Switched Dynamical Latent Force Models for Modelling Transcriptional Regulation

Andrés F. López-Lopera and Mauricio A. Álvarez.

Faculty of Engineering, Universidad Tecnológica de Pereira, Colombia, 660003.

November 24, 2015

Abstract

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. 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. 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 dynamical systems (or 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, 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.

1 Transcription networks

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]. 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 [2]. Barenco et al. [3] have proposed a 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. [4] 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 [4] can also be extended to R protein concentrations and continuous genes profiles [2, 4], obtaining the following expression

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

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 (1) 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 $\{u_r(t)\}_{r=1}^R$ [3]. 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 [4–6].

2 Latent force model approach for SIM networks

As described in section 1, the SIM approximation proposed by [3] can be seen as an LFM where the mechanistic model is governed by a first-order non-homogeneous ODE given in Equation (1) 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 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 (1) for $R = 1$. According to the appendix A, the solution of the first-order ODE follows

$$y_d(t) = [1 - c_d(t)]\frac{B_d}{\gamma_d} + c_d(t)y_d(0) + f_d(t, u), \qquad (2)$$

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

$$k_{y_d, y_{d'}}(t, t') = c_d(t)c_{d'}(t')\sigma_{y_d, y_{d'}} + k_{f_d, f_{d'}}(t, t'), \qquad (3)$$

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

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

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 [2, 4]. The resulting covariance $k_{f_d, f_{d'}}(t, t')$ follows

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

where

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

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

and

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

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 (3). Appendix B describes all the procedure for computing the covariances of the SIM framework.

3 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 (1). 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 [7].

3.1 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 [7]. Figure 1 shows a cartoon representation of output $z_d(t)$ switching its behaviour between points t_0, t_1, t_2 and t_3.[1] For each interval (t_{q-1}, t_q), only the latent force $u_{q-1}(t)$ is active.
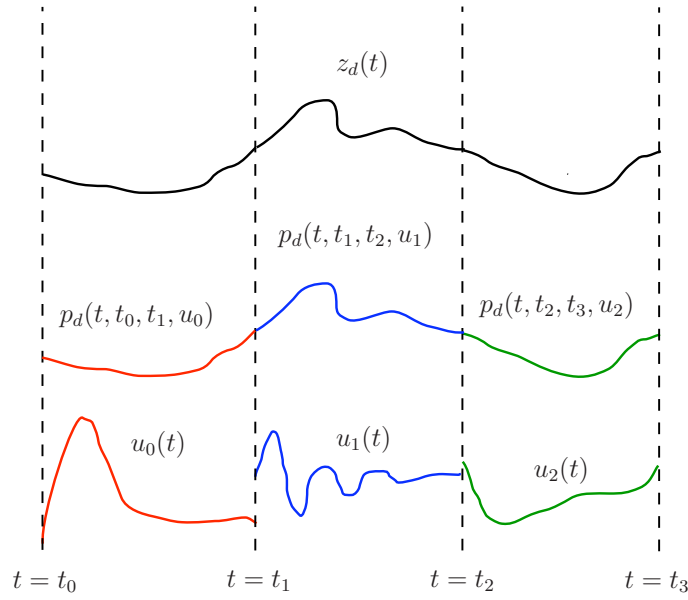

Figure 1: Cartoon representation for the output of a switched dynamical LFM proposed by Álvarez et al. in [7].

We can use the basic SIM model introduced in section 2 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

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

where $p_d(t, t_{q-1}, t_q, u_{q-1})$ uses the model in Equation (2)

$$p_d(t, t_{q-1}, t_q, u_{q-1}) = z_d(t)|_{t_{q-1}} = c_d(t - t_{q-1})z_d(t_{q-1}) + f_d(t, t_{q-1}, t_q, u_{q-1}). \qquad (4)$$

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 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

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

[1]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.

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 [4].

3.2   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 [7].

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 2. 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

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

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

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

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 $k_{f_d, f_{d'}}^{(q)}(t, t') = \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q-1}, t_q, u_{q-1})\}$.

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

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

where $\mathrm{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = 0$, because there is not correlation between the forces u_{q-1} and $u_{q'-1}$. Also, the covariance $\mathrm{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), z_{d'}(t_{q'-1})\}$ 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

$$
\mathrm{cov}\{z_d(t), z_{d'}(t')\} = c_d(t - t_{q-1})c_{d'}(t' - t_{q'-1})k_{z_d, z_{d'}}(t_{q-1}, t_{q'-1}) + c_d(t - t_{q-1})\,\mathrm{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}.
$$

According to the Equation (4), 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 $\mathrm{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), p_d(t_{q'-1}, t_{q'-2}, t_{q'-1}, u_{q'-2})\}$. Now, we have to compute the term $\mathrm{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$. This term is equal to

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

The term $\mathrm{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ 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

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

Here, the covariance $\mathrm{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ 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 $\mathrm{cov}\{z_d(t_{q-n}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = \mathrm{cov}\{p_d(t_{q-n}, t_{q-n-1}, t_{q-n}, u_{q-n-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\}$ is such that $q = q' + n$. If $q = q' + n$, we obtain the following recursive expression

$$
\mathrm{cov}\{z_d(t_{q-1}), f_{d'}(t', t_{q'-1}, t_{q'}, u_{q'-1})\} = \left[\prod_{i=1}^{n-1} c_d(t_{q-i} - t_{q-i-1})\right] k_{f_d, f_{d'}}^{(q'-1)}(t_{q-n}, t'). \tag{5}
$$

3.3   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

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

where $\mathrm{cov}\{z_d(t_{q-1}), u_{q-1}(t')\} = 0$ 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

$$\text{cov}\{z_d(t, t_{q-1}, t_q), u_{q'-1}(t')\} = \text{cov}\{p_d(t, t_{q-1}, t_q, u_{q-1}), u_{q'-1}(t')\}$$
$$= c_d(t - t_{q-1})\,\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} + \text{cov}\{f_d(t, t_{q-1}, t_q, u_{q-1}), u_{q'-1}(t')\},$$

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

$$\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} = \text{cov}\{p_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), u_{q'-1}(t')\}$$
$$= c_d(t_{q-1} - t_{q-2})\underbrace{\text{cov}\{z_d(t_{q-2}), u_{q'-1}(t')\}}_{B} + \text{cov}\{f_d(t_{q-1}, t_{q-2}, t_{q-1}, u_{q-2}), u_{q'-1}(t')\}.$$

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 $k_{f_d, u_{q'-1}}^{(q'-1)}(t_{q-1}, t')$. 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

$$\text{cov}\{z_d(t_{q-2}), u_{q'-1}(t')\} = \text{cov}\{p_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), u_{q'-1}(t')\}$$
$$= c_d(t_{q-2} - t_{q-3})\underbrace{\text{cov}\{z_d(t_{q-3}), u_{q'-1}(t')\}}_{B'} + \text{cov}\{f_d(t_{q-2}, t_{q-3}, t_{q-2}, u_{q-3}), u_{q'-1}(t')\}.$$

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 $k_{f_d, u_{q'-1}}^{(q'-1)}(t_{q-2}, t')$. 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

$$\text{cov}\{z_d(t_{q-1}), u_{q'-1}(t')\} = \left[\prod_{i=1}^{n-1} c_d(t_{q-i} - t_{q-i-1})\right] k_{f_d, u_{q'-1}}^{(q'-1)}(t_{q-n}, t'). \tag{6}$$

3.4 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 [7]. With a set of input points, $\mathbf{t} = \{t_n\}_{n=1}^{N}$, the marginal-likelihood is given as $p(\mathbf{z}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}, \mathbf{K_{z,z}} + \boldsymbol{\Sigma})$, where $\mathbf{z} = [\mathbf{z}_1^\top, \cdots, \mathbf{z}_D^\top]^\top$, with $\mathbf{z}_d = [z_d(t_1), \cdots, z_d(t_N)]^\top$, $\mathbf{K_{z,z}}$ is a $D \times D$ block-partitioned matrix with blocks $\mathbf{K_{z_d, z_{d'}}}$. 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 3.2. Appendix C shows more details about the maximum log-likelihood of a joint Gaussian process. In appendices D and E, we compute the derivatives necessary to perform the gradient-descent method.

4 Experimental Results

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.

4.1 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 3. 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.

4.1.1 Toy experiment 1: covariance examples

In Figure 2, 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 2(b) shows that the samples of the outputs are continuous across the switching points (assumption proposed in the definition of the model in section 3.1). This condition of continuity is prescribed in the smooth transition across the switching points of the covariance function as we can see in Figure 2(a).
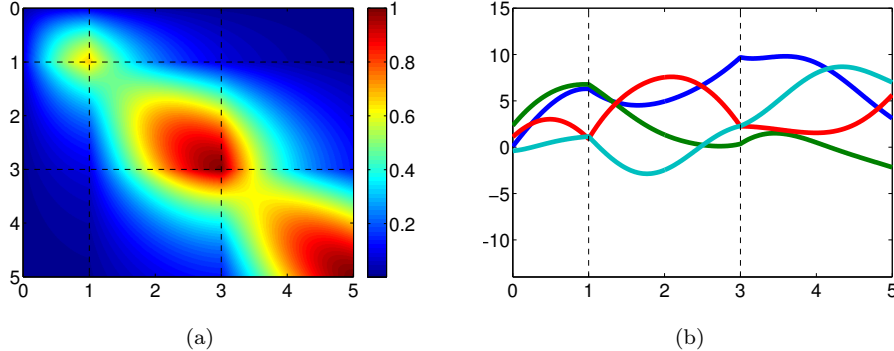

(a)   (b)

Figure 2: 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.

In Figure 3, we show similar results than Figure 2, when we change a single parameter of the covariance example of Figure 2(a). In the caption of the sub-figures we describe the parameter 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 3(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 4.2).

4.1.2 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 4.

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

(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) $S_{1,3} = -20$

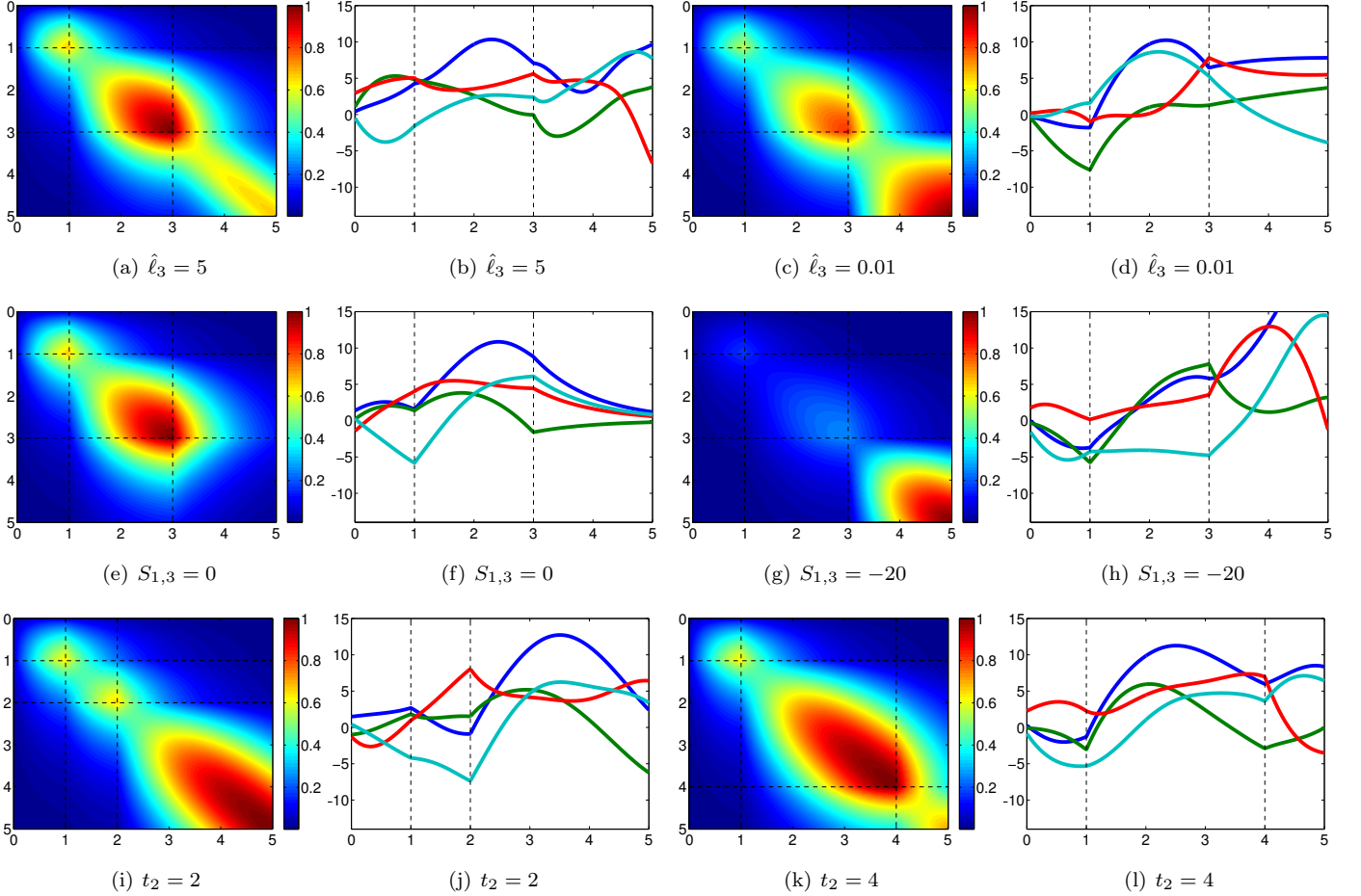(i) $t_2 = 2$ (j) $t_2 = 2$ (k) $t_2 = 4$ (l) $t_2 = 4$

Figure 3: 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 2(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.

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 4 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.

4.1.3 Toy experiment 3: hyper-parameter estimation example

In this experiment, we use the toy example B described in the section 4.1.2. 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 3.4. 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 5 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 column, we show the latent force and the outputs of the model with the initial set of hyper-parameters. In the last column, we show the results using the estimated parameters after 100 iterations. For the iteration 100, the model estimate approximately the set of hyper-parameters,

(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

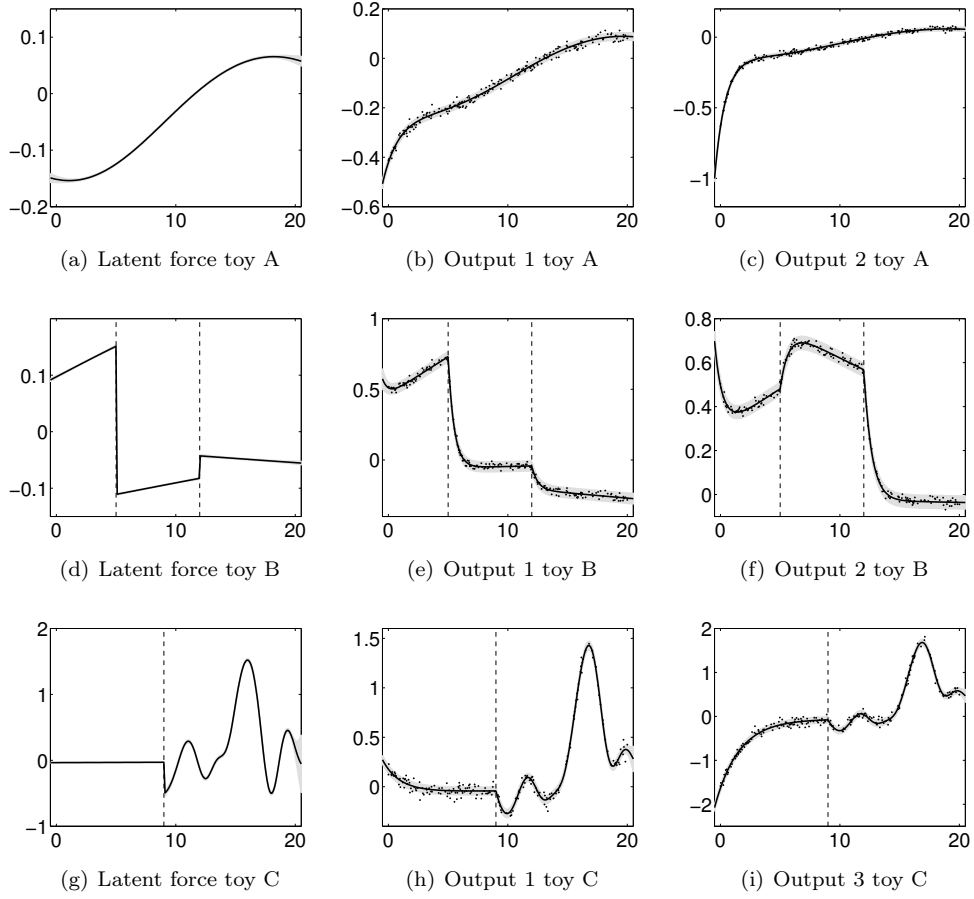(g) Latent force toy C   (h) Output 1 toy C   (i) Output 3 toy C

Figure 4: 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.

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) [8]. In the last row, we show the real values (RV) of the hyper-parameters of the covariance functions.

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) [8]. 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.
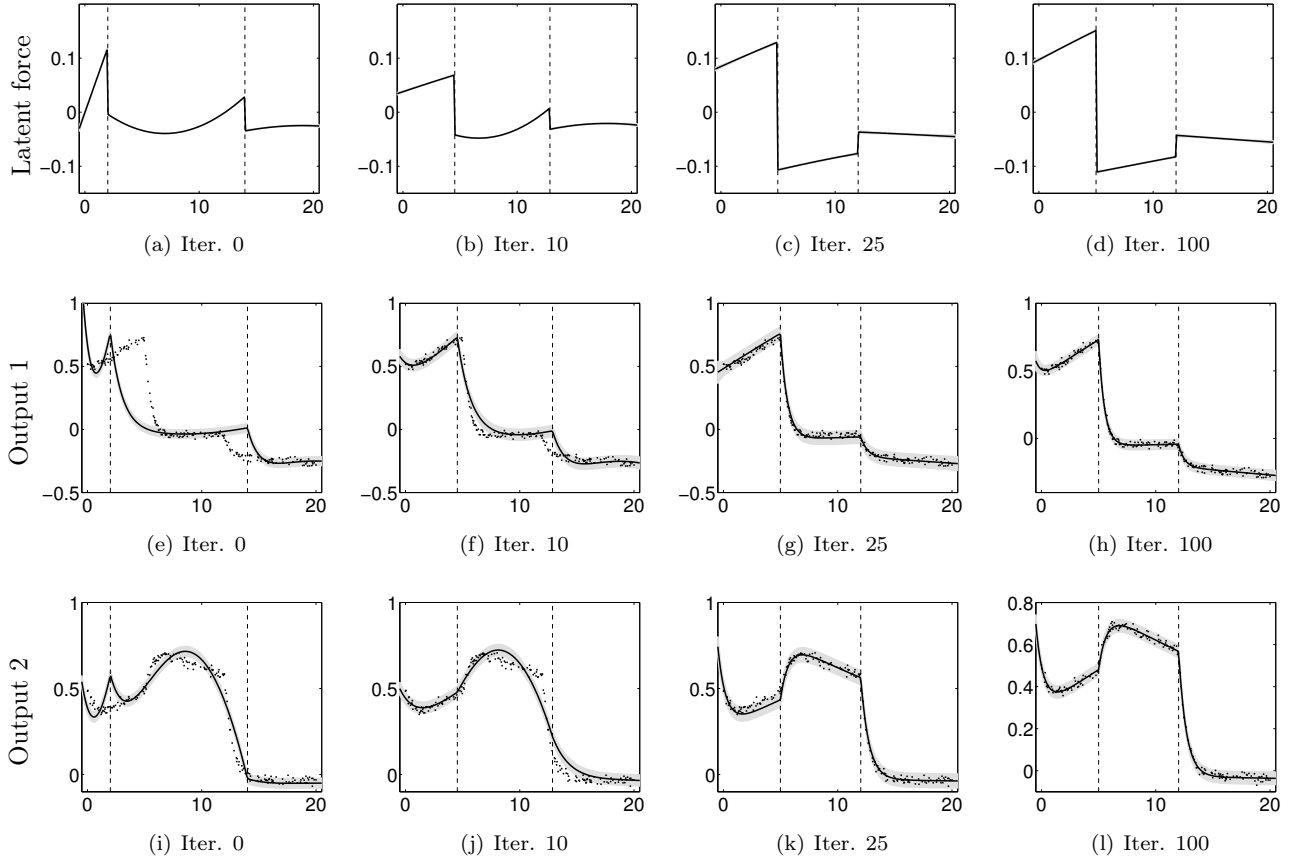
Figure 5: Toy experiment 3. Mean and two standard deviations for the predictions over the latent force and the two outputs in the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

4.1.4 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

10

(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
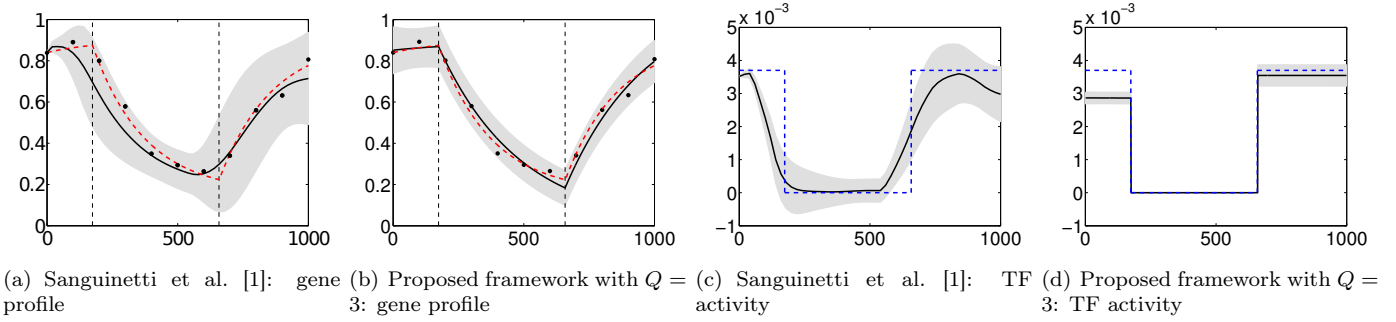
Figure 6: Toy experiment 4. Mean and two standard deviations for the predictions over the latent force and the output in the test set. Vertical dashed lines indicate the real value of the switching points. Dots indicate training data. Red dashed lines indicate test data. Blue dashed lines indicate synthetic 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.

have to be strictly positive, in this experiment we restricted the sensitivity parameters to be also positive.[2]

Employing both inference methods independently in this experiment, Figure 6 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 6(a) shows the reconstructed outputs using Sanguinetti et al. [1], and Figure 6(b) shows the result employing our proposed framework with $Q = 3$.[3] 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 6(c) and 6(d) show the scaled TF activity in the same order as we made for the outputs. From Figure 6(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 6(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.

4.2 Modelling transcriptional regulation results: E.coli dataset

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 3.4). 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.

E.coli is a robust organism that can adapt remarkably well to changes in its environment [9]. 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

---

[2]This assumption does not always guarantee that the latent force $u(t)$ be strictly a positive function. It is because the Gaussian process over the latent forces is not restricted to only generate positive functions [8]. However, this condition seems to be enough for our model.

[3]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 $Q = 3$.
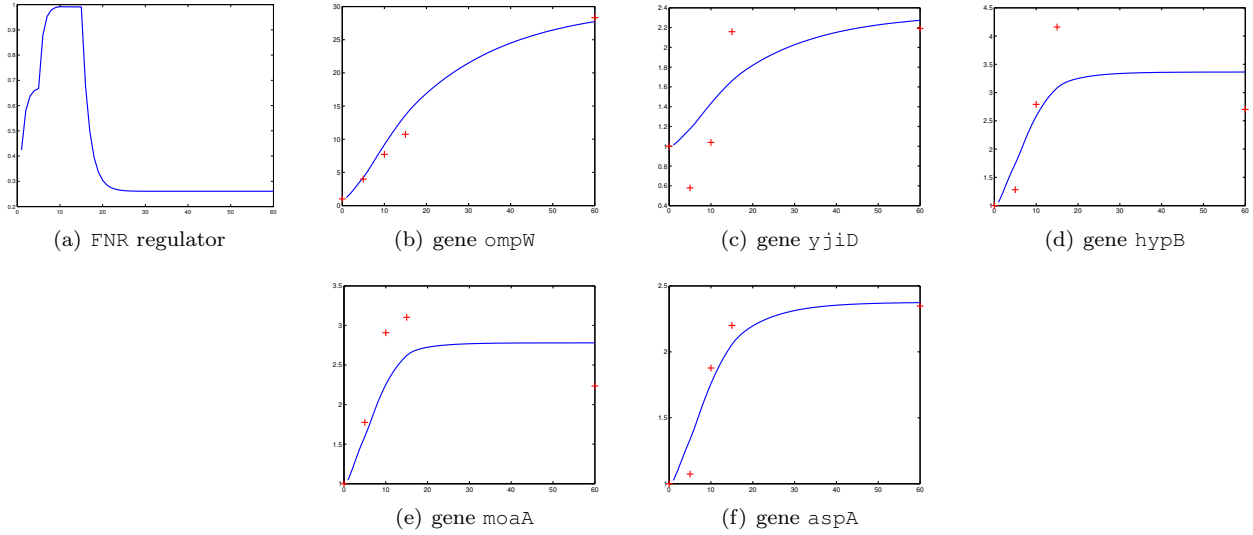
Figure 7: Microaerobic shift in E.coli results employing Sanguinetti et al [1]. Mean for the predictions over the FNR regulator and the genes profiles from the E.coli dataset. Red crosses indicate training data.
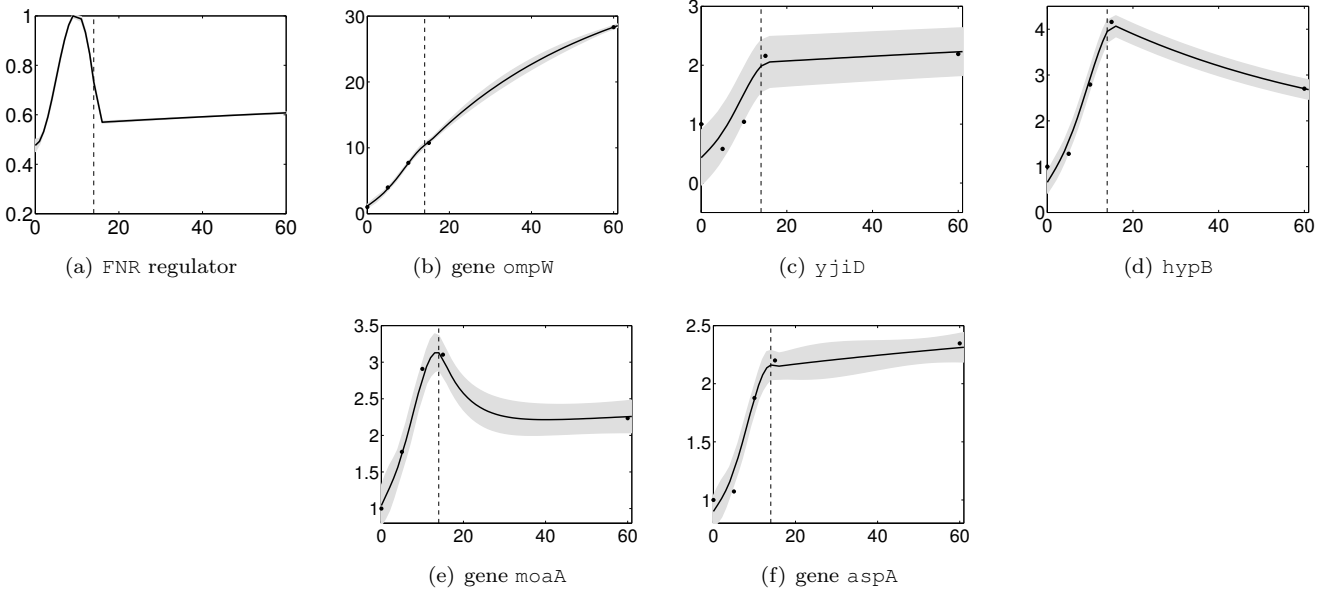

Figure 8: Microaerobic shift in E.coli results employing our proposed framework. Mean and two standard deviations for the predictions over the FNR regulator and the genes profiles in the test set. Dashed lines indicate the final value of the switching points after optimization. Dots indicate training data.

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 7 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 $t_1 = 39$.[4] However, we expect that the second estimated switching point be near to the 15 min where the measures are no 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 8 we show the results obtained employing our proposed framework after the inference procedure. According the figures 7(a) and 8(a), we notice the same problem we detailed in the experiment of section 4.1.4: 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.

---

[4]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 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 [10].

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

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 $u(t) = \exp\{\int_0^t p(\tau)d\tau\}$ [10]. According to this, the complete solution of the non-homogeneous condition is given by

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

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 (1), 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 $k_{y_d, y_{d'}}(t, t')$

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

where $\text{cov}\{y_d(0), f_{d'}(t', u)\} = \text{cov}\{f_d(t, u), y_{d'}(0)\} = 0$, 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 $\sigma_{y_d, y_{d'}} = \text{cov}\{y_d(0), y_{d'}(0)\}$ are entries of the covariance matrix K_{IC}. Covariance $k_{f_d, f_{d'}}$ is described further below.

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

$$k_{y_d,u}(t,t') = \text{cov}\{y_d(t), u(t')\}$$
$$= c_d(t)\,\text{cov}\{y_d(0), u(t')\} + \text{cov}\{f_d(t,u), u(t')\},$$

where $\text{cov}\{y_d(0), u(t')\} = 0$, 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 $k_{f_d,u}(t,t') = \text{cov}\{f_d(t,u), u(t')\}$. 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

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

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

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

where (using David's order for notation)

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

Then we have the following expression for the covariance

$$k_{f_d,f_{d'}}(t,t') = \frac{S_d S_{d'} \ell\sqrt{\pi}}{2}\exp\{-(\gamma_d t + \gamma_{d'} t')\}\underbrace{\left[h(\gamma_{d'},\gamma_d,t,t') + h(\gamma_d,\gamma_{d'},t',t)\right]}_{A},$$

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

$$A = \frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\left[\exp\{-\gamma_{d'}(t'-t)\}\mathcal{H}(\gamma_{d'},t',t) - \exp\{-(\gamma_d t + \gamma_{d'} t')\}\mathcal{H}(\gamma_{d'},t',0)\right]$$
$$+ \frac{\exp\{\nu_{d'}^2\}}{\gamma_d+\gamma_{d'}}\left[\exp\{-\gamma_d(t-t')\}\mathcal{H}(\gamma_d,t,t') - \exp\{-(\gamma_d t + \gamma_{d'} t')\}\mathcal{H}(\gamma_d,t,0)\right]$$
$$= \hat{h}(\gamma_{d'},\gamma_d,t,t') + \hat{h}(\gamma_d,\gamma_{d'},t',t),$$

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

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

where

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

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

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

15

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

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

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

where $k_{u,u}(t,t') = \text{cov}\{u(t), u(t')\} = \exp\{(t-t')^2/\ell^2\}$. 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

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

Then we have the following expression for the covariance

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

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

C   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 $\boldsymbol{\theta}$ [11]. Using some properties of the derivatives of matrices [12], we obtain

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

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 D, 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 E, we compute the derivatives of the kernel w.r.t. the other hyper-parameters (e.g. decay rate and length-scales).

D   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

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

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

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

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

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

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

Derivative for $\Upsilon_t^q(\gamma_{d'}, t', t)$ follows

$$\Upsilon_t^q(\gamma_{d'}, t', t) = \gamma_{d'} \Upsilon^q(\gamma_{d'}, t', t) - \frac{2}{\sqrt{\pi} \ell_q} \exp\left\{ -\left(\frac{t'-t}{\ell_q}\right)^2 \right\} + \frac{2}{\sqrt{\pi} \ell_q} \exp\{-\gamma_{d'} t'\} \exp\left\{ -\left(\frac{t}{\ell_q}\right)^2 \right\}.$$

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

$$\Upsilon_t^q(\gamma_{d'}, t, t') = -\gamma_{d'} \Upsilon^q(\gamma_{d'}, t, t') + \frac{2}{\sqrt{\pi} \ell_q} \exp\left\{ -\left(\frac{t-t'}{\ell_q}\right)^2 \right\}.$$

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

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

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

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

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

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

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

$$\Upsilon_{tt'}^{q}(\gamma_{d'},t',t) = -\gamma_{d'}\Upsilon_{t}^{q}(\gamma_{d'},t',t) + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t'-t)}{\ell_q^2}\exp\left\{-\left(\frac{t'-t}{\ell_q}\right)^2\right\},$$

$$\Upsilon_{tt'}^{q}(\gamma_d,t,t') = \gamma_d\Upsilon_{t}^{q}(\gamma_d,t,t') + \frac{2}{\sqrt{\pi}\ell_q}\frac{2(t-t')}{\ell_q^2}\exp\left\{-\left(\frac{t-t'}{\ell_q}\right)^2\right\} - \frac{2\gamma_d}{\sqrt{\pi}\ell_q}\exp\{-\gamma_d t\}\exp\left\{-\left(\frac{t'}{\ell_q}\right)^2\right\}.$$

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.

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

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

E 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

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

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

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

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

$$\frac{\partial}{\partial \ell_q} \Upsilon(\gamma_{d'}, t', t) = 2\nu_{qd'} \Upsilon(\gamma_{d'}, t', t) \frac{\partial \nu_{qd'}}{\partial \ell_q} - \exp\{\nu_{qd'}^2\} \exp\{-\gamma_{d'}(t' - t)\}$$

$$\times \left[\left(\frac{t' - t}{\ell_q^2} + \frac{\partial \nu_{qd'}}{\partial \ell_q} \right) \exp\left\{ -\left(\frac{t' - t}{\ell_q} - \nu_{qd'} \right)^2 \right\} + \left(\frac{t}{\ell_q^2} - \frac{\partial \nu_{qd'}}{\partial \ell_q} \right) \exp\left\{ -\left(\frac{t}{\ell_q} + \nu_{qd'} \right)^2 \right\} \right],$$

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 $\Upsilon(\gamma_{d'}, t', t)$.

References

[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] 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.

[3] 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.

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

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

[9] 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.

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

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

[12] 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.