Modelling transcriptional regulation with Gaussian processes

Vinny Davies and **Dirk Husmeier**
School of Mathematics and Statistics
University of Glasgow, United Kingdom

February 22, 2013

Abstract

A challenging problem in systems biology is the quantitative modelling of transcriptional regulation. Transcription factors (TFs), which are the key proteins at the centre of the regulatory processes, may be subject to post-translational modification, rendering them unobservable at the mRNA level, or they may be controlled outside of the subsystem being modelled. In both cases, a mechanistic model description of the regulatory system needs to be able to deal with latent activity profiles of the key regulators. A promising approach to deal with these difficulties is based on using Gaussian processes to define a prior distribution over the latent TF activity profiles. Inference is based on the principles of non-parametric Bayesian statistics, consistently inferring the posterior distribution of the unknown TF activities from the observed expression levels of potential target genes. The present work provides explicit solutions to the differential equations needed to model the data in this manner, as well as the derivatives needed for effective optimisation. The work further explores identifiability issues not fully shown in previous work and looks at how this can cause difficulties with inference. We subsequently look at how the method works on two different TFs, including looking at how the model works with a more biologically realistic mechanistic model. Finally we analyse the effect of more biologically realistic non-Gaussian noise on the biologically realistic model showing how this can cause a reduction in the accuracy of the inference.

1 INTRODUCTION

Quantitative modelling of transcriptional regulation is a highly topical and challenging research problem in contemporary systems biology. Over the last few years, a whole plethora of methods have been developed that aim to infer transcriptional regulatory networks from gene expression and DNA sequence data; see work by Lin and Husmeier [9] for a concise review. A particular challenge is that transcription factors (TFs), the regulatory proteins at the heart of transcriptional regulation, are frequently subject to post-translational modification, which may affect their DNA binding capability. Consequently, gene expression levels of TFs contain only limited information about their actual activities. Various authors have applied latent variable and factor analysis models to infer

genome-wide patterns of TF-gene interactions from high-throughput transcription profiles while allowing for the fact that the concentrations of active TFs may actually be unknown [9, 11, 13, 14].

An alternative and complementary research objective is to develop less abstract mechanistic models, based on differential equations, for smaller biochemical systems. Recent advances in computational statistics have demonstrated that kinetic parameters and interaction structures can be reliably inferred from completely observed concentration profiles of the molecular components involved [17]. However, the application of these methods to transcriptional regulation is impeded by the fact that concentrations of active TFs are often not directly observable, as discussed above. In addition, the dynamics of TF activities may be controlled outside of the subsystem being modelled, e.g. regulated by external protein signalling pathways. In both cases, the dynamics of TFs must be inferred indirectly, based on the response of the genes whose expression they control.

A promising approach to deal with these difficulties was proposed by Gao et al. [4], inspired by the work of Barenco et al. [1]. The authors advocate the use of Gaussian processes to define prior distributions over the latent TF activity profiles. Inference is soundly based on the principles of non-parametric Bayesian statistics, consistently inferring the posterior distribution of the unknown TF activities from the observed expression levels of potential target genes, and inferring regulatory network structures after marginalizing over the unknown TF activity profiles. The work has more recently been extended to incorporate a model of translation, which improves the quality of TF profile inference when TFs are primarily regulated at the transcriptional level [5], and to generalize the approach to multiple co-regulating transcription factors [16].

Problems with this method however come in the form of identifiability issues. Using a Gaussian process prior over the TF due to data being unobservable leads to a large level of flexibility within the estimates of the TF profile. Under these modelling assumptions some of the characteristic of the TF profile can be incorporated into the parameters of the model leading to a misestimation of the TF profile. We show how this is possible and show how not correcting for this can lead to problems with the inference.

The choice of a non-parametric prior distribution from the Gaussian process family is not a restrictive modelling assumption, as Gaussian processes have been shown to possess universal approximation capability [10]. Somewhat more restrictive is the assumption of Gaussian noise, which can be found in all previous applications [4, 5, 16]. A common approach is to assume that transcriptional data follows a log-normal distribution and to subject transcription profiles to a logarithmic transformation. However, Durbin et al. [3] and Huber et al. [6] showed that mRNA concentrations obtained from microarray experiments are only asymptotically log-normally distributed, in the limiting case of high transcription rates, and that the noise distribution for intermediate rates is of a more complex form. We look at how this can affect the inference of the model proposed by Gao et al. [4] which assumes an idealised Gaussian noise.

Table 1: Notation description

Symbols	Description
G	number of genes
T	number of time points
$x_i(t)$	true gene expression level of gene i
$y_i(t)$	noisy measurements of gene expression level of gene i
B_i	basal transcription rate of gene i
S_i	sensitivity to binding of TF of gene i
D_i	decay rate of gene i
$\boldsymbol{\theta}_c$	hyper parameters of the Gaussian process prior
$\boldsymbol{\theta}'$	parameters in covariance function $\mathbf{K}_{\mathbf{f},\mathbf{f}}$
$\boldsymbol{\theta}$	parameters in eq. (5)
$\mathbf{f}$	TF activity

2 METHODOLOGY

An overview of the notation used in the following exposition can be found in Table 1. A linear model of gene expression was proposed by Barenco et al. [1]

$$\frac{dx_i(t)}{dt} = B_i + S_i f(t) - D_i x_i(t) \tag{1}$$

where $i \in \{1, \ldots, G\}$ is a set of genes regulated by the same transcription factor TF, $x_i(t)$ are the (unknown) true gene expression levels at time point t, $f(t)$ is the (unknown) TF activity, B_i is the basal transcription rate of gene i, S_i is the sensitivity to binding of TF, and D_i is a decay rate. We assume that (noisy) measurements of $x_i(t)$ can be obtained, and $f(t)$ is unobservable. Eq. (1) has the analytic solution

$$x_i(t) = \frac{B_i}{D_i} + S_i \int_0^t \exp(-D_i(t-u)) f(u) du \tag{2}$$

where transient terms have been ignored.

2.1 GP Inference

Lawrence et al. [8] and Gao et al. [4] proposed a non-parametric Bayesian approach to inference in this model by placing a Gaussian process prior on the unknown TF activities $\mathbf{f} = (f(t_1), \ldots, f(t_T))$ at time points $\mathbf{t} = (t_1, \ldots, t_T)$

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}}) \tag{3}$$

that is, the prior probability of the TF activities, $p(\mathbf{f})$, is a zero-mean multivariate Gaussian distribution with covariance matrix $\mathbf{K}_{\mathbf{f},\mathbf{f}}$. The linear form of eq. (2) implies that the joint prior distribution of the expression profiles of all regulated genes

$$\mathbf{x}_i = [x_i(t_1), \ldots, x_i(t_T)]; i = 1, \ldots, G \tag{4}$$

is described by a Gaussian process prior with a covariance matrix, $\mathbf{K}$, that depends on the hyper parameters $\boldsymbol{\theta}_c$ and the parameters that characterise the transcriptional regulation processes via eq. (1):

$$p(\mathbf{x}_i|\boldsymbol{\theta}') = \mathcal{N}\left(\frac{\mathbf{B_i}}{\mathbf{D_i}}, \mathbf{K}\right); \quad \mathbf{K} = \mathbf{K}(\boldsymbol{\theta}') \tag{5}$$
$$\boldsymbol{\theta}' = (\boldsymbol{\theta}_c, B_1, \ldots, B_G, S_1, \ldots, B_G, D_1, \ldots, D_G)$$

where $\mathbf{B_i}/\mathbf{D_i}$ is divided in a point-wise manner. See [4, 5, 8] for explicit expressions.

In order to derive an expression for the covariance matrix $\mathbf{K}$, we firstly define the linear operator relating to the latent function f and the gene expression level $x_j(t)$ as shown by Lawrence et al. [8]

$$L_j[f](t) = S_j \exp(-D_j t) \int_0^t f(u) \exp(D_j u) du. \tag{6}$$

This contains all the time varying elements of eq. (2) and implies a covariance of

$$\operatorname{cov}(L_j[f](t), L_k[f](t')) = L_j \otimes L_k[k_{f,f}](t, t'). \tag{7}$$

This ultimately results in the elements of the covariance matrix being expressed as

$$\begin{aligned} k_{x_j,x_k}(t,t') = S_j S_k \exp(-D_j t - D_k t') \int_0^t \exp(D_j u) \\ \times \int_0^{t'} \exp(D_k u') k_{f,f}(u,u') du' du \end{aligned} \tag{8}$$

where $k_{f,f}(t,t')$ is the covariance matrix associated with $f(t)$. If we then let this be the Radial Basis Function (RBF) kernel, such as given in eq. (9), where l is the length scale and a^2 is the amplitude parameter, then the integral becomes tractable and can be solved using Laplace transformations (see Appendix A).

$$k_{f,f}(t,t') = a^2 \exp\left(-\frac{(t-t')^2}{l^2}\right). \tag{9}$$

From Appendix A, the resulting covariance elements are given by

$$k_{x_j,x_k}(t,t') = a^2 S_j S_k \frac{\sqrt{\pi} l}{2} [h_{kj}(t',t) + h_{jk}(t,t')] \tag{10}$$

where

$$\begin{aligned} h_{kj}(t',t) = \frac{\exp(\gamma_k^2)}{D_j + D_k} \Bigg\{ \exp[-D_k(t'-t)] \left[\operatorname{erf}\left(\frac{t'-t}{l} - \gamma_k\right) + \operatorname{erf}\left(\frac{t}{l} + \gamma_k\right) \right] \\ - \exp[-(D_j t + D_k t')] \left[\operatorname{erf}\left(\frac{t'}{l} - \gamma_k\right) + \operatorname{erf}(\gamma_k) \right] \Bigg\}. \end{aligned}$$

Here we define $\gamma_k = \frac{D_k l}{2}$ and $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy$. This then allows us to compute the likelihood dependent on parameters $\boldsymbol{\theta}'$.

In order to infer the unknown TF activity, a covariance matrix must be found relating $x_j(t)$ and $f(t)$. This is similarly given as

$$k_{x_j,f}(t,t') = S_j \exp(-D_j t) \int_0^t \exp(D_j u) k_{f,f}(u,t') du. \tag{11}$$

Again when the RBF kernel is used the integral is solvable with a Laplace transformation (see Appendix B), with the covariance matrix given as

$$\begin{aligned} k_{x_j,f}(t,t') = a^2 S_j \frac{\sqrt{\pi} l}{2} \exp(\gamma_j^2) \exp[-D_j(t-t')] \\ \times \left[\operatorname{erf}\left(\frac{t-t'}{l} - \gamma_j\right) + \operatorname{erf}\left(\frac{t'}{l} + \gamma_j\right) \right]. \end{aligned}$$

Standard regression techniques, such as shown by Rasmussen and Williams [12], then lead to f being modelled such that

$$f|\mathbf{x} \sim \mathcal{N}(K_{f,\mathbf{x}}K_{\mathbf{x},\mathbf{x}}^{-1}(\mathbf{x} - \tfrac{\mathbf{B}}{\mathbf{D}}), K_{f,f} - K_{f,\mathbf{x}}K_{\mathbf{x},\mathbf{x}}^{-1}K_{\mathbf{x},f}) \tag{12}$$

where $\mathbf{x}$ denotes the vector of observed variables all genes, $\mathbf{K}$ is the matrix which arises from evaluating the covariance functions between all time points and $\mathbf{B}/\mathbf{D}$ is the point-wise division of the vectors of $\mathbf{B}$ and $\mathbf{D}$ where the parameters come from the ODE of the genes they correspond to.

To relate the unknown true gene expression profiles $\mathbf{x}_i$ to noisy measurements $\mathbf{y}_i = [y_i(t_1), \ldots, y_i(t_T)]; i = 1, \ldots, G$, the standard approach (e.g. [2], Section 6.4.2) assumes additive Gaussian noise of constant variance σ^2:

$$p(\mathbf{y}|\mathbf{x}, \sigma^2) \;=\; \mathcal{N}(\mathbf{y}|\mathbf{x}, \sigma^2\mathbf{I}) \tag{13}$$

where $\mathbf{I}$ is the identity matrix. The marginalisation over $\mathbf{y}$ is analytically tractable and gives, with the definition $\boldsymbol{\theta} = (\boldsymbol{\theta}', \sigma^2)$:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}) \;&=\; \int p(\mathbf{y}|\mathbf{x}, \sigma^2)p(\mathbf{x}|\boldsymbol{\theta}')d\mathbf{x} \;=\; \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})) \\ \mathbf{C}(\boldsymbol{\theta}) &= \mathbf{K}(\boldsymbol{\theta}') + \sigma^2\mathbf{I} \end{aligned} \tag{14}$$

giving a log likelihood of

$$\ln(\mathbf{y}|\boldsymbol{\theta}) = -\tfrac{T}{2}\ln(2\pi) - \tfrac{1}{2}\ln|\mathbf{C}(\boldsymbol{\theta})| - \left(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}\right)^\top \mathbf{C}(\boldsymbol{\theta})^{-1}\left(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}\right). \tag{15}$$

Inference of the parameters $\boldsymbol{\theta}$ can then be achieved in a maximum likelihood or Bayesian framework, as described in standard textbooks on Gaussian process modelling [2, 12]. For non-linear optimisation methods, the gradient of the log likelihood with respect to the parameters $\boldsymbol{\theta}$ helps to improve the optimisation and these are derived in Appendix C.

Implementing these methods is often harder than suggested within the original papers from which they were proposed. For instance when inferring parameters within either a maximum likelihood or Bayesian framework, problems can also occur in the form of ill-conditioned matrices. Errors within the computation can lead to spikes in the likelihood in the tails of the distribution and it is necessary to specify that the likelihood should be reduced to zero in the case of the covariance matrix having small or negative eigenvalues, where the negative eigenvalues are due to a limited computational precision.

2.2 Identifiability Issues

The identifiability issues within the model come from the original make up of eq. (1) proposed by Barenco et al. [1] and the fact that data from the TF is unobservable. With only data from the downstream genes and no data from the TF activity, the Gaussian process effectively learns the shape of the TF, but can fail to accurately estimate the baseline activity and exact concentration levels of the profile. This can lead to problems as the method will still continue to give accurate estimates for the Gene profiles, so it may be naturally expected that the TF profile is similarly accurate. We show how this is not always the case and similarly how this can affect some of the estimates of the model parameters.

The misestimation can happen in two ways and we show how this can happen by looking at $\tilde{f}(t)$, the unscaled profile of the TF, as was shown by Barenco et al. [1]. We can think about the true TF, $f(t)$, as being related to this unscaled profile through eq. (16) where additive and multiplicative terms, α and β respectively, have been introduced;

$$f(t) = \alpha + \beta\tilde{f}(t). \tag{16}$$

We show below how this can be manipulated into the form of eq.(1), where the $g(t)$ can be seen as the TF and incorrect parameters can be learnt for the ODEs.

$$\begin{aligned} \frac{dx_i(t)}{dt} &= B_i + S_i f(t) - D_i x_i(t) \\ &= B_i + S_i\{\alpha + \beta\tilde{f}(t)\} - D_i x_i(t) \\ &= \{B_i + S_i\alpha\} + \{S_i\beta\}\tilde{f}(t) - D_i x_i(t) \\ &= \tilde{B}_i + \tilde{S}_i\tilde{f}(t) - D_i x_i(t). \end{aligned} \tag{17}$$

The problem relating from this is that the scaling parameters α and β cannot be obviously learnt, so to address this issue we fix some of the parameters, B_1 and S_1, to a set value as close as possible to their true value. Although this is not ideal, it is biologically possible, and will ensure we get accurate estimates for the TF profile.

2.3 Noise Assumptions

Previous publications, such as [4, 5, 16], have assumed white additive Gaussian noise of the form of eq. (13). This assumption leads to a closed-form solution of the convolution integral given in eq. (14). The result is a simple and straightforward modification of the noise-free scenario, in which the noise variance is added to the diagonal elements of the covariance matrix.

Unfortunately it has been shown by both Durbin et al. [3] and Huber et al. [6] that transcriptional data are subject to noise of a more complex form;

$$\begin{aligned} y_i(t) &= c + x_i(t)\exp(\mu_t) + \varepsilon_t \\ \mu_t &\sim \mathcal{N}(0, \sigma_\mu^2);\ \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned} \tag{18}$$

Maximising the log likelihood under the assumption of white additive noise is equivalent to minimising the residual sum of squares difference between the observations and the interpolant, which is equivalent to finding the *mean*. However,

$$\begin{aligned} y_i(t) &= x_i(t)\exp(\mu_t) + \varepsilon_t \\ \mathbb{E}[y_i(t)] &= x_i(t)\mathbb{E}[\exp(\mu_t)] + \mathbb{E}[\varepsilon_t] \\ &= x_i(t)\mathbb{E}[\exp(\mu_t)] \\ &\geq x_i(t)\exp(\mathbb{E}[\mu_t]) = x_i(t)\exp(0) = x_i(t) \end{aligned} \tag{19}$$

by Jensen's inequality. Replacing the true signal $x_i(t)$ in eq. (19) by the smoothed signal $\tilde{x}_i(t) = \mathbb{E}[y_i(t)]$ from the Gaussian interpolation will lead to a

systematic distortion as a consequence of the overestimation $\tilde{x}_i(t) \geq x_i(t)$. This implies that

$$\frac{d\tilde{x}_i(t)}{dt} = B_i + S_i f(t) - D_i \tilde{x}_i(t) \leq B_i + S_i f(t) - D_i x_i(t) = \frac{dx_i(t)}{dt} \qquad (20)$$

will be systematically underestimated. We will look at how in practise this incorrect assumption of Gaussian noise effects the predictions of the model parameters and the profiles of both the downstream genes and also the TF.

3 EVALUATION ON SIMULATED DATA

To assess the model performance, we carried out evaluation procedures on two known forms of TF profile. The first TF profile is in the form of a Gaussian mixture model while the second is based on a far more biologically realistic TF profile based on Michaelis-Menten kinetics. Here, we exploit the fact that the ground truth is known and that the performance of the model can therefore be objectively assessed.

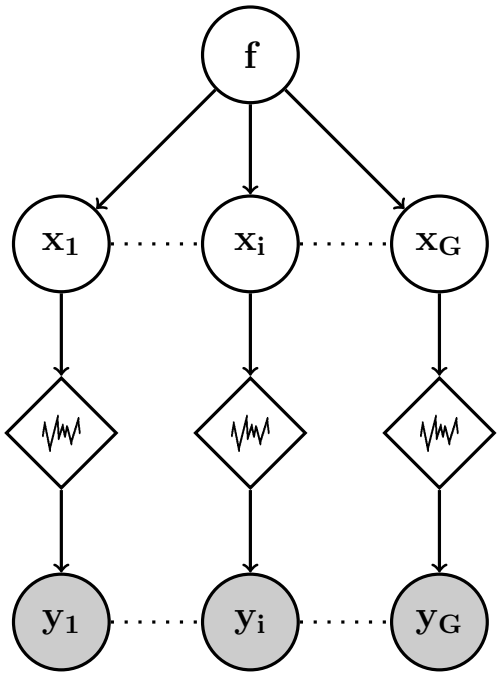


Figure 1: **TF simulated from a synthetic system**. The active form of the TF is given by f and regulates G genes with expression values $x_1,\ldots,x_i,\ldots,x_G$, whose noisy measurements are $y_1,\ldots,y_i,\ldots,y_G$.

3.1 Data Generation

We tested the performance of the proposed scheme on data simulated from the regulatory networks shown in Figure 1. A single TF regulates three downstream

genes linearly according to eq.(2). We generate data with both Gaussian noise, such as in Gao et al. [4], and also with more realistic non-Gaussian noise from eq. 18, i.e. using the model proposed in Durbin et al. [3]. The simulations require the specification of the latent TF profile.

We approached this in two different ways. In the first approach, we followed Lawrence et al. [7] and chose the TF profile to have the following fixed functional form:

$$f(t) = \sum_{j=1}^{4} a_j \exp\left(-\frac{(t-\mu_j)^2}{\sigma^2}\right) \tag{21}$$

For the parameters, we chose the same values as Lawrence et al.[7], namely $\sigma = 1.5, a_1 = a_2 = 1.5, a_3 = a_4 = 0.5,\ \mu_1 = 4, \mu_2 = 6, \mu_3 = 8.5$ and $\mu_4 = 10.5$. For our empirical evaluation studies, we generated the TF signal between time units 0 and 100, out of which we sampled 18 values regularly spaced in $[0, 100]$.

In the second approach, we chose a more realistic simulation, where we explicitly modelled the post-translational modification of the TF. To this end, we combined the regulatory network of the previous figure, Figure 1, with a simple protein signalling pathway taken from work by Vyshemirsky and Girolami [17]; see Figure 2. The active form of the TF is Rpp, which is derived from the native form of the protein, R, via phosphorylation. The phosphorylation is catalysed by the enzyme S via formation of the dimer RS, where the enzyme S can irreversibly decay into its degraded form dS. The kinetics of these processes are described by the following system of mass action and Michaelis-Menten type differential equations:

$$\begin{aligned}
\frac{d[S]}{dt} &= -k_1 \cdot [S] - k_2 \cdot [S] \cdot [R] + k_3 \cdot [RS] \\
\frac{d[dS]}{dt} &= k_1 \cdot [S] \\
\frac{d[R]}{dt} &= -k_2 \cdot [S] \cdot [R] + k_3 \cdot [RS] + \frac{V \cdot [Rpp]}{k_m + [Rpp]} \\
\frac{d[RS]}{dt} &= k_2 \cdot [S] \cdot [R] - k_3 \cdot [RS] - k_4 \cdot [RS] \\
\frac{d[Rpp]}{dt} &= k_4 \cdot [RS] - \frac{V \cdot [Rpp]}{k_m + [Rpp]}
\end{aligned} \tag{22}$$

where the square brackets, [.], denote concentrations, and the kinetic parameters were set in the same way as Vyshemirsky and Girolami [17]: $k_1 = 0.07, k_2 = 0.6, k_3 = 0.05, k_4 = 0.3, V = 0.017, k_m = 0.3$.

In this way we emulate the scenario described in the Introduction, where the TF activity profiles are unobservable due to post-translational modification, and the processes leading to the formation of active TFs are controlled outside of the subsystem being modelled (which only consists of the TF and the three downstream genes). Again, we generated the TF signal between time units 0 and 100, and sampled 18 values regularly spaced in $[0, 100]$.

From the TF profiles, the three downstream genes are then generated by solving eq. (2) with the kinetic parameter values given in Table 2. All values are the same for both TFs with the exception of D_1 which was changed in order to give more reasonable profiles in the case of the second TF. In the case of the Gaussian mixture model, eq. (2) can be solved explicitly using similar techniques

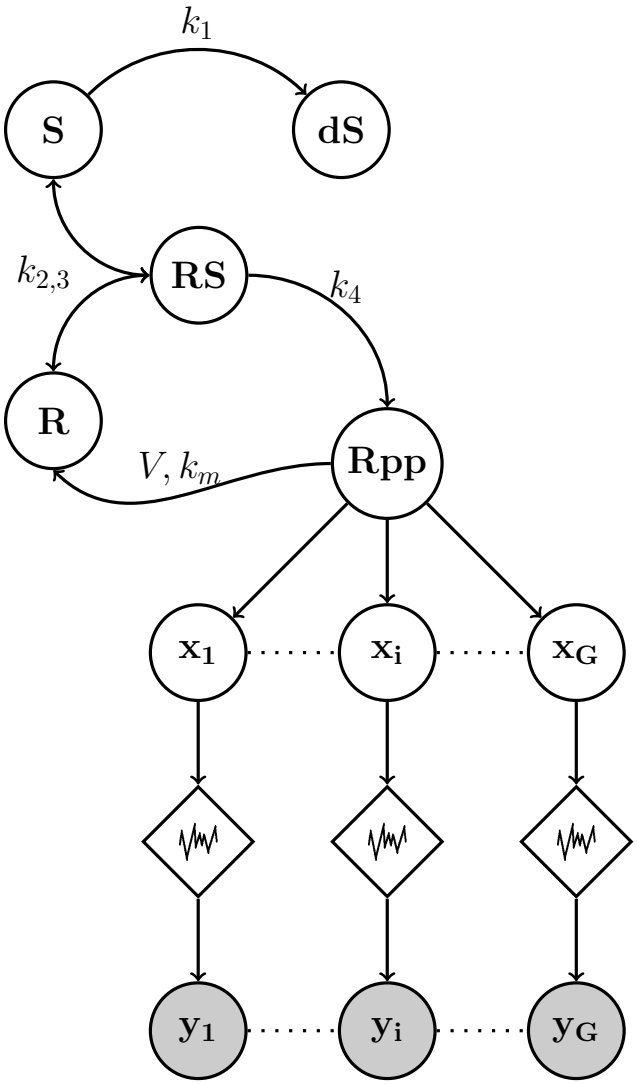


Figure 2: **Transcriptional regulation by a TF that is subject to post-translation modification**. The active form of the TF is Rpp, which is derived from the native form of the protein, R, via phosphorylation. The phosphorylation is catalysed by the enzyme S via formation of the dimer RS, where the enzyme S can irreversibly decay into its degraded form dS. The TF regulates G genes with expression values $x_1,\ldots,x_i,\ldots,x_G$, whose noisy measurements are $y_1,\ldots,y_i,\ldots,y_G$.

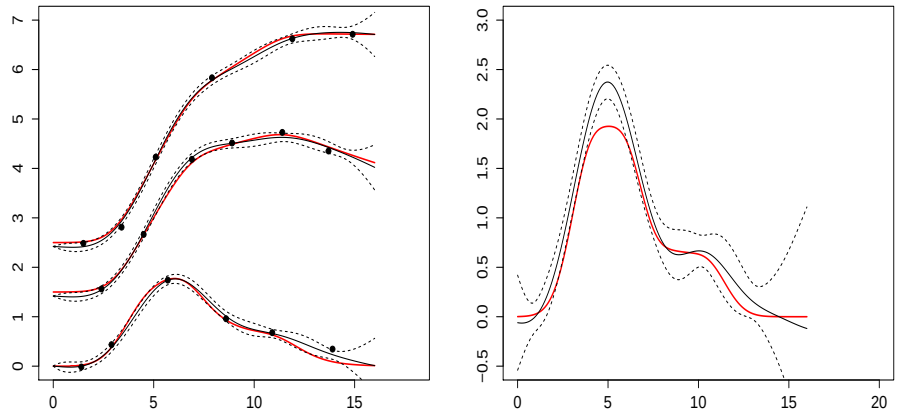
to those used for finding the covariance matrix. Where we explicitly modelled the post-translational modification of the TF standard numerical approaches such as Simpson's method can be used, where only grid values of the TF are required to find the true genes profiles. Data points can then be simulated for the downstream genes with either Gaussian or non-Gaussian noise of different sizes as required.

Table 2: Kinetic Parameter values used to generate Downstream Genes

	B_i	S_i	D_i
$i = 1$	0.0	1.0	1.0/0.2
$i = 2$	7.5×10^{-2}	0.4	5×10^{-2}
$i = 3$	2.5×10^{-3}	0.4	1×10^{-3}

3.2 Evaluation Procedure

We emulate a real experiment, in which only the expression time series of the downstream genes are observed, and the objective is to infer both the unknown TF activities as well as the kinetic parameters, i.e. the set of $\{B_i, S_i, D_i\}_{i=1,2,\ldots,G}$ with $G = 3$ in eq. (1). For the TF activity levels coming from Figure 2 we simulated 10 independent data sets generated under identical conditions for each data set, where each data set is given different values for σ_μ and σ_ε. In each case both standard deviations are given the same values and we use a roughly log scale so $\sigma_\mu, \sigma_\varepsilon = \{0.01, 0.03, 0.1, 0.3\}$. We then choose values for σ to be a numerical estimate of this standard deviation in Gaussian form, such as to give a fair comparison between methods. We represent the results in the form of box plots of the errors of the parameter and profile estimates.



(a) Downstream Genes (b) Transcriptional Factor

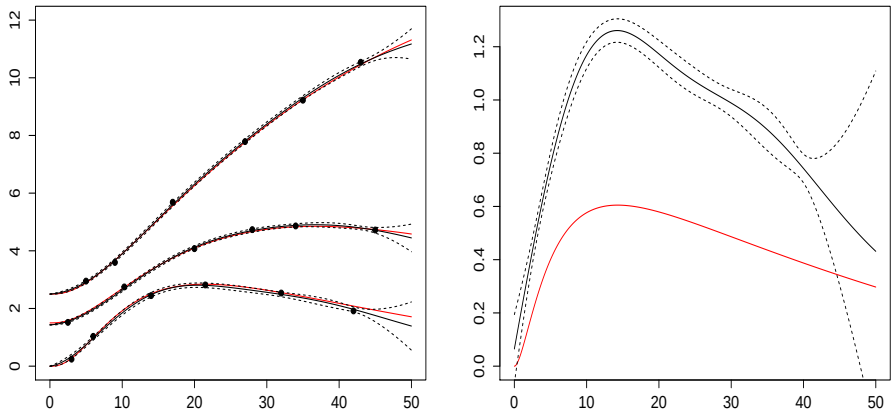
Figure 3: A summary of the expected distribution of data for the simulated (a) Downstream Genes and (b) Transcriptional Factor, based on the Gaussian mixture model TF. Solid red lines indicate the true mean of the distributions, the solid black lines give the best prediction and the dotted lines indicate the region within two standard deviations. Solid points indicate the observed data from the gene profiles. They are modelled with Gaussian noise.

4 SIMULATION RESULTS

We firstly explore the model for data with Gaussian noise similarly to Lawrence et al. [8] and Gao et al. [4], comparing this for both of the TFs and looking at how the predictions are affected by the issues with identifiability. We then use the data simulated with the more biologically realistic noise from eq. (18), looking at how this effects our ability to make predictions on the more realistic TF activity from Figure 2. We look at this over a variety of different values for σ_μ and σ_ε with the aim of comparing the effect of this as this values increases.

4.1 Gaussian Noise

Scientific interest leads to the desire to estimate both the original TF profile and the kinetic parameters of the down steam genes. Figure 3a shows the observed points available (in the form of black dots), with no points observed directly observable from the TF profile in Figure 3b. Using the Gaussian noise model as in Lawrence et al. [8] leads to an accurate estimate in this case with only a limited number of points. Estimates were found to vary in accuracy and be dependent on the number of observed points as well as their spacing and location. More points, evenly spaced will ultimately lead to the best prediction. However if points are not observed accurately or are not observed at crucial points on the profile then estimates can lose the shape of the profile. For instance in this case without accurate points on the bumps, inference can miss the double peak of the profile.

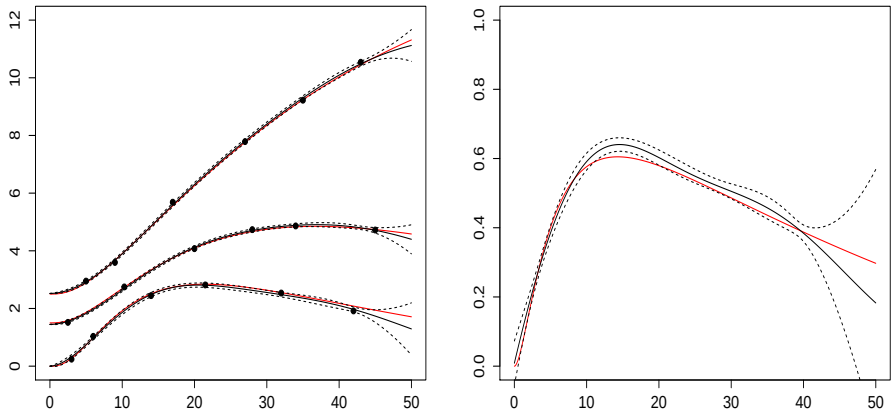


(a) Downstream Genes (b) Transcriptional Factor

Figure 4: A summary of the expected distribution of data for the simulated (a) Downstream Genes and (b) Transcriptional Factor, based on the TF that is subject to post-translation modification. Solid red lines indicate the true mean of the distributions, the solid black lines give the best prediction and the dotted lines indicate the region within two standard deviations. Solid points indicate the observed data from the gene profiles. They are modelled with Gaussian noise.

However further simulations on the TF activity profile generated for Figure 2 showed that there appeared to be identifiability issues with regards to predicting $f(t)$. This was further investigated and it was found that there were both issues with both of the TF activity profiles used in this paper. It was found that the accurate results given in Figure 4 were coincidental and it was still possible this model could still deviate away from the true profile of $f(t)$. In a more practical sense this is a worrying problem as it can mean that predictions can easily be interpreted as accurate when this is not the case.

This was further investigated and it can be seen in eq. (17) and eq. (19) how these identifiability issues occur. In practical terms it comes from the original form of the ODEs given in eq. (1), where part of the size and shape of $f(t)$ can be easily absorbed into the parameter as seen in eq. (17). For this reason we choose to fix both B_1 and S_1 to their true values. The effect of this can be seen in the difference between Figures 4 and 5. Although in both Figures the profiles of the downstream gene activity can be estimated accurately, in Figure 4b we can see how this is no longer the case with the TF profile. Fixing B_1 and S_1 fixes this problem, as shown in Figure 5b, and these parameters remain fixed for the remainder of this paper.

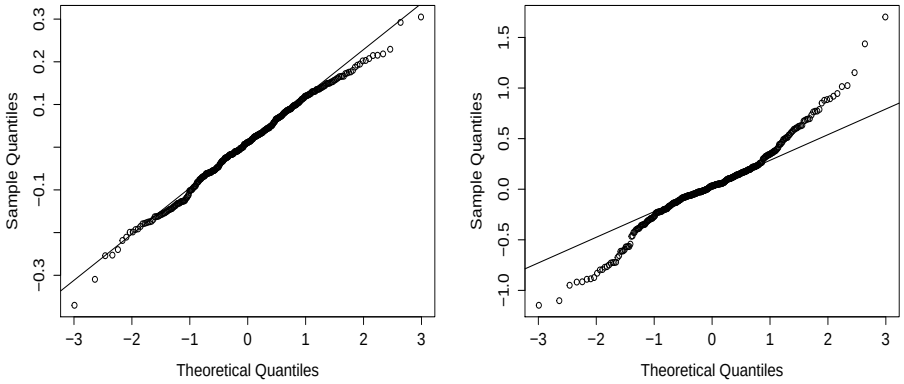


(a) Downstream Genes (b) Transcriptional Factor

Figure 5: A summary of the expected distribution of data for the simulated (a) Downstream Genes and (b) Transcriptional Factor, based on the TF that is subject to post-translation modification. Solid red lines indicate the true mean of the distributions, the solid black lines give the best prediction and the dotted lines indicate the region within two standard deviations. Solid points indicate the observed data from the gene profiles. They are modelled with Gaussian noise and in this case the parameters $S_1 = 1$ and $B_1 = 0$ are fixed.

4.2 Non-Gaussian Noise

Durbin et al. [3] and Huber et al. [6] showed that transcriptional data is subject to noise of the form given in eq. 18. However previous work, [4, 5, 16], has been completed under the assumption of Gaussian noise to make the method tractable. But intuitively there is a significant difference between these types of noise and this can be seen in the form of QQ-plots given in Figure 6. In Figure 6b the non-Gaussian noise from eq. 18 shows considerable deviations from Gaussian noise, violating the modelling assumptions.

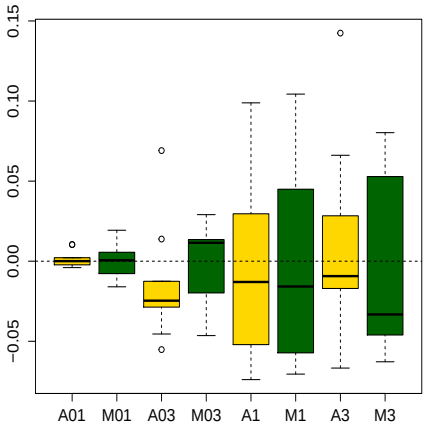


(a) Gaussian Noise (b) Non-Gaussian Noise

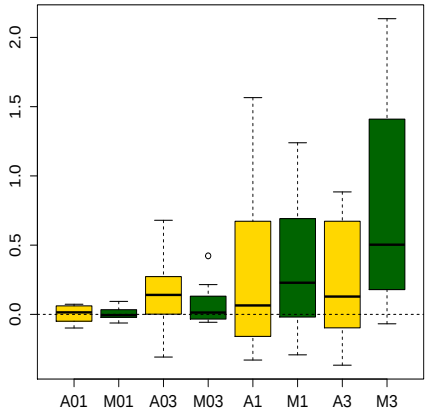
Figure 6: QQ-plots showing the deviations from Gaussian error for data generated with Gaussian and non-Gaussian noise.

Figure 7 gives box plots of the errors in the estimation of the parameters from eq. (1). It shows that there is very little difference between the estimates based on data with Gaussian and non-Gaussian noise when the standard deviation is low. However when the standard deviation becomes greater we find that the non-Gaussian noise begins to have a far more negative influence on the inference. Often the estimates have a strong bias away from the true parameter values and there also appears to be an increase in the variance of the estimates over the situation where we have Gaussian noise. This could be partially down to use of numerical estimates for the variance of the equivalent Gaussian noise model, however part of this increase is still likely to come from the noise model.

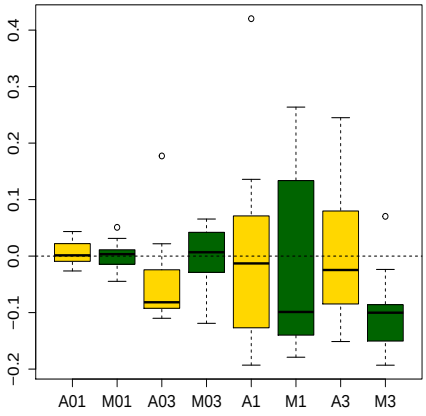
Figure 8 shows box plots illustrating the accuracy of estimates for the gene and TF profiles generated from the model based on Michaelis-Menten kinetics. The results show that for the gene profile, x_1 in Figure 8a, where the expression levels are low, estimates are very similar and the differences occur relatively randomly. For the genes with higher expression levels however, the differences are very much more pronounced and our results have shown that the non-Gaussian noise can cause a strong bias within the estimates, where this effect is exaggerated when the standard deviation get larger. Interestingly Figure 8c shows only an increased variance for the non-Gaussian noise and only a small bias when the standard deviation is increased. This seems unusual as this would likely be the gene with the most differences in terms of the comparison between Gaussian and non-Gaussian noise and in hindsight more simulations should possibly be run to give more comprehensive results.



(a) Basel Transcription Rate, $\mathbf{B_i}$



(b) Sensitivity, $\mathbf{S_i}$

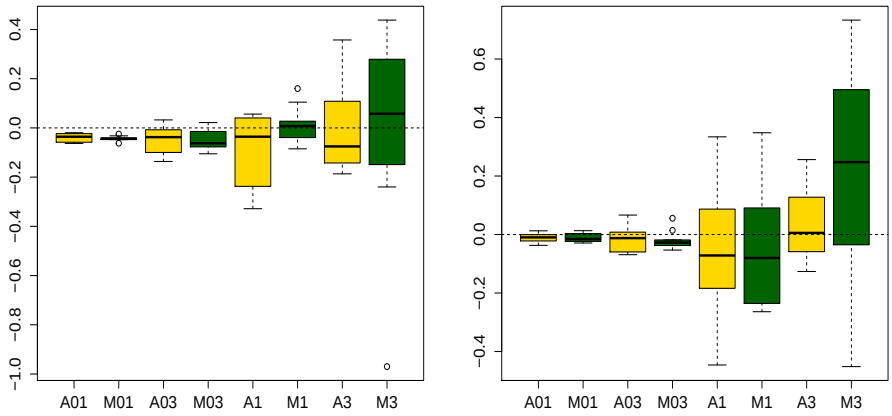


(c) Decay Rate, $\mathbf{D_i}$

Figure 7: Box plots of the error of ODEs parameters. Box plots for the Gaussian and non-Gaussian noise are given in yellow and green respectively, also indicated by 'A' and 'M' relating to where the noise is additive or multiplicative. The standard deviations used for σ_μ and σ_ε are given under each box plot and represent the values $\{0.01,0.03,0.1,0.3\}$.

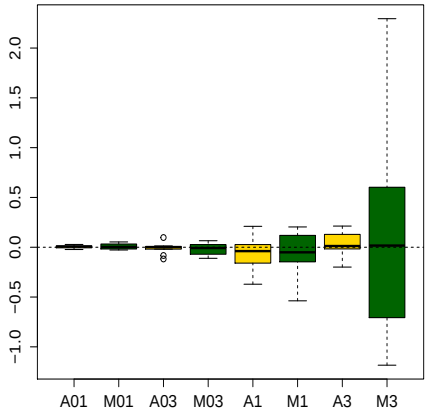
5 DISCUSSION AND FURTHER WORK

In this paper we have looked at the method proposed by Gao et al. [4] and shown how this can be used on a a more realistic TF profile based on Michaelis-Menten kinetics. We have looked at some of the issues associated with the method as well as providing explicit solutions to the equations required to fit the model. We have further looked at how a more realistic non-Gaussian noise proposed by Durbin et al. [3] affects the ability of the model to provide accurate parameter and profile estimates, showing how the estimates can give a strong bias when gene expression levels are high and this is not accounted for.

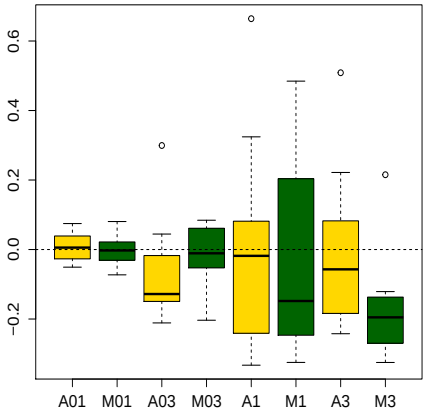


(a) Gene Profile, x_1

(b) Gene Profile, x_2



(c) Gene Profile, x_3



(d) TF Profile, f

Figure 8: Box plots of the error of gene and TF profiles. Box plots for the Gaussian and non-Gaussian noise are given in yellow and green respectively also indicated by 'A' and 'M' relating to where the noise is additive or multiplicative. The standard deviations used for σ_μ and σ_ε are given under each box plot and represent the values $\{0.01,0.03,0.1,0.3\}$

This noise form has been shown in a variety of microarray data experiments and ideally we would want to take account of this if possible. The work of Snelsen et al. [15] has shown potential for this through the use of the warped Gaussian process, however this method has proved difficult to implement when it is necessary to fix parameters due to the identifiability issues shown by Barenco et al. [1]. There is potential for finding a way of implementing this method into our current method, however the level of variance in microarray data must be taken into account when considering the worth of such work. Consultation with biologists in this field would provide an idea as to what level of variance are likely to be, at which point our results may help to give guidance as to whether such work is necessary.

Other further work in this area leads to looking at non-linear examples through the use of a MAP-Laplace approximation, such as in Lawrence et al. [8]. In this case eq. (1) and eq. (2) are extended to the form;

$$\frac{dx_i(t)}{dt} = B_i + S_i g_i(f(t)) - D_i x_i(t) \tag{23}$$

$$x_i(t) = \alpha_i \exp(-D_i t) + \frac{B_i}{D_i} + S_i \int_0^t \exp(-D_i(t-u)) g_i(f(u)) du \tag{24}$$

where $g(.)$ is a non-linear function of the concentration of the active TF. Note also that the transient terms now become necessary for this model.

One example of a biologically meaningful non-linear function is

$$g_i(f(t)) = \frac{\exp(f(t))}{\exp(f(t)) + \gamma_i} \tag{25}$$

where an additional parameter γ_i has to be estimated for each downstream gene. This helps to correct two key issues with the more simple linear model. Firstly it takes into account that the concentrations must always remain positive by modelling the log concentrations, meaning that $\exp(f(t))$ now models the concentration. Secondly it takes into account that the concentrations begin to become saturated at higher levels of TF activity.

References

[1] 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*, 7(3):R25, 2006.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Singapore, 2006.

[3] B. Durbin, J. Hardin, D. Hawkins, and D. Rocke. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, 18(suppl1):S105–S110, 2002.

[4] P. Gao, A. Honkela, M. Rattray, and N. D. Lawrence. Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. *Bioinformatics*, 24:i70–i75, 2008.

[5] A. Honkela, C. Girardot, E. H. Gustafson, Y.-H. Liu, E. E. M. Furlong, N. D. Lawrence, and M. Rattray. Model-based method for transcription factor target identification with limited data. *PNAS*, 107(17):7793–7798, 2010.

[6] W. Huber, A. von Heydebreck, H. Sultmann, A. Poustka, and M. Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(suppl1):S96–104, 2002.

[7] N. Lawrence, M. Girolami, M. Rattray, and G. Sanguinetti. *Learning and Inference in Computational Systems Biology; Chapter 9*. MIT Press, 2010.

[8] N. D. Lawrence, G. Sanguinetti, and M. Rattray. Modelling transcriptional regulation using Gaussian processes. *Advances in Neural Information Processing Systems (NIPS)*, 21:785–792, 2008.

[9] K. Lin and D. Husmeier. Modelling transcriptional regulation with a mixture of factor analyzers and variational bayesian expectation maximization. *EURASIP Journal on Bioinformatics and Systems Biology*, 2009. Article ID 601068.

[10] R. M. Neal. *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*. Springer, New York, 1996. ISBN 0-387-94724-8.

[11] I. Pournara and L. Wernisch. Factor analysis for gene regulatory networks and transcription factor activity profiles. *BMC Bioinformatics*, 8, 2007.

[12] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[13] C. Sabatti and G. M. James. Bayesian sparse hidden components analysis for transcription regulation networks. *Bioinformatics*, 22(6):739–46, 2006.

[14] G. Sanguinetti, M. Rattray, and N. D. Lawrence. A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription. *Bioinformatics*, 22(14):1753–9, 2006.

[15] E. Snelsen, C. Rasmussen, and Z. Ghahramani. Warped Gaussian processes. *Neural Information Processing Systems (NIPS)*, 16, 2004.

[16] 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*, 6(53), 2012.

[17] V. Vyshemirsky and M. A. Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):S833–839, 2008.

Appendices

A Cross-Covariance between Target Genes

The expression levels of the target genes over time, $\{x_i(t)\}_{i=1}^{G}$, are given in terms of the transcription factor in eq. (2). This gives the formula for the cross-covariance shown in eq. (8) and repeated below

$$k_{x_j,x_k}(t,t') = S_j S_k \exp(-D_j t - D_k t') \int_0^t \exp(D_j u) \times \int_0^{t'} \exp(D_k u') k_{f,f}(u,u') du' du$$

where the transient terms have been ignored. As stated, when using an RBF kernel this integral becomes tractable and can be solved using a Laplace transformation. The solution to the integral can then be derived in the following manner

$$\begin{aligned} k_{x_j,x_k}(t,t') &= S_j S_k \exp(-D_j t - D_k t') \int_0^t \exp(D_j u) \\ &\quad \times \int_0^{t'} \exp(D_k u') a^2 \exp\left(-\tfrac{1}{l^2}[u-u']^2\right) du' du \\ &= a^2 S_j S_k \exp(-D_j t - D_k t') \int_0^t \exp(D_j u) \\ &\quad \times \int_0^{t'} \exp\left(-\frac{1}{l^2}[u^2 - 2uu' + u'^2 - D_k l^2 u']\right) du' du. \end{aligned}$$

We can then complete the square and move any factors that are constant out of the integral

$$\begin{aligned} k_{x_j,x_k}(t,t') &= a^2 S_j S_k \exp(-D_j t - D_k t') \int_0^t \exp\left(\frac{1}{l^2}[u + \tfrac{1}{2} l^2 D_k]^2\right) \\ &\quad \times \exp\left(D_j u - \frac{u^2}{l^2}\right) \int_0^{t'} \exp\left\{-\frac{1}{l^2}[u' - (u + \tfrac{1}{2} l^2 D_k)]^2\right\} du' du \\ &= a^2 S_j S_k \exp(-D_j t - D_k t') \exp(\gamma_k^2) \int_0^t \exp((D_j + D_k) u) \\ &\quad \times \int_0^{t'} \exp\left\{-\frac{1}{l^2}[u' - (u + \tfrac{1}{2} l^2 D_k)]^2\right\} du' du. \end{aligned}$$

where $\gamma_k = \frac{D_k l}{2}$. At this point the manipulation becomes clearer if a substitution is used, in this case let $s = \frac{1}{l}[u' - (u + \frac{1}{2} l^2 D_k)]$ so

$$\begin{aligned} k_{x_j,x_k}(t,t') &= a^2 S_j S_k l \exp(-D_j t - D_k t') \exp(\gamma_k^2) \\ &\quad \times \int_0^t \exp((D_j + D_k) u) \int_{s'}^{s''} \exp(-s^2) ds du \end{aligned}$$

where the l comes from the change of variable and the upper and lower bounds of the second integral are given as $s'' = \frac{1}{l}[t' - (u + \frac{1}{2} l^2 D_k)]$ and $s' = \frac{1}{l}[-(u + \frac{1}{2} l^2 D_k)]$

respectively. Defining the error function to be $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)dy$ and splitting the integral gives

$$\begin{aligned} k_{x_j,x_k}(t,t') &= \frac{\sqrt{\pi}}{2} a^2 S_j S_k l \exp(-D_j t - D_k t') \exp(\gamma_k^2) \\ &\times \int_0^t \exp((D_j + D_k)u)[\text{erf}(s'') - \text{erf}(s')]du \\ &= \frac{\sqrt{\pi}}{2} a^2 S_j S_k l \exp(-D_j t - D_k t') \exp(\gamma_k^2) \\ &\times \int_0^t \exp((D_j + D_k)u) \left[\text{erf}\left(\frac{t'-u}{l} - \gamma_k\right) + \text{erf}\left(\frac{u}{l} + \gamma_k\right)\right] du. \end{aligned}$$

Solving the first integral ultimately leads to a further integral still being required to be solved. In order to make this calculation visibly easier to follow, the integral is derived separately from the main formula in two separate parts (see Appendix A.1). Once this has been solved it can be put back into the formula in order to find the solution

$$\begin{aligned} k_{x_j,x_k}(t,t') = \frac{\sqrt{\pi} a^2 S_j S_k l}{2(D_j + D_k)} \exp(-D_j t - D_k t') \exp(\gamma_k^2) \Bigg[& \\ \exp((D_j + D_k)t) \,\text{erf}\left(\frac{t'-t}{l} - \gamma_k\right) - \text{erf}\left(\frac{t'}{l} - \gamma_k\right) - \exp(-\gamma_k^2) & \\ \times \exp(\gamma_j^2) \exp((D_j + D_k)t') \left\{\text{erf}\left(\frac{t-t'}{l} - \gamma_j\right) + \text{erf}\left(\frac{t'}{l} + \gamma_j\right)\right\} & \\ + \exp((D_j + D_k)t) \,\text{erf}\left(\frac{t}{l} + \gamma_k\right) - \text{erf}(\gamma_k) & \\ - \exp(-\gamma_k^2) \exp(\gamma_j^2) \left\{\text{erf}\left(\frac{t}{l} - \gamma_j\right) + \text{erf}(\gamma_j)\right\}\Bigg] & \end{aligned}$$

where $\gamma_j = \frac{D_j l}{2}$ and $\gamma_k = \frac{D_k l}{2}$. This can then be rearrange to give the solution stated in eq. (10).

A.1 Integral Solutions

In order to calculate the cross-covariance between two target genes, the following integral must be solved

$$\int_0^t \exp((D_j + D_k)u) \left[\text{erf}\left(\frac{t'-u}{l} - \gamma_k\right) + \text{erf}\left(\frac{u}{l} + \gamma_k\right)\right] du.$$

This can derived most simply by firstly splitting it into two and then integrating by parts. In this way the first part of the integral is given by

$$I_1 = \int_0^t \exp((D_j + D_k)u) \,\text{erf}\left(\frac{t'-u}{l} - \gamma_k\right) du$$

which can be solved in the following manner

$$\begin{aligned}
I_1 &= \frac{1}{D_j+D_k}\left[\exp((D_j+D_k)u)\operatorname{erf}\left(\frac{t'-u}{l}-\gamma_k\right)\right. \\
&\left.+\frac{2}{l\sqrt{\pi}}\int\exp((D_j+D_k)u)\exp\left(-\left(\frac{t'-u}{l}-\gamma_k\right)^2\right)du\right]_0^t \\
&= \frac{1}{D_j+D_k}\left[\exp((D_j+D_k)t)\operatorname{erf}\left(\frac{t'-t}{l}-\gamma_k\right)-\operatorname{erf}\left(\frac{t'}{l}-\gamma_k\right)\right. \\
&\left.+\frac{2}{l\sqrt{\pi}}\int_0^t\exp((D_j+D_k)u)\exp\left(-\frac{1}{l^2}[t'-u-\tfrac{1}{2}l^2D_k]^2\right)du\right].
\end{aligned}$$

Again to make the formula more manageable we make a substitution, so let $P = \exp((D_j+D_k)t)\operatorname{erf}(\frac{t'-t}{l}-\gamma_k)-\operatorname{erf}(\frac{t'}{l}-\gamma_k)$. Then as before we work towards a situation where we can complete the square

$$\begin{aligned}
I_1 &= \frac{1}{D_j+D_k}\left[P+\frac{2}{l\sqrt{\pi}}\int_0^t\exp((D_j+D_k)u)\right. \\
&\left.\times\exp\left(-\frac{1}{l^2}[u^2-2u(t'-\tfrac{1}{2}l^2D_k)+(t'-\tfrac{1}{2}l^2D_k)^2]\right)du\right] \\
&= \frac{1}{D_j+D_k}\left[P+\frac{2}{l\sqrt{\pi}}\exp\left(-\frac{1}{l^2}[t'-\tfrac{1}{2}l^2D_k]^2\right)\right. \\
&\left.\times\int_0^t\exp\left(-\frac{1}{l^2}[u^2-2u(t'+\tfrac{1}{2}l^2D_j-\tfrac{1}{2}l^2D_k+\tfrac{1}{2}l^2D_k)]\right)du\right] \\
&= \frac{1}{D_j+D_k}\left[P+\frac{2}{l\sqrt{\pi}}\exp(\gamma_j^2-\gamma_k^2)\exp(t'(D_j+D_k))\right. \\
&\left.\times\int_0^t\exp\left(-\frac{1}{l^2}[u-(t'+\tfrac{1}{2}l^2D_j)]^2\right)du\right].
\end{aligned}$$

We then once again make a substitution to simplify, $z = \frac{1}{l}[u-(t'+\frac{1}{2}l^2D_j)]$. This then gives the upper and lower bounds as $z'' = \frac{1}{l}[t-(t'+\frac{1}{2}l^2D_j)]$ and $z' = \frac{1}{l}[-(t'+\frac{1}{2}l^2D_j)]$ respectively. Then splitting these as before gives the solution

$$\begin{aligned}
I_1 &= \frac{1}{D_j+D_k}\left[P+\exp(\gamma_j^2-\gamma_k^2)\exp(t'(D_j+D_k))\{\operatorname{erf}(z'')-\operatorname{erf}(z')\}\right] \\
&= \frac{1}{D_j+D_k}\left[\exp((D_j+D_k)t)\operatorname{erf}\left(\frac{t'-t}{l}-\gamma_k\right)-\operatorname{erf}\left(\frac{t'}{l}-\gamma_k\right)\right. \\
&\left.+\exp(\gamma_j^2-\gamma_k^2)\exp(t'(D_j+D_k))\left\{\operatorname{erf}\left(\frac{t-t'}{l}-\gamma_j\right)+\operatorname{erf}\left(\frac{t'}{l}+\gamma_j\right)\right\}\right].
\end{aligned}$$

The second part of the integral is given as

$$I_2 = \int_0^t\exp((D_j+D_k)u)\operatorname{erf}\left(\frac{u}{l}+\gamma_k\right)du$$

which again be solved through integration by parts

$$
\begin{aligned}
I_2 &= \frac{1}{D_j + D_k}\left[\exp((D_j + D_k)u)\,\mathrm{erf}\left(\frac{u}{l} + \gamma_k\right)\right. \\
&\left. - \frac{2}{l\sqrt{\pi}}\int \exp((D_j + D_k)u)\exp\left(-\left(\frac{u}{l} + \gamma_k\right)^2\right)du\right]_0^t \\
&= \frac{1}{D_j + D_k}\left[\exp((D_j + D_k)t)\,\mathrm{erf}\left(\frac{t}{l} + \gamma_k\right) - \mathrm{erf}(\gamma_k)\right. \\
&\left. - \frac{2}{l\sqrt{\pi}}\int_0^t \exp((D_j + D_k)u)\exp\left(-\frac{1}{l^2}[u + \tfrac{1}{2}l^2 D_k]^2\right)du\right].
\end{aligned}
$$

Again for simplicity, we denote $Q = \exp((D_j + D_k)t)\,\mathrm{erf}\left(\frac{t}{l} + \gamma_k\right) - \mathrm{erf}(\gamma_k)$.

$$
\begin{aligned}
I_2 &= \frac{1}{D_j + D_k}\left[Q - \frac{2}{l\sqrt{\pi}}\int_0^t \exp((D_j + D_k)u)\right. \\
&\left. \times \exp\left(-\frac{1}{l^2}[u^2 + l^2 D_k u + \frac{l^4 D_k^2}{4}]\right)du\right] \\
&= \frac{1}{D_j + D_k}\left[Q - \frac{2}{l\sqrt{\pi}}\exp(-\gamma_k^2)\right. \\
&\left. \times \int_0^t \exp\left(-\frac{1}{l^2}[u^2 - 2u(\tfrac{1}{2}D_j l^2 - \tfrac{1}{2}D_k l^2 + \tfrac{1}{2}D_k l^2)]\right)du\right] \\
&= \frac{1}{D_j + D_k}\left[Q - \frac{2}{l\sqrt{\pi}}\exp(\gamma_j^2 - \gamma_k^2)\int_0^t \exp\left(-\tfrac{1}{l^2}\left[u - \tfrac{1}{2}D_j l^2\right]^2\right)du\right].
\end{aligned}
$$

Again we use a similar substitution to before, letting $v = \frac{1}{l}[u - \frac{1}{2}D_j l^2]$. This gives the upper and lower bounds respectively as $v'' = \frac{1}{l}[t - \frac{1}{2}D_j l^2]$ and $v' = \frac{1}{l}[-\frac{1}{2}D_j l^2]$. Splitting the integral then leads to the solution to I_2

$$
\begin{aligned}
I_2 &= \frac{1}{D_j + D_k}\left[Q - \exp(\gamma_j^2 - \gamma_k^2)\{\mathrm{erf}(v'') - \mathrm{erf}(v')\}\right] \\
&= \frac{1}{D_j + D_k}\left[\exp((D_j + D_k)t)\,\mathrm{erf}\left(\frac{t}{l} + \gamma_k\right) - \mathrm{erf}(\gamma_k)\right. \\
&\left. - \exp(\gamma_j^2 - \gamma_k^2)\left\{\mathrm{erf}\left(\frac{t}{l} - \gamma_j\right) + \mathrm{erf}(\gamma_j)\right\}\right].
\end{aligned}
$$

B Cross-Covariance between Target Genes and the Transcription Factor

In a similar manner to finding $k_{x_j,x_k}(t,t')$ we must also find $k_{x_j,f}(t,t')$. In order to do this we must solve the integral in eq. (11), given below

$$
k_{x_j,f}(t,t') = S_j \exp(-D_j t)\int_0^t \exp(D_j u)k_{f,f}(u,t')du
$$

where the transient terms have again been ignored. Again when an RBF kernel is chosen we can use Laplace transformation to solve the integral as follows

$$\begin{aligned} k_{x_j,f}(t,t') &= S_j \exp(-D_j t) \int_0^t \exp(D_j u) a^2 \exp\left(-\frac{1}{l^2}[u-t']^2\right) du \\ &= a^2 S_j \exp(-D_j t) \int_0^t \exp\left(-\frac{1}{l^2}[u^2 - 2ut' + t'^2 - D_j l^2 u]\right) du. \end{aligned}$$

We can then complete the square and move all of the factors that are constant outside of the integral

$$\begin{aligned} k_{x_j,f}(t,t') &= a^2 S_j \exp(-D_j t) \exp\left(\frac{[t' + \frac{1}{2}D_j l^2]^2}{l^2}\right) \exp\left(-\frac{t'^2}{l^2}\right) \\ &\quad \times \int_0^t \exp\left(-\frac{1}{l^2}[u - (t' + \tfrac{1}{2}D_j l^2)]^2\right) du \\ &= a^2 S_j \exp(-D_j(t-t')) \exp(\gamma_j^2) \\ &\quad \times \int_0^t \exp\left(-\frac{1}{l^2}[u - (t' + \tfrac{1}{2}D_j l^2)]^2\right) du \end{aligned}$$

where $\gamma_j = \frac{1}{2}D_j l$. A substitution of $s = \frac{1}{l}[u - (t' + \frac{1}{2}D_j l^2)]$ then makes the manipulation easier to manage

$$k_{x_j,f}(t,t') = a^2 S_j l \exp(-D_j(t-t')) \exp(\gamma_j^2) \int_{s'}^{s''} \exp(-s^2) ds$$

where l comes from the change of variable and the upper and lower bounds are given by $s'' = \frac{1}{l}[t - t' - \frac{1}{2}D_j l^2]$ and $s' = \frac{1}{l}[-t' - \frac{1}{2}D_j l^2]$ respectively. Defining the error function $\mathrm{erf}(x)$ as before and splitting the integral allows the derivation of the final solution

$$\begin{aligned} k_{x_j,f}(t,t') =& a^2 S_j l \exp(-D_j(t-t')) \exp(\gamma_j^2) \\ &\times \left[\int_0^{s''} \exp(-s^2) ds - \int_0^{s'} \exp(-s^2) ds\right] \\ =& \frac{\sqrt{\pi}}{2} a^2 S_j l \exp(-D_j(t-t')) \exp(\gamma_j^2) [\mathrm{erf}(s'') - \mathrm{erf}(s')] \\ =& \frac{\sqrt{\pi}}{2} a^2 S_j l \exp(-D_j(t-t')) \exp(\gamma_j^2) \\ &\times \left[\mathrm{erf}\left(\frac{t-t'}{l} - \gamma_j\right) + \mathrm{erf}\left(\frac{t'}{l} + \gamma_j\right)\right]. \end{aligned}$$

C Gradients of the Log Likelihood for GP

In order to learn the hyper parameters, $\boldsymbol{\theta}$, of the model it is necessary to maximise the likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ obtained from eq. (14). This is given in the form of a multivariate Gaussian distribution, given as

$$\ln p(\mathbf{y}|\boldsymbol{\theta}) = -\frac{1}{2}\ln|\mathbf{C}(\boldsymbol{\theta})| - \frac{1}{2}\left(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}\right)^\top \mathbf{C}(\boldsymbol{\theta})^{-1}\left(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}\right) - \tfrac{N}{2}\ln(2\pi).$$

For non-linear optimisation techniques, such as the conjugate gradients and Quasi-Newton methods, the gradient of the log likelihood is helpful for giving accurate and consistent estimates. Using some basic results for the derivation of matrices, see [2], the gradient is given as follows for the parameters $a^2, l, S_1, S \ldots, S_G$ and σ^2

$$\begin{aligned}\frac{\partial}{\partial \theta_i} \ln p(\mathbf{y}|\boldsymbol{\theta}) = -\frac{1}{2} \operatorname{Tr}\left(\mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i}\right) \\ + \frac{1}{2}\left(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}\right)^\top \mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}})\end{aligned} \tag{26}$$

where the trace function is defined as to be the sum of the elements of the main diagonal of the matrix. The result is also useful for $B_1, \ldots, B_G$ and $D_1, \ldots, D_G$, although not directly applicable due to the parameter values also occurring outside of $\mathbf{C}(\boldsymbol{\theta})$.

For the linear model $\boldsymbol{\theta} = (a^2, l, B_1, \ldots, B_G, S_1, \ldots, S_G, D_1, \ldots, D_G, \sigma^2)$ where $G = 3$ in the example and a gradient function must be found for each of these hyper parameters. All of these gradient functions can be found through differentiation by parts, although the expression often becomes very large. The expression for the derivative of l is given by

$$\frac{\partial}{\partial l} k_{x_j,x_k}(t,t') = \frac{1}{l} k_{x_j,x_k}(t,t') + l_{dif1} + l_{dif_2}$$

where $k_{x_j,x_k}(t,t')$ are elements of $\mathbf{K}(\boldsymbol{\theta})$, so have no noise, and where

$$\begin{aligned}l_{dif1} = \frac{\sqrt{\pi} l a^2 S_j S_k}{2(D_j + D_k)} \Bigg[\frac{1}{2} D_k^2 l \exp(\gamma_k^2) \\ \times \Bigg\{ \exp(D_k(t-t')) \left[\operatorname{erf}\left(\frac{t'-t}{l} - \gamma_k\right) + \operatorname{erf}\left(\frac{t}{l} + \gamma_k\right)\right] \\ - \exp(-(D_j t + D_k t')) \left[\operatorname{erf}\left(\frac{t'}{l} - \gamma_k\right) + \exp(\gamma_k)\right] \Bigg\} \\ + \frac{1}{2} D_j^2 l \exp(\gamma_j^2) \\ \times \Bigg\{ \exp(D_j(t'-t)) \left[\operatorname{erf}\left(\frac{t-t'}{l} - \gamma_j\right) + \operatorname{erf}\left(\frac{t'}{l} + \gamma_j\right)\right] \\ - \exp(-(D_j t + D_k t')) \left[\operatorname{erf}\left(\frac{t}{l} - \gamma_j\right) + \exp(\gamma_j)\right] \Bigg\} \Bigg]\end{aligned}$$

and

$$
\begin{aligned}
l_{dif2} = \frac{la^2 S_j S_k}{D_j + D_k} \Bigg[& \exp(D_k(t-t')) \exp(\gamma_k^2) \\
& \times \left\{ \exp\left(- \left(\frac{t'-t}{l} - \gamma_k \right)^2 \right) \left(-\frac{D_k}{2} - \frac{t'-t}{l^2} \right) \right. \\
& \left. + \exp\left(- \left(\frac{t}{l} + \gamma_k \right)^2 \right) \left(\frac{D_k}{2} - \frac{t}{l^2} \right) \right\} \\
& - \exp(-(D_j t + D_k t')) \exp(\gamma_k^2) \\
& \times \left\{ \exp\left(- \left(\frac{t'}{l} - \gamma_k^2 \right)^2 \right) \left(-\frac{D_k}{2} - \frac{t'}{l^2} \right) + \exp(-\gamma_k^2) \frac{D_k}{2} \right\} \\
& + \exp(D_j(t'-t)) \exp(\gamma_j^2) \\
& \times \left\{ \exp\left(- \left(\frac{t-t'}{l} - \gamma_j \right)^2 \right) \left(-\frac{D_j}{2} - \frac{t-t'}{l^2} \right) \right. \\
& \left. + \exp\left(- \left(\frac{t'}{l} + \gamma_j \right)^2 \right) \left(\frac{D_j}{2} - \frac{t'}{l^2} \right) \right\} \\
& - \exp(-(D_j t + D_k t')) \exp(\gamma_j^2) \\
& \times \left\{ \exp\left(- \left(\frac{t}{l} - \gamma_j^2 \right)^2 \right) \left(-\frac{D_j}{2} - \frac{t}{l^2} \right) + \exp(-\gamma_j^2) \frac{D_j}{2} \right\} \Bigg].
\end{aligned}
$$

This can be put into eq. (26) to give the function for the gradient for l.

The gradient functions for B_1 is given by

$$
\begin{aligned}
\frac{\partial}{\partial B_1} \ln p(\mathbf{y}|\boldsymbol{\theta}) = -\frac{1}{2} \big[& (\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}})^\top \mathbf{C}(\boldsymbol{\theta})^{-1} (-\tfrac{\mathbf{1}}{\mathbf{D_1}}, \mathbf{0}, \mathbf{0}) \\
& + (-\tfrac{\mathbf{1}}{\mathbf{D_1}}, \mathbf{0}, \mathbf{0})^\top \mathbf{C}(\boldsymbol{\theta})^{-1} (\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}) \big]
\end{aligned}
$$

where the vector $(-\frac{\mathbf{1}}{\mathbf{D_1}}, \mathbf{0}, \mathbf{0})$ has the same length as $\mathbf{y}$ with each part having a length corresponding to the number of observations from the related gene. Similar gradient functions are available for B_2, B_3, etc., where the above vector is replaced by $(\mathbf{0}, -\frac{\mathbf{1}}{\mathbf{D_2}}, \mathbf{0})$ for B_2 for example when $G = 3$.

The gradient function for $S_1, \ldots, S_G$ and $D_1, \ldots, D_G$ is more complicated due to the varying occurrence of the $\mathbf{S_i}$ and $\mathbf{D_i}$ within the covariance matrix. This can be illustrated if we think of the covariance matrix as being made up of G^2 equal sized smaller matrices. Thinking in this manner means that for the parameter S_1 for example, the small matrix at the top left of the larger matrix has $S_j = S_k = S_1$. Similarly the small matrices in the same row or column as this have either $S_j = S_1$ and $S_k \neq S_1$ or $S_j \neq S_1$ and $S_k = S_1$. For all of the other matrices $S_j \neq S_1$ and $S_k \neq S_1$, which ultimately means that different elements of the matrix have different equations for calculating the gradient.

Looking specifically at the parameters, $S_1, \ldots, S_G$, where $S_j = S_i$ and $S_k = S_i$ the differential is given by

$$
\frac{\partial}{\partial S_i} k_{x_i, x_i}(t, t') = \frac{2}{S_i} k_{x_i, x_i}(t, t').
$$

Similarly for $S_j = S_i$ and $S_k \neq S_i$ or $S_j \neq S_i$ and $S_k = S_i$ the differential is given by

$$\frac{\partial}{\partial S_i} k_{x_j,x_k}(t,t') = \frac{1}{S_i} k_{x_j,x_k}(t,t').$$

Finally where $S_j \neq S_i$ and $S_k \neq S_i$, the differential is equal to zero. Combining these various elements gives the matrix for the differential and this can be put into eq. (26) to give the function for the gradient for S_i.

The gradient function for D_1 is given by

$$\begin{aligned}\frac{\partial}{\partial D_1} \ln p(\mathbf{y}|\boldsymbol{\theta}) = &-\frac{1}{2} \operatorname{Tr}\left(\mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial D_1}\right) - \frac{1}{2}(\tfrac{\mathbf{B_1}}{\mathbf{D_1^2}}, \mathbf{0}, \mathbf{0})^\top \mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}) \\ &- \tfrac{1}{2}(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}})^\top \mathbf{C}(\boldsymbol{\theta})^{-1}(\tfrac{\mathbf{B_1}}{\mathbf{D_1^2}}, \mathbf{0}, \mathbf{0}) \\ &+ \frac{1}{2}(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}})^\top \mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial D_1} \mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{y} - \tfrac{\mathbf{B}}{\mathbf{D}}). \end{aligned} \tag{27}$$

Again similarly to $B_1, \ldots, B_G$, this can be adapted for D_2, D_3, etc., where the vector $(\frac{\mathbf{B_1}}{\mathbf{D_1^2}}, \mathbf{0}, \mathbf{0})$ is replaced $(\mathbf{0}, \frac{\mathbf{B_2}}{\mathbf{D_2^2}}, \mathbf{0})$ for example for D_2 when $G = 3$.

We are now required to work out the differential of D_i to put into eq. (27). Again different elements of the matrix have varying differentials, so firstly where $D_j = D_i$ and $D_k = D_i$ we have

$$\frac{\partial k_{x_i,x_i}(t,t')}{\partial D_i} = (\tfrac{1}{2} D_i l^2 - \tfrac{1}{D_i}) k_{x_i,x_i}(t,t') + D_{i,i,dif1} + D_{i,i,dif2}$$

where

$$\begin{aligned} D_{i,i,dif1} = &\frac{\sqrt{\pi} l a^2 S_i^2}{4 D_i} \exp(\gamma_i^2) \\ &\times \left[(t-t') \exp(-D_i(t'-t)) \left\{ \operatorname{erf}\left(\frac{t'-t}{l} - \gamma_i\right) + \operatorname{erf}\left(\frac{t}{l} + \gamma_i\right) \right\} \right. \\ &+ (t+t') \exp(-D_j(t+t')) \left\{ \operatorname{erf}\left(\frac{t'}{l} - \gamma_i\right) + \operatorname{erf}(\gamma_i) \right\} \\ &+ (t'-t) \exp(-D_i(t-t')) \left\{ \operatorname{erf}\left(\frac{t-t'}{l} - \gamma_i\right) + \operatorname{erf}\left(\frac{t'}{l} + \gamma_i\right) \right\} \\ &\left. + (t+t') \exp(-D_j(t+t')) \left\{ \operatorname{erf}\left(\frac{t}{l} - \gamma_i\right) + \operatorname{erf}(\gamma_i) \right\} \right] \end{aligned}$$

and

$$
\begin{aligned}
D_{i,i,dif2} = \frac{l^2 a^2 S_i^2}{4D_i} \exp(\gamma_i^2) \\
&\times \left[\exp(-D_i(t'-t)) \left\{ -\exp\left(-\left(\frac{t'-t}{l} - \gamma_i \right)^2 \right) + \exp\left(-\left(\frac{t}{l} + \gamma_i \right)^2 \right) \right\} \right. \\
&- \exp(-D_i(t+t')) \left\{ -\exp\left(-\left(\frac{t'}{l} - \gamma_i \right)^2 \right) - \exp(-\gamma_i^2) \right\} \\
&+ \exp(-D_i(t-t')) \left\{ -\exp\left(-\left(\frac{t-t'}{l} - \gamma_i \right)^2 \right) + \exp\left(-\left(\frac{t}{l} + \gamma_i \right)^2 \right) \right\} \\
&\left. - \exp(-D_i(t+t')) \left\{ -\exp\left(-\left(\frac{t}{l} - \gamma_i \right)^2 \right) + \exp(-\gamma_i^2) \right\} \right]
\end{aligned}
$$

Similarly when $D_j = D_i$ and $D_k \neq D_i$, then the differential is given by

$$
\frac{\partial k_{x_i,x_k}(t,t')}{\partial D_i} = -\frac{1}{D_i + D_k} k_{x_i,x_k}(t,t') + D_{i,k,dif1} + D_{i,k,dif2} + D_{i,k,dif3}
$$

where

$$
\begin{aligned}
D_{i,k,dif1} = \frac{\sqrt{\pi} l a^2 S_i S_k}{2(D_i + D_k)} \left\{ t \exp(\gamma_k^2) \exp(-(D_i t + D_k t')) \right. \\
&\times \left[\operatorname{erf}\left(\frac{t'}{l} - \gamma_k \right) + \operatorname{erf}(\gamma_k) \right] + \tfrac{1}{2} D_i l^2 \exp(\gamma_i^2) \\
&\left\{ \exp(-D_i(t-t')) \left[\operatorname{erf}\left(\frac{t-t'}{l} - \gamma_i \right) + \operatorname{erf}\left(\frac{t'}{l} + \gamma_i \right) \right] \right. \\
&\left. \left. - \exp(-(D_i t + D_k t')) \left[\operatorname{erf}\left(\frac{t}{l} - \gamma_i \right) + \operatorname{erf}(\gamma_i) \right] \right\},
\end{aligned}
$$

$$
\begin{aligned}
D_{i,k,dif2} = \frac{\sqrt{\pi} l a^2 S_i S_k}{2(D_i + D_k)} \exp(\gamma_i^2) \left\{ (t'-t) \exp(-D_i(t-t')) \right. \\
&\times \left[\operatorname{erf}\left(\frac{t-t'}{l} - \gamma_i \right) + \operatorname{erf}\left(\frac{t'}{l} + \gamma_i \right) \right] + t \exp(-(D_i t + D_k t')) \\
&\left. \times \left[\operatorname{erf}\left(\frac{t}{l} - \gamma_i \right) + \operatorname{erf}(\gamma_i) \right] \right\}
\end{aligned}
$$

and

$$
\begin{aligned}
D_{i,k,dif3} = \frac{l^2 a^2 S_i S_k}{2(D_i + D_k)} \exp(\gamma_i^2) \left\{ \exp(-D_i(t-t')) \right. \\
&\times \left[-\exp\left(-\left(\frac{t-t'}{l} - \gamma_i \right)^2 \right) + \exp\left(-\left(\frac{t'}{l} + \gamma_i \right)^2 \right) \right] \\
&\left. + \exp(-(D_i t + D_k t')) \left[\exp\left(-\left(\frac{t}{l} - \gamma_i \right)^2 \right) - \exp(-\gamma_i^2) \right] \right\}
\end{aligned}
$$

There is also a similar expression also available for $D_j \neq D_i$ and $D_k = D_i$, where t and t' are swapped and D_j replaces D_k. Then similarly to $\mathbf{S}$, when $D_j \neq D_i$ and $D_k \neq D_i$, then this part of the differential is equal to zero. Once calculated, all parts of the differential can be combined and used within eq. (27).

Finally an expression must be found for the gradient of σ^2 and a^2. By eq. (14) it immediately become obvious that that the derivative for σ^2 is simply the identity matrix of the appropriate length. This can then be easily be used in eq. (26) to give the gradient function as required. The derivative for a^2 can also be simply used eq. (26) and this is given by

$$\frac{\partial}{\partial a^2} k_{x_j,x_k}(t,t') = \frac{1}{a^2} k_{x_j,x_k}(t,t').$$