamma_d,t,t').$$

Similar to the SIM approach, we can build the GP for the outputs, $y_d(t) \sim \mathcal{GP}(m_d(t), k_{y_d,y_{d'}}(t,t'))$ with $m_d(t) = [1 - c_d(t)]B_d/\gamma_d$. More details about the covariance functions can be found in the appendix C.

5.6 Statistical modelling for quick time-varying behaviour of protein activity: a review

The methodologies proposed in [3,4,9] can be used effectively for modelling the dynamics of simpler regulatory networks. However, as we saw at the beginning of section 5.1, changes in the external environment (e.g. sudden oxygen starvation), or also changes in the internal cellular condition (e.g. damage to the cell), can cause the quick activation of TFs. The activation of TFs in turn can lead to sudden changes in protein concentration profiles [1, 5, 20]. Given the difficulty to measure experimentally the TF activities, and the necessity to capture discrete changes in dynamics, the quick adaptation of statistical models to the discontinuities in the dynamical systems (or the latent driving force) is often key to represent the protein activities [1]. In order to cope with this type of drawbacks, TFs are often post-transcriptionally regulated via fast processes such as dimerization and phosphorylation, so TFs can be turned on or off as soon as the environmental signal is received [5]. In this case, this is clearly a problem in discrete-time models as the piecewise

constant assumption cannot be justified [1,7]. In 2009, Chechik et al. introduced a flexible parametric model based on the product of two sigmoid functions which can capture two transitions in the response dynamics of gene expression to environmental perturbations [21]. This model was applied to gene expression time courses in Saccharomyces cerevisiae (Yeast) after diverse environmental perturbations, evidencing that their model constitutes an improvement over other general functional forms (e.g. polynomials) [21]. However, although the magnitude and the timing of the response arise as meaningful parameters, their model is not directly connected to mechanism and is still not general enough to explain a wider range of possible dynamic pattern observed in gene expression, including oscillations [7]. Also in 2009, Sanguinetti et al. presented a method for continuous-time TF activity inference which specifically models quick response to stress signals [1]. The authors modelled the latent process proposed in [4] as a Markovian stochastic dynamical process, improving the representation of the transition between two TF states. This approach was implemented to describe the microaerobic shift in Escherichia coli (E.coli), where the sudden oxygen starvation in the environment provokes the necessity of routinely adaptation in the bacterium. The proposal in [1] was extended further in 2010 by Opper et al. to simultaneously infer the activities of multiple interacting TFs, and it was applied to describe the control of ribosomal protein production in Yeast metabolic cycle [20]. More recently, in 2013, a multi-switch approach was proposed by Jenkins et al. [7] to handle an arbitrary number of transitions where the transcription rate is changed. The framework proposed consists in a piecewise linear ODE model of RNA dynamics, which can be fitted efficiently with a reversible jump MCMC sampler for estimating the gene-specific parameters. In [7], the authors focus on modelling the wild-type behaviour of a selection of 200 circadian genes of the model plant Arabidopsis thaliana. Jenkins et al. showed that their approach allows them to investigate whether genes with similar switch event times also have correlated promoter motifs. Results of both proposal of Sanguinetti et al. and Jenkins et al. support the idea that using a mechanistic model to identify transcriptional switch points is likely to strongly contribute to efforts in elucidating and understanding key biological processes (e.g. transcription and degradation). However, the computational overheads from [1,7] are significant due to complex inference algorithms.

It is evident that modelling transcription networks requires the study of a continuous dynamical problem in which the quick time-varying behaviour of the input stimulus can be taken into account. The continuous-time method in [3] using GPs is more appealing, simpler and computationally efficient approach for inference purposes. In 2009, Álvarez et al. introduced a novel hybrid methodology using GPs and differential equations, known as latent force models (LFM). LFM combine data driven modelling with physical models of systems [9,10]. Here, the statistical inference of continuous quantities is possible without a discretization process, and it still being computationally efficient. Álvarez et al. [11] also proposed a switched dynamical LFM version for the problem of determining robot motor primitives using a second-order ODE, obtaining versatile representation for robot movements that can capture discrete changes and non-linearities in the dynamics.

Inspired by the methodology proposed by Álvarez et al. in [11], the development of this project presents a switched dynamical LFM when modelling transcriptional regulation of gene expression. To deal with discontinuities in the dynamical systems (or the latent driving force), we introduce an extension of the SIM approach, that switches between different protein concentrations, and different dynamical systems. This creates a versatile representation for transcription networks that can capture discrete changes and non-linearities in the dynamics. In the next section we introduce the mathematical formulation of a multi-output switched dynamical LFM based on Gaussian processes, that allows the prediction of the solution for the first-order non-homogeneous ODE. We also describe statistical inference procedure for computing the posterior probability distribution over the Gaussian process variables, and for hyper-parameter estimation of the covariance functions.

6 Materials and Methods

In this section we present the materials and methods used for the development of the project, as well as the formulation of the proposed switched dynamical latent force model for modelling transcriptional regulation. First, we define the assumptions which are taken into account for the definition of the model. Next, we present the mathematical formulation of the proposed framework when modelling transcriptional regulation of gene expression. We also compute the covariance function for the outputs, and the covariance function between the outputs and latent forces as we did for the SIM approach in the background. Finally, we describe the dataset and the procedure used in this work.

6.1 Switched dynamical LFM approach for SIM networks

In this section, we will consider switching the system between different latent forces for the mechanistic model proposed in Equation (2). This allows us to change the dynamical system and the driving force for each segment. By constraining the displacement at each switching time to be the same, the output functions remain continuous. This approach was first introduced by Álvarez et al. for a second-order non-homogeneous ODE in order to determine robot motor primitives [11].

6.1.1 Definition of the model

We assume that the input space is divided in a series of non-overlapping intervals $[t_{q-1}, t_q]_{q=1}^{Q}$. During each interval, only one force $u_{q-1}(t)$ out of Q forces is active, this is, there are $\{u_{q-1}(t)\}_{q=1}^{Q}$ forces. The force $u_{q-1}(t)$ is activated after time t_{q-1} (switched on) and deactivated (switched off) after time t_q [11]. Figure 2 shows a cartoon representation of output $z_d(t)$ switching its behaviour between points t_0, t_1, t_2 and t_3.[2] For each interval (t_{q-1}, t_q), only the latent force $u_{q-1}(t)$ is active.

We can use the basic SIM model introduced in section 5.4 to describe the contribution to the output due to the sequential activation of these forces. The constant $m_d(t) = [1 - c_d(t)]\frac{B_d}{\gamma_d}$ will be included at the end of the model such as the mean function of the GP for $z_d(t)$ in each interval. A particular output $z_d(t)$ at a particular time instant t, in the interval (t_{q-1}, t_q), is expressed as

$$z_d(t, t_{q-1}, t_q) = p_d(t, t_{q-1}, t_q, u_{q-1}), \qquad \text{for} \qquad 1 \leq d \leq D,$$

where $p_d(t, t_{q-1}, t_q, u_{q-1})$ uses the model in Equation (7)

$$p_d(t, t_{q-1}, t_q, u_{q-1}) = z_d(t)|_{t_q-1} = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1}). \tag{10}$$

Notice that there are as many intervals $[t_{q-1}, t_q]_{q=1}^{Q}$ as latent forces $\{u_{q-1}(t)\}_{q=1}^{Q}$. The previous expression is assumed to be valid for describing the output only inside the interval (t_{q-1}, t_q). Here, $f_d(t, t_{q-1}, t_q, u_{q-1})$ is a function of four arguments: the first argument, t, refers to the independent variable; the second argument, t_{q-1}, and third argument t_q specify the lower and upper limits of the

[2] In our proposal, the outputs will be denoted as $z_d(t)$ to distinguish them from the outputs $y_d(t)$ of the SIM framework.

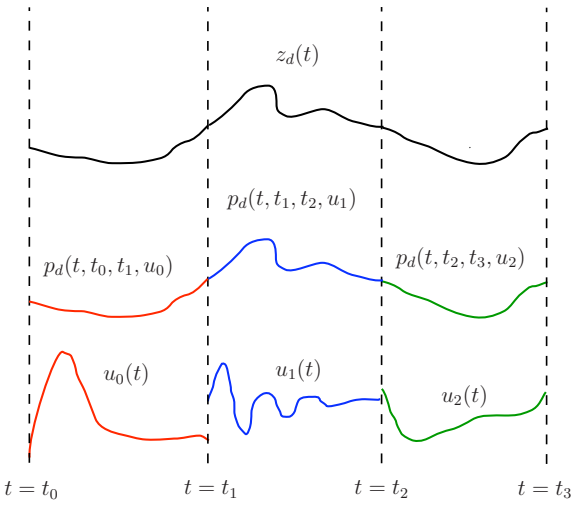


Figure 2: Cartoon representation for the output of a switched dynamical LFM proposed by Álvarez et al. in [11].

time interval to be analyzed, u_{q-1}, specifies the latent force acting in this interval. Additionally, we define a similar function for the velocity $\dot{z}_d(t)$ (which we use later for inference purpose) as

$$\dot{z}_d(t, t_{q-1}, t_q) = \xi_d(t, t_{q-1}, t_q, u_{q-1}), \qquad \text{for} \qquad 1 \leq d \leq D,$$

where

$$\xi_d(t, t_{q-1}, t_q, u_{q-1}) = v_d(t)|_{t_q-1} = g_d(t - t_{q-1}) z_d(t_{q-1}) + m_d(t, t_{q-1}, t_q, u_{q-1}),$$

with

$$m_d(t, t_{q-1}, t_q, u_{q-1}) = \frac{df_d(t, u)}{dt} = S_d \frac{d}{dt} \left(c_d(t) \int_0^t u(\tau) \exp\{\gamma_d \tau\} \, d\tau \right),$$
$$g_d(t) = \frac{dc_d(t)}{dt} = -\gamma_d \exp\{-\gamma_d t\}.$$

Given the parameters $\boldsymbol{\theta} = \{\{\gamma_d, S_d\}_{d=1}^D, \{\ell_{q-1}\}_{q=1}^Q\}$, the uncertainty in the outputs is induced by the prior over the initial conditions $z_d(t_{q-1})$ for all values of t_{q-1}, and the prior over the latent force $u_{q-1}(t)$ that is active during (t_{q-1}, t_q). We place independent GP priors over each of these latent forces $u_{q-1}(t)$, assuming independence between them.

For initial conditions $z_d(t_{q-1})$, we could assume that they are either parameters to be estimated or random variables with uncertainty governed by independent Gaussian distributions with covariance matrices K_{IC}^q. However, for the type of applications we will consider (modelling transcriptional regulation), the outputs should be continuous across the switching points. We therefore assume that the uncertainty about the initial conditions for the interval q, $z_d(t_{q-1})$, is proscribed by the GP that describes the outputs $z_d(t)$ in the previous interval $q - 1$. Finally, we assume $z_d(t_{q-1})$ are Gaussian-distributed with mean value given by $p(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2})$ and covariances $k_{z_d, z_{d'}}(t_{q-1}, t_{q'-1}) = \text{cov}\{p(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), p(t_{q'-1}, t_{q'-2}, t_{q'-1}, u_{q'-2})\}$.

In the context of transcription networks, our model assumes that the gene activity in each non-overlapping interval is governed by a single protein concentration, obtaining a switched dynamical version of the single input motif approach proposed in [3].

6.1.2 Covariance for the outputs

The derivation of the covariance function for the switching model is rather involved. For continuous output signals, we must take into account constraints at each switching time instant. This causes initial conditions for each interval to be dependent on final conditions for the previous interval and induces correlations across the intervals. This effort is worthwhile though as the resulting model is very flexible, and can take advantage of the switching dynamics to represent a range of signals [11].

In particular, we are interested in the computation of the covariance functions for the outputs, and between the outputs and latent forces, as well as we did for the SIM framework described in section 5.4. We need to compute the covariance $\text{cov}\{z_d(t, t_{q-1}, t_q), z_{d'}(t', t_{q'-1}, t_{q'})\}$ for $z_d(t, t_{q-1}, t_q)$ in time interval (t_{q-1}, t_q), and $z_{d'}(t', t'_{q'-1}, t_{q'})$ in time interval $(t_{q'-1}, t_{q'})$. Our model assumes independence between the latent forces $u_q(t)$, and assumes independence between the initial conditions $\mathbf{z}_{IC}$ and the latent forces $u_q(t)$. For computing the covariance, we have to analyze three regimes: $q > q'$, $q = q'$ and $q < q'$. In this section, we compute the first two cases, $q > q'$ and $q = q'$. The solution for $q < q'$ is given by $q > q'$, where we have to change the roles between q and q'.

Covariance for the interval (t_{q-1}, t_q)

Let $z_d(t) = p_d(t, t_{q-1}, t_q, u_{q-1}) = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})$, and $q = q'$, the covariance is given by

$$\begin{aligned}
\text{cov}\{z_d(t), z_{d'}(t')\} =& \text{cov}\{p_d(t, t_{q-1}, t_q, u_{q-1}), p_d(t', t_{q-1}, t_q, u_{q-1})\} \\
=& \mathbb{E}\{[c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})] \\
& \times [c_{d'}(t' - t_{q-1})z_{d'}(t_{q-1}) + f_{d'}(t', t_{q-1}, t_q, u_{q-1})]\} \\
=& c_d(t - t_{q-1})c_{d'}(t' - t_{q-1}) \,\text{cov}\{z_d(t_{q-1}), z_{d'}(t_{q-1})\} \\
& + c_d(t - t_{q-1}) \,\text{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\} \\
& + c_{d'}(t' - t_{q-1}) \,\text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), z_{d'}(t_{q-1})\} \\
& + \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\},
\end{aligned}$$

where $\text{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\}$ and $\text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), z_{d'}(t_{q-1})\}$ are zero, assuming independence between the initial conditions $\mathbf{z}_{IC}^q$ and the latent forces $u_{q-1}(t)$. Finally, we obtain

$$\begin{aligned}
\text{cov}\{z_d(t), z_{d'}(t')\} =& c_d(t - t_{q-1})c_{d'}(t' - t_{q-1}) \,\text{cov}\{z_d(t_{q-1}), z_{d'}(t_{q-1})\} \\
& + \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\} \\
=& c_d(t - t_{q-1})c_{d'}(t' - t_{q-1})k_{z_d, z_{d'}}(t_{q-1}, t_{q-1}) + k^{(q)}_{f_d, f_{d'}}(t, t'),
\end{aligned}$$

where the covariance $k_{z_d, z_{d'}}(t_{q-1}, t_{q-1}) = \text{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), p_{d'}(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2})\}$ (considering the outputs should be continuous across the switching points), and the covariance $k^{(q)}_{f_d, f_{d'}}(t, t') = \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\}$.

Covariance for the interval (t_{q-1}, t_q) and $(t_{q'-1}, t_{q'})$

When $q > q'$, we have to take into account the correlation between the initial conditions $z_d(t_{q-1})$ and the latent force $u_{q'-1}(t')$. This correlation appears due to the contribution of $u_{q'-1}(t')$ for generating the initial conditions, $z_d(t_{q-1})$. Let $z_d(t) = p_d(t, t_{q-1}, t_q, u_{q-1}) = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})$, and $q > q'$, the covariance is given by

$$\begin{aligned}
\operatorname{cov}\{z_d(t), z_{d'}(t')\} =& \operatorname{cov}\{p_d(t, t_{q-1}, t_q, u_{q-1}), p_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
=& \mathbb{E}\{[c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})] \\
& \times [c_{d'}(t' - t_{q'-1})z_{d'}(t_{q'-1}) + f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})]\} \\
=& c_d(t - t_{q-1})c_{d'}(t' - t_{q'-1}) \operatorname{cov}\{z_d(t_{q-1}), z_{d'}(t_{q'-1})\} \\
& + c_d(t - t_{q-1}) \operatorname{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
& + c_{d'}(t' - t_{q'-1}) \operatorname{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), z_{d'}(t_{q'-1})\} \\
& + \operatorname{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\},
\end{aligned}$$

where $\operatorname{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = 0$, because there is not correlation between the forces u_{q-1} and $u_{q'-1}$. Also, the covariance $\operatorname{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), z_{d'}(t_{q'-1})\}$ is zero, since $q > q'$, there is not correlation between force u_{q-1} and any force u_{k-1}, for $k \leq q' - 2$. We can then rewrite the above expression as

$$\begin{aligned}
\operatorname{cov}\{z_d(t), z_{d'}(t')\} =& c_d(t - t_{q-1})c_{d'}(t' - t_{q'-1})k_{z_d, z_{d'}}(t_{q-1}, t_{q'-1}) \\
& + c_d(t - t_{q-1}) \operatorname{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}. \qquad (11)
\end{aligned}$$

According to the Equation (10), the term $k_{z_d, z_{d'}}(t_{q-1}, t_{q'-1})$ is given by the covariance of previous interval and is equal to $\operatorname{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), p_d(t_{q'-1}, t_{q'-2}, t_{q'-1}, u_{q'-2})\}$. Now, we have to compute the term $\operatorname{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$. This term is equal to

$$\begin{aligned}
=& \operatorname{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
=& \mathbb{E}\{[c_d(t_{q-1} - t_{q-2})z_d(t_{q-2}) + f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2})] f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
=& c_d(t_{q-1} - t_{q-2}) \underbrace{\operatorname{cov}\{z_d(t_{q-2}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}}_{A} \\
& + \operatorname{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}.
\end{aligned}$$

The term $\operatorname{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ is only different from zero for $q = q'+1$ and it would reduce to $k_{f_d, f_{d'}}^{(q'-1)}(t_{q-1}, t')$. The term A, if $q \leq q' + 1$, is equal to zero because there is not correlation between the force and the initial condition. For $q > q' + 1$, the term in A is equal to

$$\begin{aligned}
=& \operatorname{cov}\{p_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
=& \mathbb{E}\{[c_d(t_{q-2} - t_{q-3})z_d(t_{q-3}) + f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3})] f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} \\
=& c_d(t_{q-2} - t_{q-3}) \underbrace{\operatorname{cov}\{z_d(t_{q-3}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}}_{A'} \\
& + \operatorname{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}.
\end{aligned}$$

Here, the covariance $\text{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ is different to zero for $q = q' + 2$, and it can be reduced to $k_{f_d,f_{d'}}^{(q'-1)}(t_{q-2}, t')$. Term A' follows the same form that term A. If, $q > q' + 2$, the recursion is repeated until the most inner term in $\text{cov}\{z_d(t_{q-n}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = \text{cov}\{p_d(t_{q-n}, t_{q-n-1}, t_{q-n}, u_{q-n-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ is such that $q = q' + n$. If $q = q' + n$, we obtain the following recursive expression

$$\text{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = \left[\prod_{i=1}^{n-1} c_d(t_{q-i} - t_{q-i-1})\right] k_{f_d,f_{d'}}^{(q'-1)}(t_{q-n}, t'). \tag{12}$$

6.1.3 Covariances between outputs and latent functions

For inference purposes, we also need the cross-covariances between the outputs $z_d(t, t_{q-1}, t_q)$ and the latent force $u_{q'-1}(t)$. If $q < q'$, then this covariance is zero. We are left with the cases $q = q'$ and $q > q'$.

Covariance between $z_d(t, t_{q-1}, t_q)$ and $u_{q'-1}(t')$, with $q = q'$

Let $z_d(t) = p_d(t, t_{q-1}, t_q, u_{q-1}) = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})$, and $q = q'$, the covariance is given by

$$\begin{aligned}
\text{cov}\{z_d(t, t_{q-1}, t_q), u_{q-1}(t')\} &= \text{cov}\{p_d(t, t_{q-1}, t_q, u_{q-1}), u_{q-1}(t')\} \\
&= \mathbb{E}\{[c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})]u_{q-1}(t')\} \\
&= c_d(t - t_{q-1})\,\text{cov}\{z_d(t_{q-1}), u_{q-1}(t')\} + \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), u_{q-1}(t')\},
\end{aligned}$$

where $\text{cov}\{z_d(t_{q-1}), u_{q-1}(t')\} = 0$ because there is not correlation between the initial condition and the latent force.

Covariances between $z_d(t, t_{q-1}, t_q)$ and $u_{q'-1}(t')$, with $q > q'$

Let $z_d(t) = p_d(t, t_{q-1}, t_q, u_{q-1}) = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})$, and $q > q'$, the covariance is given by

$$\begin{aligned}
\text{cov}\{z_d(t, t_{q-1}, t_q), u_{q'-1}(t')\} =& \text{cov}\{p_d(t, t_{q-1}, t_q, u_{q-1}), u_{q'-1}(t')\} \\
=& \mathbb{E}\{[c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1})]u_{q'-1}(t')\} \\
=& c_d(t - t_{q-1})\,\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} \\
&+ \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), u_{q'-1}(t')\},
\end{aligned}$$

where $\text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), u_{q'-1}(t')\} = 0$ for q strictly greater than q'. Now, we have to compute $\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\}$, obtaining

$$\begin{aligned}\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} &= \text{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), u_{q'-1}(t')\} \\ &= \mathbb{E}\{[c_d(t_{q-1} - t_{q-2})z_d(t_{q-2}) + f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2})]u_{q'-1}(t')\} \\ &= c_d(t_{q-1} - t_{q-2})\underbrace{\text{cov}\{z_d(t_{q-2}), u_{q'-1}(t')\}}_{B} \\ &\quad + \text{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), u_{q'-1}(t')\}.\end{aligned}$$

The term $\text{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), u_{q'-1}(t')\}$ is different from zero for $q = q' + 1$, and follows $k^{(q'-1)}_{f_d, u_{q'-1}}(t_{q-1}, t')$. The term B, if $q \leq q' + 1$, is zero because there is not dependency between forces. For $q > q' + 1$, the term in B follows

$$\begin{aligned}\text{cov}\{z_d(t_{q-2}), u_{q'-1}(t')\} &= \text{cov}\{p_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), u_{q'-1}(t')\} \\ &= \mathbb{E}\{[c_d(t_{q-2} - t_{q-3})z_d(t_{q-3}) + f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3})]u_{q'-1}(t')\} \\ &= c_d(t_{q-2} - t_{q-3})\underbrace{\text{cov}\{z_d(t_{q-3}), u_{q'-1}(t')\}}_{B'} \\ &\quad + \text{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), u_{q'-1}(t')\}.\end{aligned}$$

The term $\text{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), u_{q'-1}(t')\}$ is different from zero for $q = q' + 2$, and follows $k^{(q'-1)}_{f_d, u_{q'-1}}(t_{q-2}, t')$. The term B' follows the same recursive form that the term B. If $q > q' + 2$, the recursion is repeated until the most inner term in $\text{cov}\{z_d(t_{q-n}), u_{q'-1}(t')\} = \text{cov}\{p_d(t_{q-n}, t_{q-n-1}, t_{q-n}, u_{q-n-1}), u_{q'-1}(t')\}$ is such that $q = q' + n$. For $q = q' + n$, we obtain

$$\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} = \left[\prod_{i=1}^{n-1} c_d(t_{q-i} - t_{q-i-1})\right] k^{(q'-1)}_{f_d, u_{q'-1}}(t_{q-n}, t'). \tag{13}$$

6.2 Hyper-parameter estimation of the covariance functions

Given the number of outputs D and the number of intervals Q, we estimate the parameters $\boldsymbol{\theta} = \{\{\gamma_d, S_d\}_{d=1}^D, \{\ell_{q-1}\}_{q=1}^Q, \{t_q\}_{q=1}^{Q-1}\}$ by maximizing the marginal-likelihood of the joint Gaussian process $\{z_d^q(t)\}_{d=1}^D$ using gradient-descent methods [11,22]. With a set of input points, $\mathbf{t} = \{t_n\}_{n=1}^N$, the marginal-likelihood is given as $p(\mathbf{z}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}, \mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma})$, where $\mathbf{z} = [\mathbf{z}_1^\top, \cdots, \mathbf{z}_D^\top]^\top$, with $\mathbf{z}_d = [z_d(t_1), \cdots, z_d(t_N)]^\top$, $\mathbf{K}_{\mathbf{z},\mathbf{z}}$ is a $D \times D$ block-partitioned matrix with blocks $\mathbf{K}_{\mathbf{z}_d,\mathbf{z}_{d'}}$. The entries in each of these blocks are evaluated using $k_{z_d,z_{d'}}(t, t')$. Furthermore, $k_{z_d,z_{d'}}(t, t')$ is computed according the section 6.1.2. Appendix D shows more details about the maximum log-likelihood of a joint Gaussian process. In appendices E and F, we compute the derivatives necessary to perform the gradient-descent method.

6.3 Biological datasets

In this part, we describe all the datasets used to validate the performance of the proposed framework when modelling transcriptional regulation of gene expression in different applications (e.g. Escherichia coli baterium, and Yeast metabolisms). The datasets were collected from the state-of-art related to transcription networks [3, 4, 9, 23, 24].

6.3.1 E.coli dataset

Escherichia coli (E.coli) is a robust organism that can adapt remarkably well to changes in its environment [23]. The sudden oxygen starvation in the environment provokes the necessity of routinely adaptation in the bacterium. Here, the bacterium can be expelled from the host's gut and very rapidly moves from an environment with virtually no oxygen to an aerobic environment [1]. This change entails a whole shift in the metabolism of the bacterium, from a nitric metabolism to a much more energetically favourable aerobic metabolism. E.coli dataset contains the measurements of the expression levels of 5 genes: `ompW`, `yjiD`, `hypB`, `moaA`, and `aspA`. There are measurements at five different time instants: 0, 5, 10, 15 and 60 min. The arrays measure the change in concentration of mRNA relative to the initial point. The gene expressions are driven by the `FNR` regulator, which is completely unknown.

6.3.2 Yeast dataset

This dataset describes the biological dynamics of ribosomal protein production in the metabolism of a specific type of eukaryotic microorganism, also known as Yeast respiration [20]. The metabolic cycle was assayed using microarrays by Tu et al. in [24]. Yeast dataset contains the measurements of the expression levels of 3178 genes, measured in 36 time points sampled at 25 min intervals. Experimental studies of Yeast respiration have shown that there are two important transcriptional regulators for controlling the production of ribosomal proteins, `FHL1` and `RAP1` [20, 25, 26]. According the ChIP-on-chip data,[3] there are ten genes which depend on both transcription proteins. The `FHL1` protein regulates solely three genes: `TOS4`, `YLR030W` and `TKL2`. On the other hand, `RAP1` protein regulates solely two genes: `YOR359W` and `PFK27`. The remaining five genes code for ribosomal proteins and are jointly regulated by `FHL1` and `RAP1`, although the precise nature of the control is not known. These genes are `RPL9A`, `RPL13A`, `RPL17B`, `RPL30` and `RPS16B` [20].

[3] ChIP-on-chip is a technology that combines chromatin immunoprecipitation with DNA microarray. ChIP-on-chip is used to investigate interactions between proteins and DNA in vivo [20, 26].

7 Experimental Results and Discussion

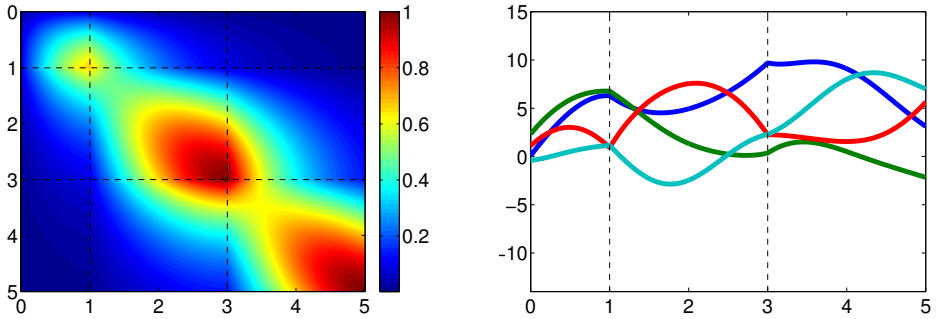
We now show and discuss the results of the proposed framework with artificial data and real biological data. First, we present the results obtained in different simulation experiments using the proposed switched dynamical latent force model for the first-order ODE. Finally, we present the results of the model when modelling transcriptional regulation of gene expression.

7.1 Toy examples

Using the model proposed in this work, we generate samples from the GP with zero mean and covariance function as explained in section 6. We implement several toy examples in order to evaluate the performance of the model under different conditions. We work with the inverse of the length-scale defined as $\ell_q = \sqrt{2/\hat{\ell}_q}$ for stability of the model.

7.1.1 Toy experiment 1: covariance examples

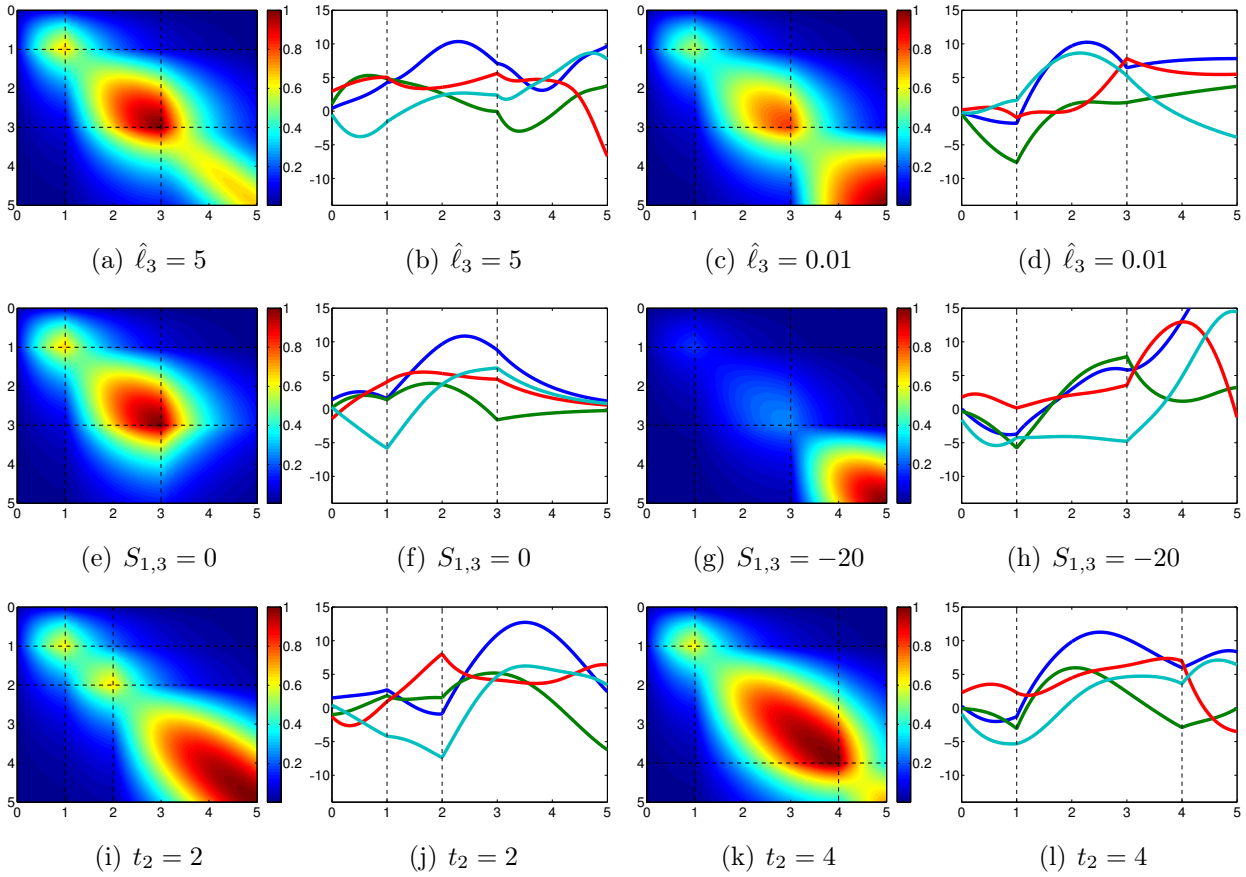
In Figure 3, we compute the covariance function $k_{z_d,z_{d'}}(t,t')$, and some samples generated from the zero mean GP with the same covariance function. For this experiment, we compute the covariance function from a model with $D = 1$ and $Q = 3$, with switching points $t_0 = -0.1$, $t_1 = 1$ and $t_2 = 3$. For the output, we fix the decay value $\gamma_d = 1$. We also restrict the latent forces to have the same inverse of length-scale value $\hat{\ell}_1 = \hat{\ell}_2 = \hat{\ell}_3 = 1$, and fix the same values of sensitivity parameters as $S_{1,1} = S_{1,2} = S_{1,3} = 10$. From both Figures, dashed lines indicate the final value of the switching points. Figure 3(b) shows that the samples of the outputs are continuous across the switching points (assumption proposed in the definition of the model in section 6.1.1). This condition of continuity is prescribed in the smooth transition across the switching points of the covariance function as we can see in Figure 3(a).



(a) (b)

Figure 3: Toy experiment 1.a. (a) covariance function between the outputs of the model with $D = 1$ and $Q = 3$, with switching points $t_0 = -0.1$, $t_1 = 1$ and $t_2 = 3$. (b) samples generated from the GP with covariance function as explained before.

In Figure 4, we show similar results than Figure 3, when we change a single parameter of the covariance example of Figure 3(a). In the caption of the sub-figures we describe the parameter



(a) $\hat{\ell}_3 = 5$ (b) $\hat{\ell}_3 = 5$ (c) $\hat{\ell}_3 = 0.01$ (d) $\hat{\ell}_3 = 0.01$

(e) $S_{1,3} = 0$ (f) $S_{1,3} = 0$ (g) $S_{1,3} = -20$ (h) $S_{1,3} = -20$

(i) $t_2 = 2$ (j) $t_2 = 2$ (k) $t_2 = 4$ (l) $t_2 = 4$

Figure 4: Toy experiment 1.b. In first and third columns, we show covariance function between the outputs of the model, when we change one of the parameters of the covariance example of Figure 3(a). In the second and fourth columns, we generate samples from the GP with the different covariance functions. In the caption of each sub-figure, we describe the parameter which we modified with its corresponding value.

which we modified with its corresponding value. In each row we want to show the ability of the model to perform changes in: the length-scales of the forces (first row), sensitive parameter (second row), and switching points (third row). Here, we notice the flexibility of the covariance to represent different conditions per each interval, ensuring the condition of continuity in the outputs. For example, if the model requires the inference of quick time-varying outputs, it is possible to deal with this assuming a high value of the inverse of length-scale parameter (see Figure 4(a)). We can also control the sensitivity of a specific output under the effect of a specific latent force through the sensitivity parameter of the covariance function (e.g. if a specific interval does not depend on a specific latent force, we can fix a sensitive parameter equal to zero). Finally, according the last row, the covariance function has the ability to describe the changes in the switching time instants, and even more it is able to work with different number of switching points (we will show this property in section 7.2).

7.1.2 Toy experiment 2: inference examples

In this example, we implement several toy examples to evaluate the performance of the model under different conditions. We sample 5 times from each toy with each output having 500 data points, and add some noise with variance equal to five percent of the variance of each sampled output. In each of the five repetitions, we took $N = 200$ data points for training and the remaining 300 for testing. For the test step, we fix the same parameters which were used for training. We describe below the implemented toys to obtain the results from Figure 5.

Toy example A: in the first experiment, we sample from a model with $D = 2$ and $Q = 1$. In this experiment, we want to show the performance of the model over a SIM network. For the outputs, we have $\gamma_1 = 0.7$ and $\gamma_2 = 1.1$. The single latent force has the length-scale value $\hat{\ell} = 1 \times 10^{-2}$, with same sensitivity parameters $S_1 = S_2 = 1$.

Toy example B: in this experiment, we sample from a model with $D = 2$ and $Q = 3$, with switching points $t_0 = -1$, $t_1 = 5$ and $t_2 = 12$. For the outputs, we have $\gamma_1 = 2.0$ and $\gamma_2 = 1.5$. We restrict the latent forces to have the same length-scale value $\hat{\ell}_0 = \hat{\ell}_1 = \hat{\ell}_2 = 1 \times 10^{-3}$, but change the values of the sensitivity parameters as $S_{1,1} = 10$, $S_{1,2} = 1$, $S_{1,3} = 10$, $S_{2,1} = 5$, $S_{2,2} = -10$ and $S_{2,3} = 1$, where the first sub-index refers to the output d and the second sub-index refers to the force in the interval q. In this experiment, we want to show the ability of the model to detect changes in the sensitivities of the forces, while keeping the length-scales equal along the intervals.

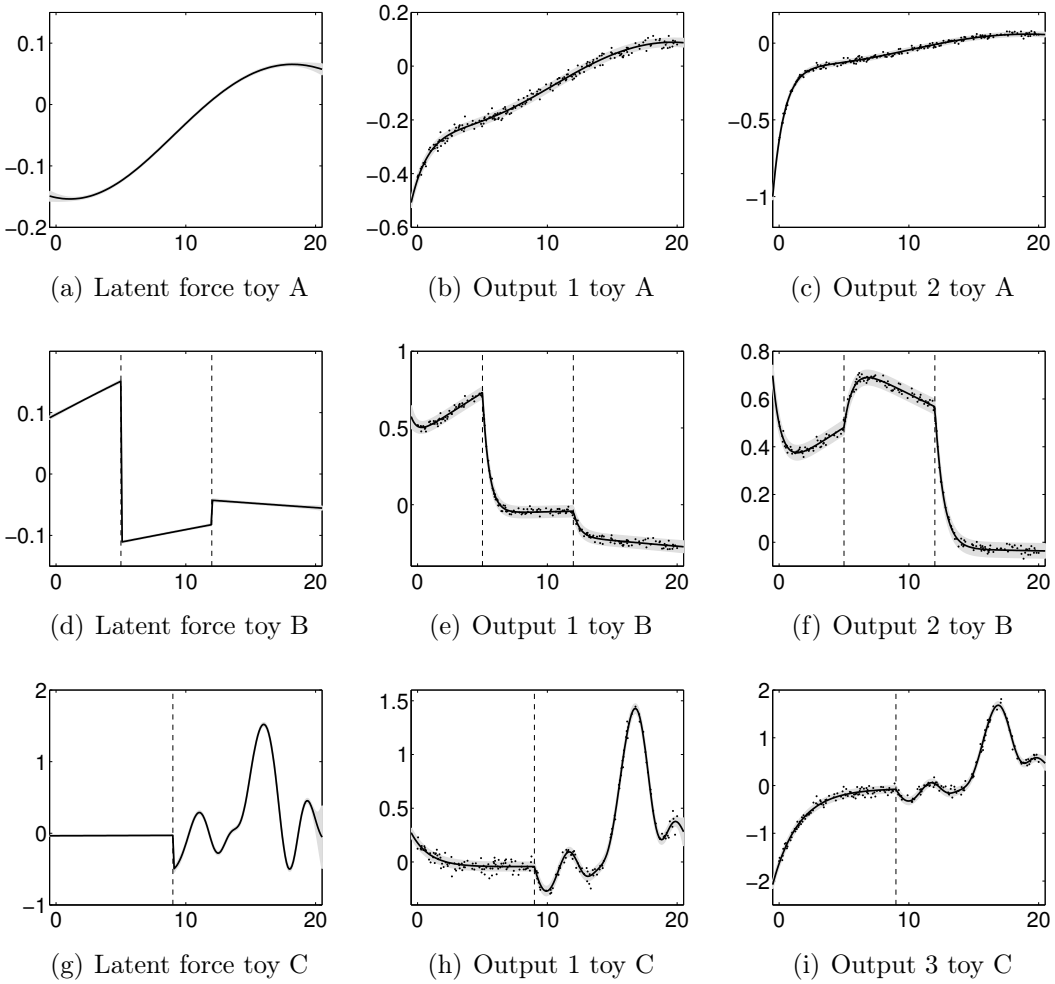
Toy example C: we sample from a model with $D = 3$ and $Q = 2$, with switching points $t_0 = -2$ and $t_1 = 9$. For the outputs, we have $\gamma_1 = 0.7$, $\gamma_2 = 1.5$, $\gamma_3 = 0.5$, and length-scales $\hat{\ell}_0 = 1 \times 10^{-3}$, and $\hat{\ell}_1 = 1$. The sensitive parameters in this case are given by $S_{1,1} = 1$, $S_{1,2} = 1$, $S_{2,1} = 5$, $S_{2,2} = 1$, $S_{3,1} = 1$ and $S_{3,2} = 1$.

Figure 5 shows the results per each toy example proposed in this experiment. The inference for the toy examples A, B and C are shown in the first, second and third row, respectively. The inference procedure evidences that the model reconstruct the different outputs of every toy example proposed.

7.1.3 Toy experiment 3: hyper-parameter estimation example

In this experiment, we use the toy example B described in the section 7.1.2. In this case, we change the initial set of hyper-parameters, and perform the gradient-descend method in order to maximize the marginal-likelihood of the joint Gaussian process described in section 6.2. We have to highlight that the hyper-parameter estimation of the covariance functions is a non-convex optimization problem. Depending on the initial set of parameters, we can find local optimum solutions which can infer properly the output profile. But the estimated latent forces can be different compared to latent driving force assumed in this experiment. In order to estimate a similar latent force obtained in the toy example B, we set the following initial set of hyper-parameters. For the outputs, we start $\gamma_1 = 1.0$ and $\gamma_2 = 0.5$, with switching points $t_0 = -1$, $t_1 = 2$ and $t_2 = 14$. We restrict the latent forces to have the same length-scale value $\hat{\ell}_0 = \hat{\ell}_1 = \hat{\ell}_2 = 1 \times 10^{-3}$, and we set the values of the sensitivity parameters as $S_{1,1} = 10$, $S_{1,2} = 1$, $S_{1,3} = 10$, $S_{2,1} = 5$, $S_{2,2} = -10$ and $S_{2,3} = 1$.

Figure 6 shows the sequence of the convergence using the first sample of the model until to obtain the estimated set of hyper-parameter for some iterations of the gradient method. In the first row, we



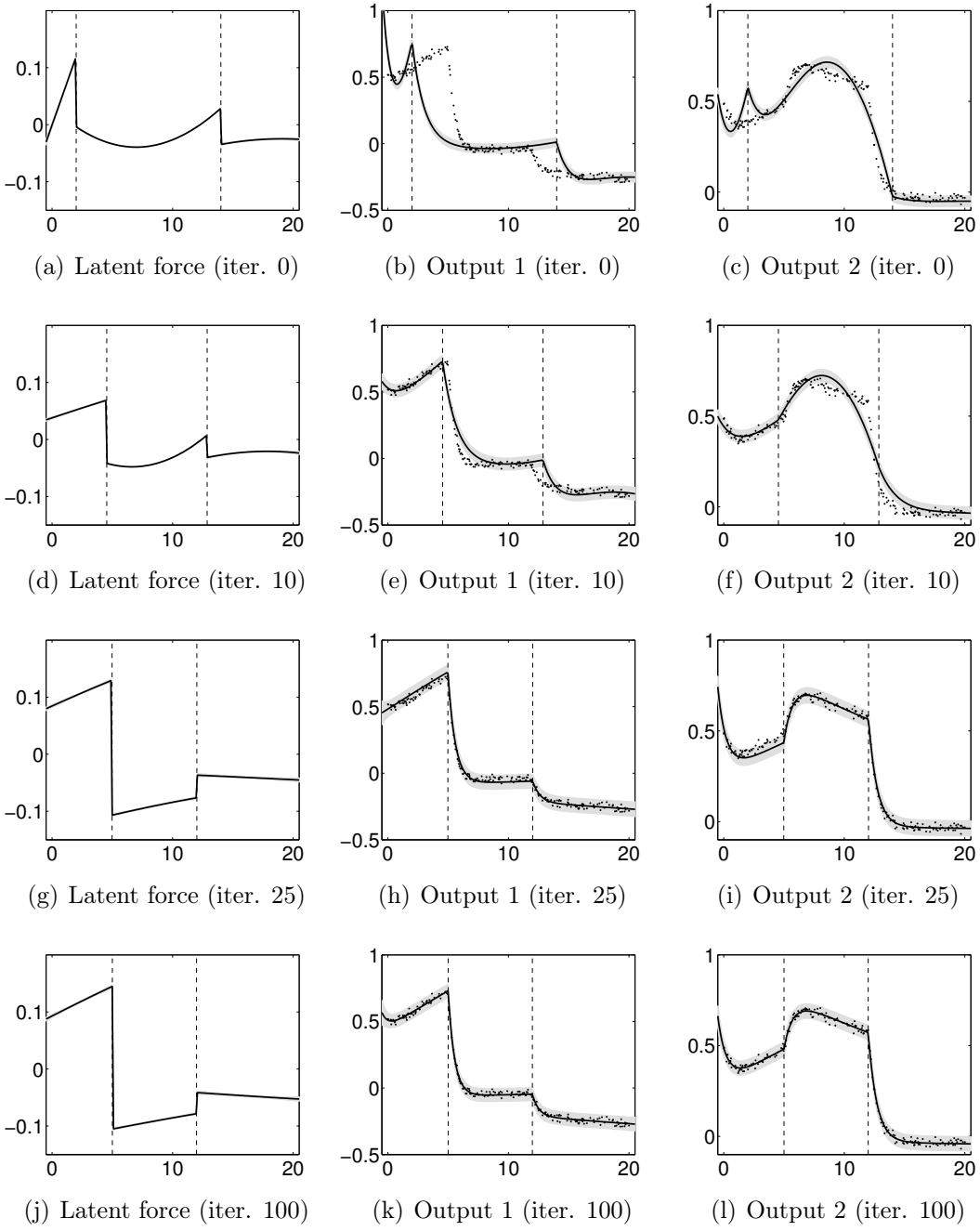
(a) Latent force toy A (b) Output 1 toy A (c) Output 2 toy A

(d) Latent force toy B (e) Output 1 toy B (f) Output 2 toy B

(g) Latent force toy C (h) Output 1 toy C (i) Output 3 toy C

Figure 5: Toy experiment 2. Mean and two standard deviations for the predictions over the latent force and two of the three outputs in the test set. Dashed lines indicate the final value of the switching points. Dots indicate training data.

show the latent force and the outputs of the model with the initial set of hyper-parameters. In the last row, we show the results using the estimated parameters after 100 iterations. For the iteration 100, the model estimate approximately the set of hyper-parameters, and recover the corresponding outputs and latent forces proposed in the toy example B. Table 1 shows the values of the estimated parameters for different iterations. In the first column, we show the iterations (Iter.). The other columns correspond to the parameters of the model. We can see that the optimization algorithm only modify the parameters which are different of the original set, aiming to converge to the true ones. The real values (RV) of the parameters are showed in the last row of the table. In the last two columns, we measure two error metrics for all the five samples, namely, the mean standardized log loss (MSLL), and the mean standardized mean square error (SMSE) [15]. We compute the mean and the standard deviation of the error results obtained per each sample ($\mu \pm \sigma$). We conclude that the optimization algorithm reduce the error produced in the inference procedure in each iteration, promoting the lowest possible error in the iteration 100.



(a) Latent force (iter. 0) (b) Output 1 (iter. 0) (c) Output 2 (iter. 0)

(d) Latent force (iter. 10) (e) Output 1 (iter. 10) (f) Output 2 (iter. 10)

(g) Latent force (iter. 25) (h) Output 1 (iter. 25) (i) Output 2 (iter. 25)

(j) Latent force (iter. 100) (k) Output 1 (iter. 100) (l) Output 2 (iter. 100)

Figure 6: Toy experiment 3. Mean and two standard deviations for the predictions over the latent force and the two outputs in the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

Iter.	Hyper-parameters of the model										Error	
	γ_1	t_1	t_2	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	γ_2	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$	SMSE	MSLL
0	1.000	2.000	14.00	10.00	1.000	10.00	0.500	5.000	-10.00	1.000	0.274 ± 0.120	152.5 ± 89.11
1	0.958	2.140	13.31	10.00	0.996	10.00	0.503	4.997	-10.00	0.999	0.207 ± 0.091	104.1 ± 66.95
5	0.730	4.337	13.36	10.01	0.987	9.999	0.541	4.975	-10.00	1.002	0.070 ± 0.037	21.93 ± 12.82
10	0.840	4.536	12.87	10.02	1.061	9.997	0.619	4.959	-10.00	1.029	0.034 ± 0.018	7.916 ± 3.053
25	1.640	4.977	12.01	10.03	1.234	9.982	1.402	4.932	-9.999	1.180	0.016 ± 0.017	-1.077 ± 1.337
50	1.926	5.003	11.98	10.02	1.108	9.988	1.399	4.948	-10.01	1.122	0.010 ± 0.017	-2.604 ± 0.609
100	1.917	4.997	11.98	10.02	1.106	9.988	1.415	4.946	-10.01	1.121	0.006 ± 0.007	-2.785 ± 0.403
RV	2.000	5.000	12.00	10.00	1.000	10.00	1.500	5.000	-10.00	1.000		

Table 1: Toy experiment 3. Performance of the gradient-descend method in order to maximize the marginal-likelihood of the joint Gaussian process. The first column shows iterations (Iter.). The other columns correspond to the parameters of the model. The last two columns measure the mean standardized log loss (MSLL) and the mean standardized mean square error (SMSE) [15]. In the last row, we show the real values (RV) of the hyper-parameters of the covariance functions.

7.1.4 Toy experiment 4: methodology comparison

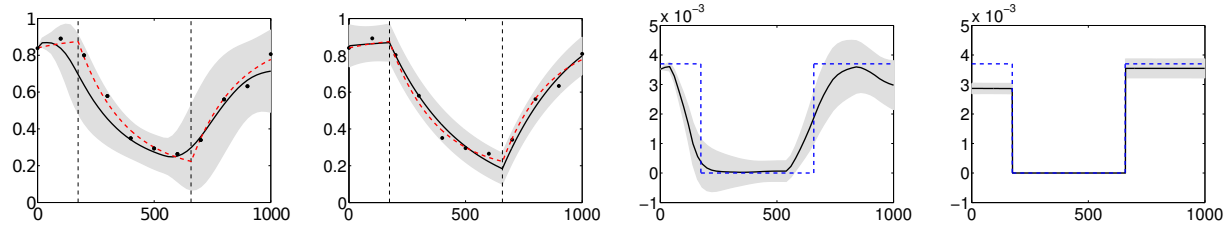
Sanguinetti et al. [1] presented an inference algorithm of transcription factors (TF) based on variational approximation. This approximation exploits the causal structure of the SIM framework to derive a forward-backward algorithm for the joint posterior over the outputs $z_d(t)$ and latent function $u(t)$. Their aim was to model a biological situation where a rapid response to a signal makes the TF activity quickly switch between the saturation level and zero. To encode the fact that $u(t)$ can perform an arbitrary number of switches between its two states, Sanguinetti et al. placed a prior on it in the form of a two-states Markov jump process [1]. Here, the inference task consists of two parts: state inference and the parameter estimation. In the first part, they use the noisy observations $\hat{z}_d(t)$ to infer the posterior distribution over the true state of the system (both $z_d(t)$ and $u(t)$). In the second part, they learn the model parameters.

In this experiment, our aim is to reconstruct the output from the synthetic data proposed in [1], in order to assess the validity of our approximation for modelling transcriptional regulation. We also want to evidence the ability of the proposed framework to detect the switching time instants when the quick time-varying behaviour of the TF activity is taken into account. The dataset is composed by a single output $z_d(t)$, driven by a single latent force $u(t)$. The latent force $u(t)$ represents a TF protein which transits from active to inactive state, and it is composed by 1000 time points. The synthetic TF is defined by

$$u(t) = \begin{cases} 1, & t \in [0, 169] \cup [660, 1000] \\ 0, & t \in [170, 659] \end{cases}.$$

The differential equation parameters were chosen as $B = 8 \times 10^{-4}$ (basal transcription), and $\gamma = 5 \times 10^{-3}$ (decay rate). The authors fixed the sensitivity parameter $S = 3.7 \times 10^{-3}$, which is the same during all the process. We compare the results obtained employing the model based on Markov jump processes [1], and using our proposed framework. The hyper-parameters were optimized with the corresponding estimation modules of the models. Because TF concentrations have to be strictly positive, in this experiment we restricted the sensitivity parameters to be also positive.[4]

[4]This assumption does not always guarantee that the latent force $u(t)$ be strictly a positive function. It is



(a) Sanguinetti et al. [1]: gene profile (b) Proposed framework with $Q = 3$: gene profile (c) Sanguinetti et al. [1]: TF activity (d) Proposed framework with $Q = 3$: TF activity

Figure 7: Toy experiment 4. Mean and two standard deviations for the predictions over the latent force and the output in the test set. Vertical dashed lines indicate the real value of the switching points. Dots indicate training data. Red dashed lines indicate test data. Blue dashed lines indicate synthetic latent force employed to generate the test data. The reconstructed output profiles (first two columns), and their corresponding scaled TF activities (last two columns), are showed using Sanguinetti et al. [1], and employing our proposed framework. The TF activities were scaled according their sensitivity parameters.

Employing both inference methods independently in this experiment, Figure 7 shows the reconstructed output profiles (first two columns), and their corresponding TF activities (last two columns). The red dashed lines indicate test data, and blue dashed lines indicate synthetic latent force employed to generate the test data. Figure 7(a) shows the reconstructed outputs using Sanguinetti et al. [1], and Figure 7(b) shows the result employing our proposed framework with $Q = 3$.[5] We conclude that our approach improves the inference of the gene profile (output), respect to the results obtained in [1]. On the other hand, we find differences between the magnitudes of the inferred TF activities, and the scale of the sensitive parameters, due to the nature of both models. For example, our framework estimates three sensitivity values $S_1 = 0.30, S_2 = 8.8 \times 10^{-5}$, and $S_3 = 1.1$ (one sensitivity parameter per each interval), while than Sanguinetti et al. only require one sensitivity parameter $S = 2.8 \times 10^{-3}$. However, we can compare both methods if we scale the latent forces $u(t)$ using their corresponding sensitivity parameters. Figures 7(c) and 7(d) show the scaled TF activity in the same order as we made for the outputs. From Figure 7(c), we see that the model proposed in [1] has problems to represent the TF activity when the TF transits from active to inactive state, and vice versa. It is due to the smoothed tendency of their model. From Figure 7(d), we conclude that our framework estimates correctly the switching time instants of the process, outperforming the representation of the quick time-varying behaviour of the TF activity. Related to the hyper-parameters estimation, we fixed a basal transcription equal to zero, and we obtained an estimated decay rate equal to $\gamma = 3.2 \times 10^{-3}$ (similar value than Sanguinetti et al. obtained, $(\gamma = 4.0 \times 10^{-3})$ [1]. Respect to the inverse of length-scale parameters, we obtained $\hat{\ell}_0 = 5.8 \times 10^{-8}$, $\hat{\ell}_0 = 1.8 \times 10^{-6}$ and $\hat{\ell}_0 = 6.5 \times 10^{-10}$, justifying why the latent functions are completely flat.

because the Gaussian process over the latent forces is not restricted to only generate positive functions [15]. However, this condition seems to be enough for our model.

[5]We fixed the number of intervals $Q = 3$ at the beginning of the experiment, because it was easy to see this quantity from the output profile. However, we performed the model with different numbers of intervals, and we obtained a maximum value of the log-likelihood for $Q = 3$.

7.2 Modelling transcriptional regulation

In this section, we employ the proposed framework when modelling transcriptional regulation of gene expression, evaluating its performance using the datasets described in section 6.3. We train a model for each dataset according to the nature of the biological applications. In each experiment, we initialize manually the set of hyper-parameters, aiming to achieve the lower possible error in the reconstruction of the gene profiles. After having fixed the initial set of hyper-parameters, we perform the optimization algorithm based on the gradient-descend method in order to maximize the marginal log-likelihood (section 6.2). For the outputs, we assume that the basal transcription is prescribed in the initial condition and the sensitivity parameters, allowing us to make zero the terms $\{B_d\}_{d=1}^{D}$ for any biological dataset. We also work with the inverse of the length-scale defined as $\ell_q = (2/\hat{\ell}_q)^{1/2}$ for stability of the model.

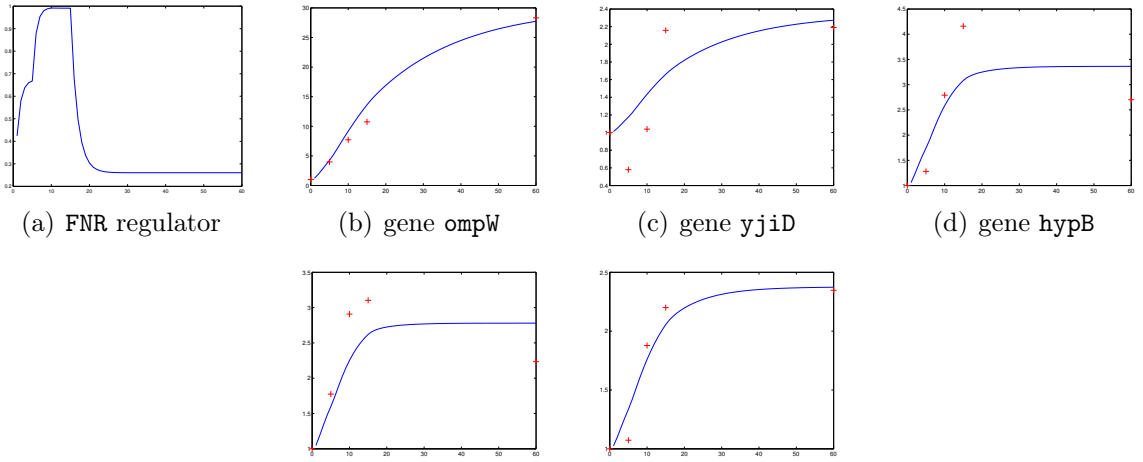
7.2.1 E.coli data results

As we described in section 7.2, E.coli is a robust organism that can adapt remarkably well to changes in its environment [23]. The sudden oxygen starvation in the environment provokes the necessity of routinely adaptation in the bacterium. This change entails a whole shift in the metabolism of the bacterium, from a nitric metabolism to a much more energetically favourable aerobic metabolism. In this experiment, we take into account the five gene expressions available from the dataset (`ompW`, `yjiD`, `hypB`, `moaA`, and `aspA`). According to Sanguinetti et al in [1], the system appears to undergo a sharp transition between inactive and active state at around 3 to 6 min. For the authors, this results in an interesting prediction: removing oxygen for a period shorter than 3 min will not lead to an `FNR`-mediated transcriptional response. Therefore, one may view this as an indirect measurement of the time it takes E.coli to commit itself to change its metabolic regime between aerobic and nitric. More details about the biological phenomenon can be found in [1]. In this experiment we are interested in the reconstruction of the five gene activities, and in the inference of the `FNR` activity. Due that the activity of `FNR` regulator is completely unknown, we compare the results obtained with our framework respect to the results obtained by Sanguinetti et al in [1].

Figure 8 shows the results obtained by [1]. Sub-figure (a) shows the activity of the `FNR` regulator. Sub-figures (b), (c), (d) (e) and (f) show the activity of genes: `ompW`, `yjiD`, `hypB`, `moaA` and `aspA` (respectively). We notice that the fit of their model to the expression activities `yjiD`, `hypB` and `moaA` are not as good as in the other two cases. In particular, `hypB` expression markedly decreases from 15 min to 60 min, which is incompatible with the other profiles, and can hardly be accommodated by their model [1]. The authors justified this problem due to the effect of the TF `IHF`, which also activates `hypB` but is repressed by `FNR`. Concluding that, after a certain amount of time, the SIM approximation breaks down in this case. Similar explanations can be provided for gene `yjiD` and gene `moaA`.

For the implementation of our framework, we implemented a model with $D = 5$ (each output represent a different gene activity). Due to the large lapse between the two final time instants, we fix the number of intervals $Q = 2$, with switching points $t_0 = -1$, and $t_1 = 39$.[6] However, we expect that the second estimated switching point be near to the 15 min where the measures are no

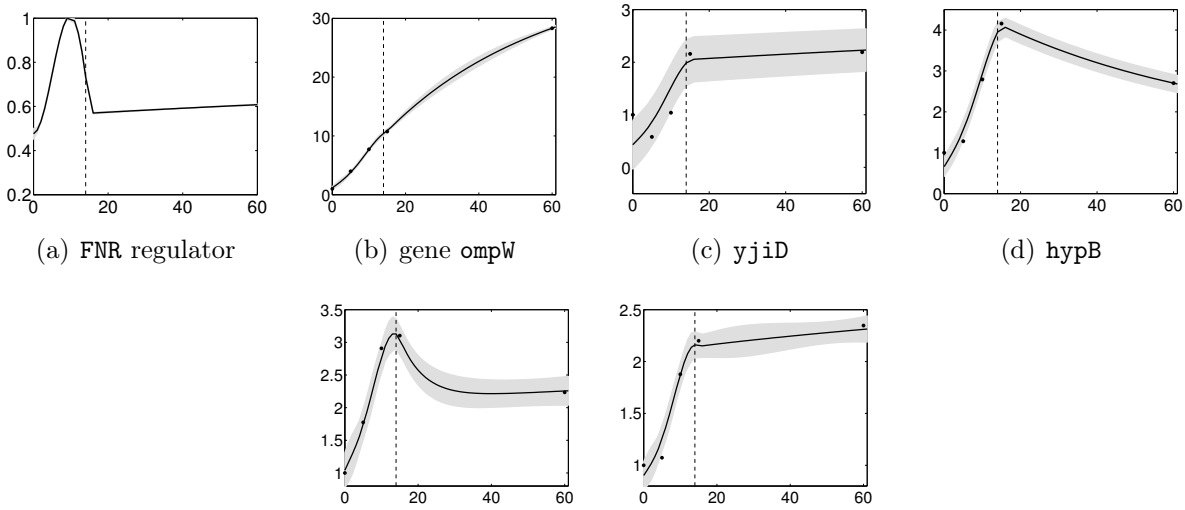
[6] We also implemented the model for different number of intervals Q, obtaining worse results. Due to the lack of data between 15 and 60 min, our framework with $Q > 2$ tends to over-fit the model in the interval between 0 to 15 min.



(a) FNR regulator (b) gene ompW (c) gene yjiD (d) gene hypB

(e) gene moaA (f) gene aspA

Figure 8: Microaerobic shift in E.coli results employing Sanguinetti et al [1]. Mean for the predictions over the FNR regulator and the genes profiles from the E.coli dataset. Red crosses indicate training data.



(a) FNR regulator (b) gene ompW (c) yjiD (d) hypB

(e) gene moaA (f) gene aspA

Figure 9: Microaerobic shift in E.coli results employing our proposed framework. Mean and two standard deviations for the predictions over the FNR regulator and the genes profiles in the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

longer taken periodically. We initialize the parameters with the same decay $\{\gamma_d\}_{d=1}^{D} = 1$ rate, same sensitivity parameters $\{S_{d,q}\}_{d=1,q=1}^{D,Q} = 1$, and same inverse of length-scales $\{\hat{\ell}_{q-1}\}_{q=1}^{Q} = 1 \times 10^{-5}$. As well as we made with the results obtained by Sanguinetti et al., in Figure 9 we show the results obtained employing our proposed framework after the inference procedure. According the figures 8(a) and 9(a), we notice the same problem we detailed in the experiment of section 7.1.4: difference

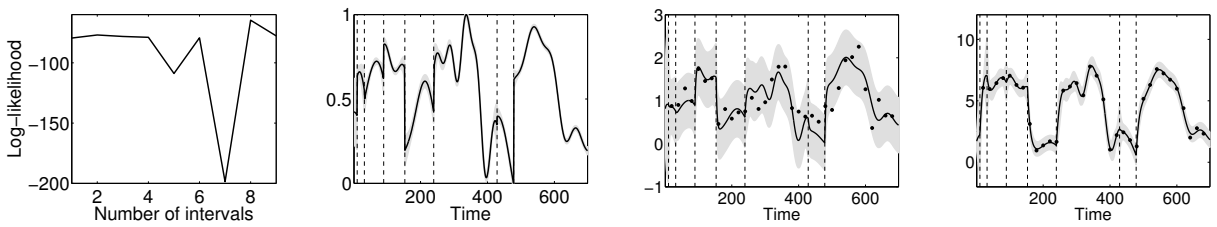
between the magnitude of the FNR activities. However, we observe that both protein activities follow similar dynamics where the FNR regulator is repressed after the 18 min, instant that our model selects as second switching point. According the genes profiles, we observe that our framework is able to represent the genes expression activities, even the expression activities of the genes: yjiD, hypB and moaA.

7.2.2 Yeast data results

To test our model on this data, we use independently both transcriptional regulators controlling the production of ribosomal proteins, FHL1 and RAP1, obtaining two independent biological problems. Due to there are some gene profiles which depend on both TF proteins (e.g. RPL9A, RPL13A) [20], we only consider the gene activities which are regulated solely by one of the TF regulators in each experiment. This is because our framework assumes that only one TF is acting over the biological system. To fix the number of outputs D, we consider the number of gene profiles in each problem. In both experiments, we train different models using different values of Q, including the latent force model without switches ($Q = 1$). Finally, we show the results obtained using the models which produced the better log-likelihood performance.

For the TF FHL1, we associate the genes profiles TOS4, YLR030W and TKL2, obtaining a total number of outputs $D = 3$. We initialize the parameters with the same decay $\{\gamma_d\}_{d=1}^{D} = 0.5$, same sensitivity parameters $\{S_{d,q}\}_{d=1,q=1}^{D,Q} = 2$, and same inverse of length-scales $\{\hat{\ell}_{q-1}\}_{q=1}^{Q} = 1 \times 10^{-3}$. Figure 10 shows the log-likelihood performance, the inferred TF activity and two of the three genes profiles. Sub-figure (a) shows peaks for the log-likelihood at $Q = 8$, concluding that our model performs better when we fix eight intervals. According to the gene profiles, we conclude once again that our model can represent the quick time-varying behaviour of the TF regulator. Our framework also estimates a set of switching points in which the protein activity changes are considerable. Related to the FHL1 profile, due to the difficulty to obtain experimental measures of the TF activity, we cannot validate directly our model. However, our model infers a TF activity which follows a similar dynamic than the results obtained by Opper et al. [20]. According to [20], the TF is active during three main intervals between 0 to 180 min, 200 to 380 min, and between 500 to 650 min. In other intervals, the TF protein looks like repressed. Sub-figure (b) shows that our framework presents the same properties for the inferred TF activity. Due to the approach proposed by Opper et al. is based on the model proposed [1], we justify the difference in magnitude between both TF activities due to the difference in the inferred scales between both models. It is also interesting to notice that the fit of the model to the YLR030W expression profile is not as good as in the TKL2 expression profile (see sub-figures (c) and (d), respectively). This problem is because the YLR030W profile is incompatible with the other profiles and can hardly be accommodated by the model. We can force our model to represent more appropriately the YLR030W expression profile, but it incurs in the over-fitting of the model, producing a wrong inference of the TF FHL1 activity. We also have to emphasize that the genes profiles, in which the FHL1 protein is acting together with the RAP1 protein, were not taken into account. This limits the performance of our approach because those genes could provide more information on the behaviour of the TF protein.

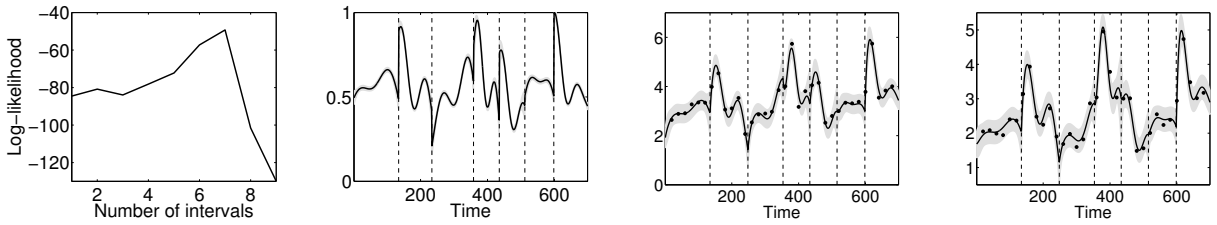
For the TF RAP1, we associate the genes profiles YOR359W and PFK27, obtaining a total number of outputs $D = 2$. We repeat a similar procedure as we made for the ribosomal protein FHL1. Figure 11 shows the results for the regulator RAP1. For this case, we obtain peaks for the log-likelihood at $Q = 7$. Once again, according to the results in [20], our inferred TF protein follows a similar dynamic, but with different magnitude. This validates partially our model. The results obtained



(a) Log-Likelihood (b) `FHL1` protein (c) gene `YLR030W` (d) gene `TKL2`

Figure 10: Yeast metabolic cycle: `FHL1` protein. Mean and two standard deviations for the predictions over the TF `FHL1` and two of the three genes profiles from the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

in [20] evidences that the TF `RAP1` presents several activation peaks at the instants 150, 210, 400 and 600 min (approximately). Sub-figure (b) shows that our model detects similar instant times in which the activation peaks occur. It is interesting to notice that the estimated switching points are near to the instants in which the peaks occur in order to improve the inference of the TF `RAP1` activity.



(a) Log-Likelihood (b) `RAP1` protein (c) gene `YOR359W` (d) gene `PFK27`

Figure 11: Yeast metabolic cycle: `RAP1` protein. Mean and two standard deviations for the predictions over the TF `RAP1` and the two genes profiles from the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

Finally, according the results given in section 7.2, we conclude that our proposed framework can be applied potentially when modelling transcriptional regulation of gene expression. Unfortunately, we could not validate directly our models with experimental TF activities because the measurement of TF profiles is difficult to obtain experimentally. We validated the performance of our models, comparing our results respect to different results presented in the state-of-art. We evaluated the performance of our approach for different number of outputs, and for a large number of intervals. Finally, we also evidenced the necessity to extend our framework in which we can take into account that several transcription regulators are acting in each interval.

8 Conclusions and future works

In this project we introduced a switched dynamical latent force model when modelling transcriptional regulation of gene expression. We used the first-order non-homogeneous ordinary differential equation and Gaussian process prior to model gene activities in transcription networks as well as its protein concentration. The results show that the proposed framework can model a multi-output transcription network, as well as the dynamic of switching protein concentration problems. According to the general objective and the specific objectives describe in section 4, next we summarize the contributions of the project.

Specific objective 1. The mathematical formulation of a multi-output switched dynamical latent force model based on Gaussian processes, that allows the prediction of the solution for the first-order non-homogeneous ordinary differential equation.

Specific objective 2. The development of an statistical inference procedure for computing the posterior probability distribution over the Gaussian process variables, and for hyper-parameter estimation of the covariance functions. Respect to the statistical inference procedure, the model provides both the solution to direct problems (e.g. inference of gene activities), as well as inverse problems (e.g. inference of protein concentrations). Respect to the hyper-parameter estimation, we developed a methodology based on maximum log-likelihood of the joint GP using a gradient-descent method.

Specific objective 3. The validation of the performance of the proposed framework when modelling transcriptional regulation of gene expression. We proposed a methodology which can describe the gene activity taking into account the switching time instants in protein concentrations due to external or internal stimuli (e.g. damage to the cell, sudden oxygen starvation, or ribosomal protein production).

We highlight that the algorithms of the proposed framework were programmed in Matlab®. The programming codes are based on the GPmat Matlab Toolbox provided by the University of Sheffield. This model could be implemented in other applications which imply the use of the first-order non-homogeneous ODE such as socio-economic problems (e.g. the population growth, and compound interest), and some particular mechanical systems (e.g. free fall movements, and Newton's law of cooling). On the other hand, there are some limitations of the proposed framework which we propose as future works. Next, we list the activities for future works.

Future work 1. Become the model a completely non-parametric model, in which the number of intervals Q could be another hyper-parameter to be estimated.

Future work 2. Generalize the model when several latent forces are acting per each interval. Here, we could also propose a sparse approach which selects the main latent forces from a fix number of forces which are taken into consideration per each interval.

Future work 3. In the proposed framework, we assumed that the hyper-parameters of the mechanistic system (e.g. decay rate) were stationary parameters per each output. We could extend this approach assuming non-stationary hyper-parameters, which constantly can change per each interval.

A Solution of the first-order non-homogeneous ODE

Let the first-order non-homogeneous ordinary differential equation (ODE) given by

$$\frac{dy_d(t)}{dt} + \gamma_d y_d(t) = B_d + \sum_{r=1}^{R} S_{d,r} u_r(t),$$

where the outputs $y_d(t)$ are driven by a set of forces $\{u_r(t)\}_{r=1}^{R}$ with sensitivities $\{S_{d,r}\}_{r=1}^{R}$, and constants γ_d and B_d. Next, we compute the homogeneous (transient-state), the non-homogeneous (steady-state), and the complete solutions of first-order ODE.

Homogeneous solution: the homogeneous solution $y_d^h(t)$ has to satisfy the expression $\dot{y}_d(t) + \gamma_d y_d(t) = 0$. Assuming that $y_d^h(t) = K_1 \exp\{-\gamma_d t\}$, it is possible to prove the solution satisfies the homogeneous condition [27].

$$\frac{dy_d^h(t)}{dt} + \gamma_d y_d^h(t) = \frac{d}{dt}\Big[K_1 \exp\{-\gamma_d t\}\Big] + \gamma_d \Big[K_1 \exp\{-\gamma_d t\}\Big] = 0.$$

Here, the constant K_1 depends of the initial condition of the system.

Non-homogeneous solution: according to the superposition principle and the integrating factor theory, the non-homogeneous solution can be separated in two particular solutions of the form $\dot{y}_d(t) + p(t) y_d(t) = q(t)$, which its solution is given by $y_d(t) = \frac{1}{u(t)} \int_0^t u(\tau) q(\tau) d\tau$, with $u(t) = \exp\{\int_0^t p(\tau) d\tau\}$ [27]. According to this, the complete solution of the non-homogeneous condition is given by

$$\begin{aligned} y_d^p(t) &= \exp\{-\gamma_d t\} \left[\int_0^t B_d \exp\{\gamma_d \tau\} d\tau + \int_0^t \exp\{\gamma_d \tau\} \left(\sum_{r=1}^{R} S_{d,r} u_r(\tau) \right) d\tau \right] \\ &= \Big[1 - \exp\{-\gamma_d t\}\Big] \frac{B_d}{\gamma_d} + \sum_{r=1}^{R} S_{d,r} \exp\{-\gamma_d t\} \int_0^t \exp\{\gamma_d \tau\} u_r(\tau) d\tau. \end{aligned}$$

Complete solution: the complete solution of the first-order non-homogeneous ODE is given by the sum of the homogeneous solution and non-homogeneous solutions, obtaining

$$y_d(t) = y_d^h(t) + y_d^p(t) = K_1 c_d(t) + [1 - c_d(t)] \frac{B_d}{\gamma_d} + \sum_{r=1}^{R} f_d(t, u_r),$$

with $f_d(t, u_r) = S_{d,r} c_d(t) \int_0^t \exp\{\gamma_d \tau\} u_r(\tau) d\tau$ and $c_d(t) = \exp\{-\gamma_d t\}$. Finally, evaluating the initial condition and organizing the terms, we obtain

$$y_d(t) = [1 - c_d(t)] \frac{B_d}{\gamma_d} + c_d(t) y_d(0) + \sum_{r=1}^{R} f_d(t, u_r).$$

B Covariances for the SIM approximation

Let $y_d(t) = [1 - c_d(t)]\frac{B_d}{\gamma_d} + c_d(t)y_d(0) + f_d(t,u)$, the solution of the ODE proposed in the Equation (2), with $c_d(t) = \exp\{-\gamma_d t\}$, $f_d(t,u) = S_d c_d(t)\int_0^t u(\tau)\exp\{\gamma_d\tau\}\,d\tau$, and $R = 1$. Next, we compute the covariance of the outputs and the covariance between outputs and latent functions. Also, we compute the covariances $k_{f_d,f_{d'}}(t,t')$ and $k_{f_d,u}(t,t')$

Covariance for the outputs $k_{y_d,y_{d'}}(t,t')$

$$\begin{aligned}
k_{y_d,y_{d'}}(t,t') &= \operatorname{cov}\{y_d(t), y_{d'}(t')\} \\
&= \mathbb{E}\{y_d(t)y_{d'}(t')\} - \mathbb{E}\{y_d(t)\}\mathbb{E}\{y_{d'}(t')\} \\
&= \mathbb{E}\{c_d(t)c_{d'}(t')y_d(0)y_{d'}(0) + c_d(t)y_d(0)f_{d'}(t',u) + c_{d'}(t')f_d(t,u)y_{d'}(0) + f_d(t,u)f_{d'}(t',u)\} \\
&= c_d(t)c_{d'}(t')\operatorname{cov}\{y_d(0), y_{d'}(0)\} + c_d(t)\operatorname{cov}\{y_d(0), f_{d'}(t',u)\} \\
&\quad + c_{d'}(t')\operatorname{cov}\{f_d(t,u), y_{d'}(0)\} + \operatorname{cov}\{f_d(t,u), f_{d'}(t',u)\},
\end{aligned}$$

where $\operatorname{cov}\{y_d(0), f_{d'}(t',u)\} = \operatorname{cov}\{f_d(t,u), y_{d'}(0)\} = 0$, because the initial conditions are independent of the latent force. Finally, the resulting covariance is given by $k_{y_d,y_{d'}}(t,t') = c_d(t)c_{d'}(t')\sigma_{y_d,y_{d'}} + k_{f_d,f_{d'}}(t,t')$, where $\sigma_{y_d,y_{d'}} = \operatorname{cov}\{y_d(0), y_{d'}(0)\}$ are entries of the covariance matrix K_{IC}. Covariance $k_{f_d,f_{d'}}$ is described further below.

Covariance between outputs and latent functions $k_{y_d,u}(t,t')$

$$\begin{aligned}
k_{y_d,u}(t,t') &= \operatorname{cov}\{y_d(t), u(t')\} \\
&= \mathbb{E}\{y_d(t)u(t')\} - \mathbb{E}\{y_d(t)\}\mathbb{E}\{u(t')\} \\
&= \mathbb{E}\{c_d(t)y_d(0)u(t') + f_d(t,u)u(t')\} \\
&= c_d(t)\operatorname{cov}\{y_d(0), u(t')\} + \operatorname{cov}\{f_d(t,u), u(t')\},
\end{aligned}$$

where $\operatorname{cov}\{y_d(0), u(t')\} = 0$, because the initial conditions are independent of the latent force. Finally, the resulting covariance is $k_{y_d,u}(t,t') = k_{f_d,u}(t,t')$, with $k_{f_d,u}(t,t') = \operatorname{cov}\{f_d(t,u), u(t')\}$. The covariance $k_{f_d,u}(t,t')$ is described further below.

Covariance between $f_d(t)$ and $f_{d'}(t')$

The covariance $k_{f_d,f_{d'}}(t,t')$ is given by

$$\begin{aligned}
k_{f_d,f_{d'}}(t,t') &= \operatorname{cov}\{f_d(t,u), f_{d'}(t',u)\} \\
&= \mathbb{E}\left\{\left[S_d c_d(t)\int_0^t u(\tau)\exp\{\gamma_d\tau\}\,d\tau\right]\left[S_{d'}c_{d'}(t')\int_0^{t'} u(\tau')\exp\left\{\gamma_{d'}\tau'\right\}d\tau'\right]\right\} \\
&= \mathbb{E}\left\{S_d S_{d'}c_d(t)c_{d'}(t')\int_0^t u(\tau)\exp\{\gamma_d\tau\}\int_0^{t'} u(\tau')\exp\left\{\gamma_d\tau'\right\}d\tau' d\tau\right\} \\
&= S_d S_{d'}c_d(t)c_{d'}(t')\int_0^t \exp\{\gamma_d\tau\}\int_0^{t'}\exp\left\{\gamma_d\tau'\right\}k_{u,u}(t,t')d\tau' d\tau,
\end{aligned}$$

where $k_{u,u}(t,t') = \text{cov}\{u(t), u(t')\} = \exp\{(t-t')^2/\ell^2\}$. The double integral of the previous equation follows the general form

$$H(a,b,u,v) = \int_0^v \exp\{bz\} \int_0^u \exp\{az'\} \exp\left\{ -\frac{(z-z')^2}{\sigma^2} \right\} dz'dz,$$

which its solution is given by

$$H(a,b,u,v) = \frac{\sigma\sqrt{\pi}}{2}\Big[h(a,b,v,u) + h(b,a,u,v)\Big], \qquad a+b \neq 0$$

where (using David's order for notation)

$$h(\zeta,\rho,\nu,\varphi) = \frac{\exp\{(\zeta\sigma/2)^2\}}{\zeta+\rho}\Big[\exp\{(\zeta+\rho)\nu\}\,\mathcal{H}(\zeta,\varphi,\nu) - \mathcal{H}(\zeta,\varphi,0)\Big]$$
$$\mathcal{H}(\zeta,\varphi,\nu) = \text{erf}\left\{\frac{\varphi-\nu}{\sigma} - \frac{\sigma\zeta}{2}\right\} + \text{erf}\left\{\frac{\nu}{\sigma} + \frac{\sigma\zeta}{2}\right\}.$$

Then we have the following expression for the covariance

$$\begin{aligned} k_{f_a,f_{a'}}(t,t') &= \frac{S_d S_{d'}\ell\sqrt{\pi}}{2} c_d(t)c_{d'}(t')\Big[h(\gamma_{d'},\gamma_d,t,t') + h(\gamma_d,\gamma_{d'},t',t)\Big] \\ &= \frac{S_d S_{d'}\ell\sqrt{\pi}}{2} \underbrace{\exp\{-(\gamma_d t + \gamma_{d'}t')\}\Big[h(\gamma_{d'},\gamma_d,t,t') + h(\gamma_d,\gamma_{d'},t',t)\Big]}_{A}, \end{aligned}$$

with $\gamma_{d'} + \gamma_d \neq 0$. Reorganizing the term A,

$$\begin{aligned} A &= \exp\{-(\gamma_d t + \gamma_{d'}t')\}\,\frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\Big[\exp\{(\gamma_d+\gamma_{d'})t\}\mathcal{H}(\gamma_{d'},t',t) - \mathcal{H}(\gamma_{d'},t',0)\Big] \\ &\quad + \exp\{-(\gamma_d t + \gamma_{d'}t')\}\,\frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\Big[\exp\{(\gamma_d+\gamma_{d'})t'\}\,\mathcal{H}(\gamma_d,t,t') - \mathcal{H}(\gamma_d,t,0)\Big] \\ &= \frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\Big[\exp\{-\gamma_{d'}(t'-t)\}\mathcal{H}(\gamma_{d'},t',t) - \exp\{-(\gamma_d t + \gamma_{d'}t')\}\,\mathcal{H}(\gamma_{d'},t',0)\Big] \\ &\quad + \frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\Big[\exp\{-\gamma_d(t-t')\}\,\mathcal{H}(\gamma_d,t,t') - \exp\{-(\gamma_d t + \gamma_{d'}t')\}\,\mathcal{H}(\gamma_d,t,0)\Big] \\ &= \hat{h}(\gamma_{d'},\gamma_d,t,t') + \hat{h}(\gamma_d,\gamma_{d'},t',t), \end{aligned}$$

with $\nu_{d'} = \ell\gamma_{d'}/2$. Finally, the covariance follows

$$k_{f_a,f_{a'}}(t,t') = \frac{S_d S_{d'}\ell_q\sqrt{\pi}}{2}\Big[\hat{h}(\gamma_{d'},\gamma_d,t,t') + \hat{h}(\gamma_d,\gamma_{d'},t',t,)\Big],$$

where

$$\hat{h}(\gamma_{d'},\gamma_d,t,t') = \frac{1}{\gamma_d+\gamma_{d'}}\Big[\Upsilon(\gamma_{d'},t',t) - \exp\{-\gamma_d t\}\,\Upsilon(\gamma_{d'},t',0)\Big],$$

with $\nu_{d'} = \ell\gamma_{d'}/2$, and $\Upsilon(\gamma_{d'},t',t)$ is given by

$$\Upsilon(\gamma_{d'},t',t) = \exp\{\nu_{d'}^2\}\exp\{-\gamma_{d'}(t'-t)\}\Big[\operatorname{erf}\left\{\frac{t'-t}{\ell_q}-\nu_{d'}\right\}+\operatorname{erf}\left\{\frac{t}{\ell}+\nu_{d'}\right\}\Big].$$

Covariance between $f_d(t)$ and latent force $u(t')$

The covariance $k_{f_d,u}(t,t')$ is given by

$$\begin{aligned}
k_{f_d,u}(t,t') &= \operatorname{cov}\{f_d(t,u),u(t')\}\\
&= \mathbb{E}\left\{\left[S_d c_d(t)\int_0^t u(\tau)\exp\{\gamma_d\tau\}\,d\tau\right]u(t')\right\}\\
&= \mathbb{E}\left\{S_d c_d(t)\int_0^t \exp\{\gamma_d\tau\}\,u(\tau)u(t')d\tau\right\}\\
&= S_d c_d(t)\int_0^t \exp\{\gamma_d\tau\}\,k_{u,u}(\tau,t')d\tau,
\end{aligned}$$

where $k_{u,u}(t,t') = \operatorname{cov}\{u(t),u(t')\} = \exp\{(t-t')^2/\ell^2\}$. The integral of the previous equation follows the general form

$$G(a,u) = \int_0^u \exp\{az\}\exp\left\{-\frac{(z-z')^2}{\sigma^2}\right\}dz.$$

The solution of the previous expression is given by

$$G(a,u) = \frac{\sigma\sqrt{\pi}}{2}\exp\{(a\sigma/2)^2\}\exp\{az'\}\Big[\operatorname{erf}\left\{\frac{u-z'}{\sigma}-\frac{a\sigma}{2}\right\}+\operatorname{erf}\left\{\frac{z'}{\sigma}-\frac{a\sigma}{2}\right\}\Big].$$

Then we have the following expression for the covariance

$$\begin{aligned}
k_{f_d,u}(t,t') &= \frac{S_d\ell\sqrt{\pi}}{2}\exp\{\nu_d^2\}\exp\{-\gamma_d(t-t')\}\Big[\operatorname{erf}\left\{\frac{t-t'}{\ell}-\nu_d\right\}+\operatorname{erf}\left\{\frac{t'}{\ell}+\nu_d\right\}\Big]\\
&= \frac{S_{d'}\ell\sqrt{\pi}}{2}\Upsilon(\gamma_d,t,t'),
\end{aligned}$$

with $\nu_d = \ell\gamma_d/2$.

C Covariances for the SIM approximation with R independent latent forces

Let $y_d(t) = [1 - c_d(t)]\frac{B_d}{\gamma_d} + c_d(t)y_d(0) + \sum_{r=1}^{R} f_d(t, u_r)$, the complete solution of the ODE proposed in the Equation (2), with $c_d(t) = \exp\{-\gamma_d t\}$, $f_d(t, u_r) = S_{d,r}c_d(t)\int_0^t u_r(\tau)\exp\{\gamma_d\tau\}\,d\tau$, and R independent latent functions. Next, we compute the covariance of the outputs and the covariance between outputs and latent functions.

Covariance for the outputs $k_{y_d,y_{d'}}(t,t')$

$$\begin{aligned}
k_{y_d,y_{d'}}(t,t') &= \operatorname{cov}\{y_d(t), y_{d'}(t')\} \\
&= \mathbb{E}\{y_d(t)y_{d'}(t')\} - \mathbb{E}\{y_d(t)\}\mathbb{E}\{y_{d'}(t')\} \\
&= \mathbb{E}\Big\{c_d(t)c_{d'}(t')y_d(0)y_{d'}(0) + c_d(t)\sum_{r=1}^{R} y_d(0)f_{d'}(t', u_r) + c_{d'}(t')\sum_{r=1}^{R} f_d(t, u_r)y_{d'}(0) \\
&\quad + \sum_{r=1}^{R}\sum_{p=1}^{P} f_d(t, u_r)f_{d'}(t', u_p)\Big\} \\
&= c_d(t)c_{d'}(t')\operatorname{cov}\{y_d(0), y_{d'}(0)\} + c_d(t)\sum_{r=1}^{R}\operatorname{cov}\{y_d(0), f_{d'}(t', u_r)\} \\
&\quad + c_{d'}(t')\sum_{r=1}^{R}\operatorname{cov}\{f_d(t, u_r), y_{d'}(0)\} + \sum_{r=1}^{R}\sum_{p=1}^{P}\operatorname{cov}\{f_d(t, u_r)f_{d'}(t', u_p)\},
\end{aligned}$$

where $\operatorname{cov}\{y_d(0), f_{d'}(t', u_r)\} = \operatorname{cov}\{f_d(t, u_r), y_{d'}(0)\} = 0$ for any value of r, because the initial conditions are independent of the latent forces. Also, the covariances $\operatorname{cov}\{f_d(t, u_r)f_{d'}(t', u_p)\} = 0$ if $p \neq r$, because there is not dependency between latent forces. Finally, the resulting covariance is given by $k_{y_d,y_{d'}}(t,t') = c_d(t)c_{d'}(t')\sigma_{y_d,y_{d'}} + \sum_{r=1}^{R} k_{f_{d,r},f_{d',r}}(t,t')$, where $\sigma_{y_d,y_{d'}} = \operatorname{cov}\{y_d(0), y_{d'}(0)\}$ are entries of the covariance matrix K_{IC}. The covariance $k_{f_{d,r},f_{d',r}}(t,t')$ is described in appendix B with a latent force $u_r(t)$.

Covariance between outputs and latent functions $k_{y_d,u_r}(t,t')$

$$\begin{aligned}
k_{y_d,u_r}(t,t') &= \operatorname{cov}\{y_d(t), u_r(t')\} \\
&= \mathbb{E}\{y_d(t)u_r(t')\} - \mathbb{E}\{y_d(t)\}\mathbb{E}\{u_r(t')\} \\
&= \mathbb{E}\Big\{c_d(t)y_d(0)u_q(t') + \sum_{r=1}^{R} f_d(t, u_r)u_r(t')\Big\} \\
&= c_d(t)\operatorname{cov}\{y_d(0), u_r(t')\} + \sum_{r=1}^{R}\operatorname{cov}\{f_d(t, u_r), u_r(t')\},
\end{aligned}$$

where $\operatorname{cov}\{y_d(0), u_r(t')\} = 0$ for any value of r, because the initial conditions are independent of latent forces. Finally, the resulting covariance is $k_{y_d,u_r}(t,t') = \sum_{r=1}^{R}\operatorname{cov}\{f_d(t, u_r), u_r(t')\}$. The covariance $\operatorname{cov}\{f_d(t, u_r), u_r(t')\}$ is described in appendix B with a latent force $u_r(t)$.

D Maximum log-likelihood of a joint Gaussian process

Let $p(\mathbf{z}|\boldsymbol{\theta})$ given by

$$p(\mathbf{z}|\boldsymbol{\theta}) = \mathcal{N}\left(\mathbf{z}, |\boldsymbol{\mu}, \mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma}\right) = \frac{|\mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{D/2}} \exp\left\{ -\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top (\mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma})^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}.$$

The logarithmic of the marginal-likelihood of the joint Gaussian process is given by

$$\ln p(\mathbf{z}|\boldsymbol{\theta}) = -\frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top (\mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma})^{-1}(\mathbf{z}-\boldsymbol{\mu}).$$

The maximum log-likelihood of a joint Gaussian process using gradient-descent methods requires the computation of the derivatives of $\ln p(\mathbf{z}|\boldsymbol{\theta})$ w.r.t. the hyper-parameters $\boldsymbol{\theta}$ [19, 22]. Using some properties of the derivatives of matrices [28], we obtain

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{z}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{1}{2}\operatorname{tr}\left\{\hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}}{\partial \boldsymbol{\theta}}\right\} + \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}}{\partial \boldsymbol{\theta}}\hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}) - (\mathbf{z}-\boldsymbol{\mu})^\top \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial(\mathbf{z}-\boldsymbol{\mu})}{\partial \boldsymbol{\theta}} \\
&= -\frac{1}{2}\operatorname{tr}\left\{\hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial \mathbf{K}_{\mathbf{z},\mathbf{z}}}{\partial \boldsymbol{\theta}}\right\} + \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial \mathbf{K}_{\mathbf{z},\mathbf{z}}}{\partial \boldsymbol{\theta}}\hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}) + (\mathbf{z}-\boldsymbol{\mu})^\top \hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}},
\end{aligned}$$

where $\hat{\mathbf{K}}_{\mathbf{z},\mathbf{z}} = \mathbf{K}_{\mathbf{z},\mathbf{z}} + \boldsymbol{\Sigma}$. Here, the noise covariance $\boldsymbol{\Sigma}$ does not depend on the parameters in $\boldsymbol{\theta}$, but the variance of the noise is also another parameter to be estimated. According the previous equation, for the gradient method, we have to compute the derivatives of the entries $k_{z_d,z_{d'}}(\cdot,\cdot)$ of each block-partitioned matrix in $\mathbf{K}_{\mathbf{z},\mathbf{z}}$. We also need to compute the derivatives over the mean vector $\boldsymbol{\mu}$. In appendix E, we compute the derivative of the kernel $k_{z_d,z_{d'}}$ w.r.t the independent time variable, namely, the velocity kernel. We use these derivatives for the optimization of the switching point instants. In appendix F, we compute the derivatives of the kernel w.r.t. the other hyper-parameters (e.g. decay rate and length-scales).

E Velocity kernels

In order to compute the gradients of the proposed framework w.r.t the switching points $\{t_q\}_{q=1}^{Q-1}$, we need to compute the velocity kernels. Since differentiation is a linear operation, the derivative of a Gaussian process is also a Gaussian process. To obtain the velocity kernel, we need to differentiate $k_{f_d,f_{d'}}(t,t')$ w.r.t. t and t'.

Kernel between $m_d(t)$ and $f_{d'}(t')$

The kernel can be obtained by

$$k_{m_d,f_{d'}}(t,t') = \frac{\partial}{\partial t}k_{f_d,f_{d'}}(t,t') = \sum_{q=1}^{Q}\frac{S_{d,q}S_{d',q}\ell_q\sqrt{\pi}}{2}\frac{\partial}{\partial t}k^{(q)}_{f_d,f_{d'}}(t,t') = \sum_{q=1}^{Q}\frac{S_{d,q}S_{d',q}\ell_q\sqrt{\pi}}{2}k^{(q)}_{m_d,f_{d'}}(t,t').$$

The term $k^{(q)}_{m_d,f_{d'}}(t,t')$ is given by

$$k^{(q)}_{m_d,f_{d'}}(t,t') = \hat{h}^q_t(\gamma_{d'},\gamma_d,t,t') + \hat{h}^q_t(\gamma_d,\gamma_{d'},t',t),$$

where $\hat{h}^q(\gamma_{d'},\gamma_d,t,t')$ and $\hat{h}^q(\gamma_d,\gamma_{d'},t',t)$ are given by

$$\hat{h}^q(\gamma_{d'},\gamma_d,t,t') = \frac{1}{\gamma_d+\gamma_{d'}}\left[\Upsilon^q(\gamma_{d'},t',t) - \exp\left\{-\gamma_d t\right\}\Upsilon^q(\gamma_{d'},t',0)\right],$$
$$\hat{h}^q(\gamma_d,\gamma_{d'},t',t) = \frac{1}{\gamma_d+\gamma_{d'}}\left[\Upsilon^q(\gamma_d,t,t') - \exp\left\{-\gamma_{d'} t'\right\}\Upsilon^q(\gamma_d,t,0)\right].$$

Their partial derivative with respect to t are given by

$$\hat{h}^q_t(\gamma_{d'},\gamma_d,t,t') = \frac{1}{\gamma_d+\gamma_{d'}}\left[\Upsilon^q_t(\gamma_{d'},t',t) + \gamma_d\exp\left\{-\gamma_d t\right\}\Upsilon^q(\gamma_{d'},t',0)\right],$$
$$\hat{h}^q_t(\gamma_d,\gamma_{d'},t',t) = \frac{1}{\gamma_d+\gamma_{d'}}\left[\Upsilon^q_t(\gamma_d,t,t') - \exp\left\{-\gamma_{d'} t'\right\}\Upsilon^q_t(\gamma_d,t,0)\right].$$

Derivative for $\Upsilon^q_t(\gamma_{d'},t',t)$ follows

$$\begin{aligned}\Upsilon^q_t(\gamma_{d'},t',t) &= \gamma_{d'}\exp\{\nu^2_{qd'}\}\exp\{-\gamma_{d'}(t'-t)\}\left[\operatorname{erf}\left\{\frac{t'-t}{\ell_q}-\nu_{qd'}\right\}+\operatorname{erf}\left\{\frac{t}{\ell_q}+\nu_{qd'}\right\}\right]\\
&\quad-\frac{2}{\sqrt{\pi}\ell_q}\exp\{\nu^2_{qd'}\}\exp\{-\gamma_{d'}(t'-t)\}\left[\exp\left\{-\left(\frac{t'-t}{\ell_q}-\nu_{qd'}\right)^2\right\}-\exp\left\{-\left(\frac{t}{\ell_q}+\nu_{qd'}\right)^2\right\}\right]\\
&= \gamma_{d'}\Upsilon^q(\gamma_{d'},t',t) - \frac{2}{\sqrt{\pi}\ell_q}\exp\left\{-\left(\frac{t'-t}{\ell_q}\right)^2\right\}+\frac{2}{\sqrt{\pi}\ell_q}\exp\{-\gamma_{d'}t'\}\exp\left\{-\left(\frac{t}{\ell_q}\right)^2\right\}.\end{aligned}$$

In similar way, the derivative for $\Upsilon_t^q(\gamma_{d'}, t, t')$ follows

$$\begin{aligned}\Upsilon_t^q(\gamma_{d'}, t, t') &= -\gamma_{d'} \exp\{\nu_{qd'}^2\} \exp\{-\gamma_{d'}(t-t')\}\left[\operatorname{erf}\left\{\frac{t-t'}{\ell_q} - \nu_{qd'}\right\} + \operatorname{erf}\left\{\frac{t'}{\ell_q} + \nu_{qd'}\right\}\right] \\ &\quad + \frac{2}{\sqrt{\pi}\ell_q} \exp\{\nu_{qd'}^2\} \exp\{-\gamma_{d'}(t-t')\} \exp\left\{-\left(\frac{t-t'}{\ell_q} - \nu_{qd'}\right)^2\right\} \\ &= -\gamma_{d'} \Upsilon^q(\gamma_{d'}, t, t') + \frac{2}{\sqrt{\pi}\ell_q} \exp\left\{-\left(\frac{t-t'}{\ell_q}\right)^2\right\}.\end{aligned}$$

Kernel between $f_d(t)$ and $m_{d'}(t')$

The kernel can be obtained by

$$k_{f_d, m_{d'}}(t, t') = \frac{\partial}{\partial t'} k_{f_d, f_{d'}}(t, t') = \sum_{q=1}^{Q} \frac{S_{d,q} S_{d',q} \ell_q \sqrt{\pi}}{2} \frac{\partial}{\partial t'} k_{f_d, f_{d'}}^{(q)}(t, t') = \sum_{q=1}^{Q} \frac{S_{d,q} S_{d',q} \ell_q \sqrt{\pi}}{2} k_{f_d, m_{d'}}^{(q)}(t, t').$$

The term $k_{f_d, m_{d'}}^{(q)}(t, t')$ is given by

$$k_{f_d, m_{d'}}^{(q)}(t, t') = \hat{h}_{t'}^q(\gamma_{d'}, \gamma_d, t, t') + \hat{h}_{t'}^q(\gamma_d, \gamma_{d'}, t', t),$$

where their partial derivative with respect to t are given by

$$\begin{aligned}\hat{h}_{t'}^q(\gamma_{d'}, \gamma_d, t, t') &= \frac{1}{\gamma_d + \gamma_{d'}}\left[\Upsilon_{t'}^q(\gamma_{d'}, t', t) - \exp\{-\gamma_d t\}\, \Upsilon_{t'}^q(\gamma_{d'}, t', 0)\right], \\ \hat{h}_{t'}^q(\gamma_d, \gamma_{d'}, t', t) &= \frac{1}{\gamma_d + \gamma_{d'}}\left[\Upsilon_{t'}^q(\gamma_d, t, t') + \gamma_{d'} \exp\left\{-\gamma_{d'} t'\right\} \Upsilon^q(\gamma_d, t, 0)\right].\end{aligned}$$

Derivatives of $\Upsilon_{t'}^q(\gamma_{d'}, t', t)$, $\Upsilon_{t'}^q(\gamma_{d'}, t', 0)$ and $\Upsilon_{t'}^q(\gamma_d, t, t')$ follow the same expression than the derivatives of $\Upsilon_t^q(\gamma_{d'}, t, t')$, $\Upsilon_t^q(\gamma_{d'}, t, 0)$ and $\Upsilon_t^q(\gamma_{d'}, t', t)$, respectively.

Kernel between $m_d(t)$ and $m_{d'}(t')$

The kernel can be obtained as

$$k_{m_d, m_{d'}}(t, t') = \sum_{q=1}^{Q} \frac{S_{d,q} S_{d',q} \ell_q \sqrt{\pi}}{2} k_{m_d, m_{d'}}^{(q)}(t, t').$$

As in the sections before, we need to obtain

$$\begin{aligned}\hat{h}_{tt'}^q(\gamma_{d'}, \gamma_d, t, t') &= \frac{1}{\gamma_d + \gamma_{d'}}\left[\Upsilon_{tt'}^q(\gamma_{d'}, t', t) + \gamma_d \exp\{-\gamma_d t\}\, \Upsilon_{t'}^q(\gamma_{d'}, t', 0)\right], \\ \hat{h}_{tt'}^q(\gamma_d, \gamma_{d'}, t', t) &= \frac{1}{\gamma_d + \gamma_{d'}}\left[\Upsilon_{tt'}^q(\gamma_d, t, t') + \gamma_{d'} \exp\left\{-\gamma_{d'} t'\right\} \Upsilon_t^q(\gamma_d, t, 0)\right].\end{aligned}$$

Derivatives $\Upsilon^q_{t'}(\gamma_{d'}, t', 0)$ and $\Upsilon^q_{t}(\gamma_{d}, t, 0)$ were obtained in previously. It remains to obtain $\Upsilon^q_{tt'}(\gamma_{d'}, t', t)$ and $\Upsilon^q_{tt'}(\gamma_{d'}, t, t')$

$$\begin{aligned}
\Upsilon^q_{tt'}(\gamma_{d'}, t', t) &= \frac{\partial}{\partial t}\Big[-\gamma_{d'}\Upsilon^q(\gamma_{d'}, t', t) + \frac{2}{\sqrt{\pi}\ell_q}\exp\Big\{-\Big(\frac{t'-t}{\ell_q}\Big)^2\Big\}\Big] \\
&= -\gamma_{d'}\Upsilon^q_t(\gamma_{d'}, t', t) + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t'-t)}{\ell_q^2}\exp\Big\{-\Big(\frac{t'-t}{\ell_q}\Big)^2\Big\}, \\
\Upsilon^q_{tt'}(\gamma_{d}, t, t') &= \frac{\partial}{\partial t}\Big[\Upsilon^q_{t'}(\gamma_{d}, t, t')\Big] \\
&= \gamma_d\Upsilon^q_t(\gamma_{d}, t, t') + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t-t')}{\ell_q^2}\exp\Big\{-\Big(\frac{t-t'}{\ell_q}\Big)^2\Big\} - \frac{2\gamma_d}{\sqrt{\pi}\ell_q}\exp\{-\gamma_d t\}\exp\Big\{-\Big(\frac{t'}{\ell_q}\Big)^2\Big\}.
\end{aligned}$$

Functions $\Upsilon^q(\gamma_{d'}, t, t')$ and $\Upsilon^q(\gamma_{d'}, t', t)$ along with their respective derivatives are the key to compute the velocity kernels. Table 2 summarizes these expressions together with their derivatives.

$\Upsilon^q(\gamma_{d'}, t', t)$	$\exp\{\nu^2_{qd'}\}\exp\{-\gamma_{d'}(t'-t)\}\Big[\operatorname{erf}\Big\{\frac{t'-t}{\ell_q} - \nu_{qd'}\Big\} + \operatorname{erf}\Big\{\frac{t}{\ell_q} + \nu_{qd'}\Big\}\Big]$
$\Upsilon^q_t(\gamma_{d'}, t', t)$	$\gamma_{d'}\Upsilon^q(\gamma_{d'}, t', t) - \frac{2}{\sqrt{\pi}\ell_q}\exp\Big\{-\Big(\frac{t'-t}{\ell_q}\Big)^2\Big\} + \frac{2}{\sqrt{\pi}\ell_q}\exp\{-\gamma_{d'}t'\}\exp\Big\{-\Big(\frac{t}{\ell_q}\Big)^2\Big\}$
$\Upsilon^q_t(\gamma_{d'}, t, t')$	$-\gamma_{d'}\Upsilon^q(\gamma_{d'}, t, t') + \frac{2}{\sqrt{\pi}\ell_q}\exp\Big\{-\Big(\frac{t-t'}{\ell_q}\Big)^2\Big\}$
$\Upsilon^q_{tt'}(\gamma_{d'}, t', t)$	$-\gamma_{d'}\Upsilon^q_t(\gamma_{d'}, t', t) + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t'-t)}{\ell_q^2}\exp\Big\{-\Big(\frac{t'-t}{\ell_q}\Big)^2\Big\}$
$\Upsilon^q_{tt'}(\gamma_{d'}, t, t')$	$\gamma_{d'}\Upsilon^q_t(\gamma_{d'}, t, t') + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t-t')}{\ell_q^2}\exp\Big\{-\Big(\frac{t-t'}{\ell_q}\Big)^2\Big\} - \frac{2\gamma_{d'}}{\sqrt{\pi}\ell_q}\exp\{-\gamma_{d'}t\}\exp\Big\{-\Big(\frac{t'}{\ell_q}\Big)^2\Big\}$

Table 2: Derivatives of $\Upsilon^q(\gamma_{d'}, t', t)$ and $\Upsilon^q(\gamma_{d'}, t, t')$ w.r.t. the time variable.

F Gradients of the kernels w.r.t. hyper-parameters

In this appendix, we compute the derivatives of the kernel $k_{z_d,z_{d'}}(t,t')$ w.r.t. the hyper-parameters of the model, including decays $\gamma_d, \gamma_{d'}$, length scales $\{\ell_{q-1}\}_{q=1}^{Q}$. We have to notice that the derivatives of the proposed framework will depend on the derivatives of functions accompanying the initial conditions, and the derivatives of the SIM approach, which will depend on the derivatives of the Upsilon function.

Gradients of functions accompanying the initial conditions

The relevant derivative for $c_d(t) = \exp\{-\gamma_d t\}$ is

$$\frac{\partial c_d(t)}{\partial \gamma_d} = -t\exp\{-\gamma_d t\} = -t c_d(t).$$

Gradients of functions $\Upsilon(\gamma_d, t', t)$ and $\Upsilon(\gamma_d, t, t')$

Let the Upsilon function given by

$$\Upsilon(\gamma_{d'}, t', t) = \exp\{\nu_{qd'}^2\}\exp\{-\gamma_{d'}(t'-t)\}\left[\operatorname{erf}\left\{\frac{t'-t}{\ell_q} - \nu_{qd'}\right\} + \operatorname{erf}\left\{\frac{t}{\ell_q} + \nu_{qd'}\right\}\right],$$

where $\nu_{qd'} = \ell_q\gamma_{d'}/2$. Its derivatives with respect to γ_d and ℓ_q, are given by

$$\begin{aligned}
\frac{\partial}{\partial\gamma_{d'}}\Upsilon(\gamma_{d'}, t', t) &= \Upsilon(\gamma_{d'}, t', t)\left[2\nu_{qd'}\frac{\partial\nu_{qd'}}{\partial\gamma_{d'}} - (t'-t)\right] - \exp\{\nu_{qd'}^2\}\exp\{-\gamma_{d'}(t'-t)\} \\
&\quad\times\left[\exp\left\{-\left(\frac{t'-t}{\ell_q} - \nu_{qd'}\right)^2\right\} - \exp\left\{-\left(\frac{t}{\ell_q} + \nu_{qd'}\right)^2\right\}\right]\frac{\partial\nu_{qd'}}{\partial\gamma_{d'}}, \\
\frac{\partial}{\partial\ell_q}\Upsilon(\gamma_{d'}, t', t) &= 2\nu_{qd'}\Upsilon(\gamma_{d'}, t', t)\frac{\partial\nu_{qd'}}{\partial\ell_q} - \exp\{\nu_{qd'}^2\}\exp\{-\gamma_{d'}(t'-t)\} \\
&\quad\times\left[\left(\frac{t'-t}{\ell_q^2} + \frac{\partial\nu_{qd'}}{\partial\ell_q}\right)\exp\left\{-\left(\frac{t'-t}{\ell_q} - \nu_{qd'}\right)^2\right\} + \left(\frac{t}{\ell_q^2} - \frac{\partial\nu_{qd'}}{\partial\ell_q}\right)\exp\left\{-\left(\frac{t}{\ell_q} + \nu_{qd'}\right)^2\right\}\right],
\end{aligned}$$

where $\frac{\partial\nu_{qd'}}{\partial\gamma_{d'}} = \frac{\ell_q}{2}$ and $\frac{\partial\nu_{qd'}}{\partial\ell_q} = \frac{\gamma_{d'}}{2}$. The derivatives for $\Upsilon(\gamma_{d'}, t, t')$ follow a similar structure than the derivatives of $\Upsilon(\gamma_{d'}, t', t)$.

Bibliography

[1] G. Sanguinetti, A. Ruttor, M. Opper, and C. Archambeau, "Switching regulatory models of cellular stress response," *Bioinformatics*, vol. 25, no. 10, pp. 1280–1286, March 2009.

[2] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, "Transcriptional regulation by the numbers: applications," *Current Opinion in Genetics & Development*, vol. 15, no. 2, pp. 125 – 135, 2005.

[3] N. D. Lawrence, G. Sanguinetti, and M. Rattray, "Modelling transcriptional regulation using Gaussian processes," 2006.

[4] M. Barenco, D. Tomescu, D. Brewer, R. Callard, J. Stark, and M. Hubank, "Ranked prediction of p53 targets using hidden variable dynamic modeling," *Genome Biology*, vol. 7, no. 3, pp. R25+, 2006.

[5] U. Alon, *An Introduction to Systems Biology*. Chapman and Hall, London, 2006.

[6] P. Gao, A. Honkela, M. Rattray, and N. D. Lawrence, "Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities," *Bioinformatics*, vol. 24, no. 10, pp. 170–175, 2008.

[7] D. J. Jenkins, B. Finkenstädt, and D. A. Rand, "A temporal switch model for estimating transcriptional activity in gene expression," *Bioinformatics*, vol. 29, no. 9, pp. 1158–1165, May 2013.

[8] C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor, "Gaussian process approximations of stochastic differential equations," in *Gaussian Processes in Practice*, ser. J. Mach. Learn. Res. Workshop Conf. Proc., vol. 1, 2007, pp. 1–16.

[9] M. A. Álvarez, D. Luengo, and N. D. Lawrence, "Latent force models." in *AISTATS*, ser. JMLR Proceedings, D. A. V. Dyk and M. Welling, Eds., vol. 5, 2009, pp. 9–16.

[10] ——, "Linear latent force models using Gaussian processes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 11, pp. 2693–2705, Nov 2013.

[11] M. A. Álvarez, J. R. Peters, N. D. Lawrence, and B. Schölkopf, "Switched latent force models for movement segmentation," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds. Curran Associates, Inc., 2010, pp. 55–63.

[12] M. K. Titsias, A. Honkela, N. D. Lawrence, and M. Rattray, "Identifying targets of multiple co-regulating transcription factors from expression time-series by Bayesian model comparison." *BMC Systems Biology*, vol. 6, p. 53, 2012.

[13] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, "Transcriptional regulation by the numbers: models," *Current Opinion in Genetics & Development*, vol. 15, no. 2, pp. 116 – 124, 2005.

[14] G. Malacinski, *Essentials of Molecular Biology.* Jones and Bartlett Publishers, Incorporated, 2005.

[15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005.

[16] J. D. Vásquez Jaramillo, M. A. Álvarez, and A. A. Orozco, "Latent force models for describing transcriptional regulation processes in the embryo development problem for the Drosophila melanogaster," in *Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE*, Aug 2014, pp. 338–341.

[17] P. A. Alvarado, M. A. Álvarez, G. Daza-Santacoloma, A. A. Orozco, and G. Castellanos-Domínguez, "A latent force model for describing electric propagation in deep brain stimulation: a simulation study," in *Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE*, Aug 2014, pp. 2617–2620.

[18] K. P. Murphy, *Machine Learning: A Probabilistic Perspective (Adaptive Computation And Machine Learning Series).* The MIT Press, 2012.

[19] C. M. Bishop, *Pattern Recognition And Machine Learning (Information Science And Statistics).* Springer, 2007.

[20] M. Opper and G. Sanguinetti, "Learning combinatorial transcriptional dynamics from gene expression data." *Bioinformatics*, vol. 26, no. 13, pp. 1623–1629, 2010.

[21] G. Chechik and D. Koller, "Timing properties of gene expression responses to environmental changes," *J. Computational Biology*, vol. 9, 2009.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall PTR, 1998.

[23] J. D. Partridge, G. Sanguinetti, D. P. Dibden, R. E. Roberts, R. K. Poole, and J. Green, "Transition of Escherichia coli from aerobic to micro-aerobic conditions involves fast and slow reacting regulatory components," *The Journal of Biological Chemistry*, vol. 282, no. 15, pp. 11 230–11 237, Apr. 2007.

[24] B. Tu, A. Kudlicki, M. Rowicka, and S. McKnight, "Logic of the Yeast metabolic cycle: Temporal compartmentalization of cellular processes," *Science*, vol. 310, no. 5751, pp. 1152–1158, Nov. 2005.

[25] C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young, "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99–104+, 2004.

[26] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, "Transcriptional regulatory networks in Saccharomyces cerevisiae," *Science*, vol. 298, pp. 799–804, 2002.

[27] R. K. Nagle, E. B. Saff, and A. D. Snider, *Fundamentals of differential equations and boundary value problems*, 6th ed. Pearson Addison-Wesley, 2012.

[28] K. B. Petersen, M. S. Pedersen, J. Larsen, K. Strimmer, L. Christiansen, K. Hansen, L. He, L. Thibaut, M. Barão, S. Hattinger, V. Sima, and W. The, "The matrix cookbook," ISP, IMM, Technical University of Denmark, Tech. Rep., 2006.