
Latent Force Models

Mauricio Alvarez

School of Computer Science
University of Manchester
Manchester, UK, M13 9PL
alvarezm@cs.man.ac.uk

David Luengo

Dep. Teoría de Señal y Comunicaciones
Universidad Carlos III de Madrid
28911 Leganés, Spain
luengod@ieee.org

Neil D. Lawrence

School of Computer Science
University of Manchester
Manchester, UK, M13 9PL
neill@cs.man.ac.uk

Abstract

Purely data driven approaches for machine learning present difficulties when data is scarce relative to the complexity of the model or when the model is forced to extrapolate. On the other hand, purely mechanistic approaches need to identify and specify all the interactions in the problem at hand (which may not be feasible) and still leave the issue of how to parameterize the system. In this paper, we present a hybrid approach using Gaussian processes and differential equations to combine data driven modelling with a physical model of the system. We show how different, physically-inspired, kernel functions can be developed through sensible, simple, mechanistic assumptions about the underlying system. The versatility of our approach is illustrated with three case studies from computational biology, motion capture and geostatistics.

1 Introduction

Traditionally, the main focus in machine learning has been model generation through a *data driven paradigm*. The usual approach is to combine a data set with a (typically fairly flexible) class of models and, through judicious use of regularization, make useful predictions on previously unseen data. There are two key problems with purely data driven approaches. Firstly, if data is scarce relative to the complexity of the system we may be unable to make accurate predictions on test data. Secondly, if the model is forced to

extrapolate, *i.e.* make predictions in a regime in which data has not been seen yet, performance can be poor.

Purely *mechanistic models*, *i.e.* models which are inspired by the underlying physical knowledge of the system, are common in many areas such as chemistry, systems biology, climate modelling and geophysical sciences, *etc.* They normally make use of a fairly well characterized physical process that underpins the system, typically represented with a set of differential equations. The purely mechanistic approach leaves us with a different set of problems to those from the data driven approach. In particular, accurate description of a complex system through a mechanistic modelling paradigm may not be possible: even if all the physical processes can be adequately described, the resulting model could become extremely complex. Identifying and specifying all the interactions might not be feasible, and we would still be faced with the problem of identifying the parameters of the system.

Despite these problems, physically well characterized models retain a major advantage over purely data driven models. A mechanistic model can enable accurate prediction even in regions where there may be no available training data. For example, Pioneer space probes have been able to enter different extra terrestrial orbits despite the absence of data for these orbits.

In this paper we advocate an alternative approach. Rather than relying on an exclusively mechanistic or data driven approach we suggest a *hybrid system* which involves a (typically overly simplistic) mechanistic model of the system which can easily be augmented through machine learning techniques. We will start by considering two dynamical systems, both simple latent variable models, which incorporate first and second order differential equations. Our inspiration is the work of (Lawrence *et al.*, 2007; Gao *et al.*, 2008) who encoded a first order differential equation in a Gaussian process (GP). However, their aim was to construct an accurate model of transcriptional regulation, whereas ours is to make use of the mechanistic model to in-

Appearing in Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS) 2009, Clearwater Beach, Florida, USA. Volume 5 of JMLR: W&CP 5. Copyright 2009 by the authors.

corporate salient characteristics of the data (*e.g.* in a mechanical system *inertia*) without necessarily associating the components of our mechanistic model with actual physical components of the system. For example, for a human motion capture dataset we develop a mechanistic model of motion capture that does not exactly replicate the *physics* of human movement, but nevertheless captures salient features of the movement. Having shown how first and second order dynamical systems can be incorporated in a GP, we finally show how partial differential equations can also be incorporated for modelling systems with multiple inputs.

2 Latent Variables and Physical Systems

From the perspective of machine learning our approach can be seen as a type of latent variable model. In a latent variable model we may summarize a high dimensional data set with a reduced dimensional representation. For example, if our data consists of N points in a Q dimensional space we might seek a linear relationship between the data, $\mathbf{Y} \in \mathbb{R}^{N \times Q}$, and a reduced dimensional representation, $\mathbf{F} \in \mathbb{R}^{N \times R}$, where $R < Q$. From a probabilistic perspective this involves an assumption that we can represent the data as

$$\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{E}, \quad (1)$$

where $\mathbf{E}$ is a matrix-variate Gaussian noise: each column, $\epsilon_{:,q}$ ($1 \leq q \leq Q$), is a multi-variate Gaussian with zero mean and covariance Σ , i.e. $\epsilon_{:,q} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. The usual approach, as undertaken in factor analysis and principal component analysis (PCA), to dealing with the unknowns in this model is to integrate out $\mathbf{F}$ under a Gaussian prior and optimize with respect to $\mathbf{W} \in \mathbb{R}^{R \times Q}$ (although it turns out that for a non-linear variant of the model it can be convenient to do this the other way around, see *e.g.* (Lawrence, 2005)). If the data has a temporal nature, then the Gaussian prior in the latent space could express a relationship between the rows of $\mathbf{F}$, $\mathbf{f}_{t_n} = \mathbf{f}_{t_{n-1}} + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and $\mathbf{f}_{t_n}$ is the n -th row of $\mathbf{F}$, which we associate with time t_n . This is known as the *Kalman filter/smoother*. Normally the times, t_n , are taken to be equally spaced, but more generally we can consider a joint distribution for $p(\mathbf{F}|\mathbf{t})$, $\mathbf{t} = [t_1 \dots t_N]^\top$, which has the form of a Gaussian process (GP),

$$p(\mathbf{F}|\mathbf{t}) = \prod_{r=1}^R \mathcal{N}(\mathbf{f}_{:,r} | \mathbf{0}, \mathbf{K}_{f_{:,r}, f_{:,r}}),$$

where we have assumed zero mean and independence across the R dimensions of the latent space. The GP makes explicit the fact that the latent variables are functions, $\{f_r(t)\}_{r=1}^R$, and we have now described them

with a process prior. The notation used, $\mathbf{f}_{:,r}$, indicates the r -th column of $\mathbf{F}$, and represents the values of that function for the r -th dimension at the times given by $\mathbf{t}$. The matrix $\mathbf{K}_{f_{:,r}, f_{:,r}}$ is the covariance function associated to $f_r(t)$ computed at the times given in $\mathbf{t}$.

Such a GP can be readily implemented. Given the covariance functions for $\{f_r(t)\}$ the implied covariance functions for $\{y_q(t)\}$ are straightforward to derive. In (Teh *et al.*, 2005) this is known as a semi-parametric latent factor model (SLFM), although their main focus is not the temporal case. Historically the Kalman filter approach has been preferred, perhaps because of its linear computational complexity in N . However, recent advances in sparse approximations have made the general GP framework practical (see (Quiñero Candela and Rasmussen, 2005) for a review).

So far the model described relies on the latent variables to provide the dynamic information. Our main contribution is to include a further dynamical system with a *mechanistic* inspiration. We now use a mechanical analogy to introduce it. Consider the following physical interpretation of (1): the latent functions, $f_r(t)$, are R forces and we observe the displacement of Q springs, $y_q(t)$, to the forces. Then we can reinterpret (1) as the force balance equation, $\mathbf{Y}\mathbf{D} = \mathbf{F}\mathbf{S} + \tilde{\mathbf{E}}$. Here we have assumed that the forces are acting, for example, through levers, so that we have a matrix of sensitivities, $\mathbf{S} \in \mathbb{R}^{R \times Q}$, and a diagonal matrix of spring constants, $\mathbf{D} \in \mathbb{R}^{Q \times Q}$. The original model is recovered by setting $\mathbf{W} = \mathbf{S}\mathbf{D}^{-1}$ and $\tilde{\epsilon}_{:,q} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^\top \Sigma \mathbf{D})$. The model can be extended by assuming that the spring is acting in parallel with a damper and that the system has mass, allowing us to write,

$$\mathbf{F}\mathbf{S} = \ddot{\mathbf{Y}}\mathbf{M} + \dot{\mathbf{Y}}\mathbf{C} + \mathbf{Y}\mathbf{D} + \epsilon, \quad (2)$$

where $\mathbf{M}$ and $\mathbf{C}$ are diagonal matrices of masses and damping coefficients respectively, $\dot{\mathbf{Y}} \in \mathbb{R}^{N \times Q}$ is the first derivative of $\mathbf{Y}$ w.r.t. time and $\ddot{\mathbf{Y}}$ is the second derivative. The second order mechanical system that this model describes will exhibit several characteristics which are impossible to represent in the simpler latent variable model given by (1), such as inertia and resonance. This model is not only appropriate for data from mechanical systems. There are many analogous systems which can also be represented by second order differential equations, *e.g.* Resistor-Inductor-Capacitor circuits. A unifying characteristic for all these models is that the system is being forced by latent functions, $\{f_r(t)\}_{r=1}^R$. Hence, we refer to them as *latent force models* (LFMs).

One way of thinking of our model is to consider puppetry. A marionette is a representation of a human (or animal) controlled by a limited number of inputs through strings (or rods) attached to the character.

This limited number of inputs can lead to a wide range of character movements. In our model, the data is the movements of the marionette, and the latent forces are the inputs to the system from the puppeteer.

Finally, note that it is of little use to include dynamical models of the type specified in (2) if their effects cannot be efficiently incorporated into the inference process. Fortunately, as we will see in the case studies, for an important class of covariance functions it is *analytically* tractable to compute the implied covariance functions for $\{y_q(t)\}_{q=1}^Q$. Then, given the data a conjugate gradient descent algorithm can be used to obtain the hyperparameters of the model which minimize the minus log-likelihood, and inference is performed based on standard GP regression techniques.

3 First Order Dynamical System

A single input module is a biological network motif where the transcription of a number of genes is driven by a single transcription factor. In (Barenco *et al.*, 2006) a simple first order differential equation was proposed to model this situation. Then (Lawrence *et al.*, 2007; Gao *et al.*, 2008) suggested that inference of the latent transcription factor concentration should be handled using GPs. In effect their model can be seen as a latent force model based on a first order differential equation with a single latent force. Here we consider the extension of this model to multiple latent forces. As a mechanistic model, this is a severe over simplification of the physical system: transcription factors are known to interact in a non linear manner. Despite this we will be able to uncover useful information. Our model is based on the following differential equation,

$$\frac{dy_q(t)}{dt} + D_q y_q(t) = B_q + \sum_{r=1}^R S_{rq} f_r(t). \quad (3)$$

Here the latent forces, $f_r(t)$, represent protein concentration (which is difficult to observe directly), the outputs, $y_q(t)$, are the mRNA abundance levels for different genes, B_q and D_q are respectively the basal transcription and the decay rates of the q -th gene, and S_{rq} are coupling constants that quantify the influence of the r -th input on the q -th output (*i.e.* the sensitivity of gene q to the concentration of protein r). Solving (3) for $y_q(t)$, we obtain

$$y_q(t) = \frac{B_q}{D_q} + \sum_{r=1}^R L_{rq} [f_r](t),$$

where we have ignored transient terms, which are easily included, and the linear operator is given by the following linear convolution operator,

$$L_{rq} [f_r](t) = S_{rq} \exp(-D_q t) \int_0^t f_r(\tau) \exp(D_q \tau) d\tau.$$

If each latent force is taken to be independent with a covariance function given by

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

then we can compute the covariance of the outputs analytically, obtaining (Lawrence *et al.*, 2007)

$$k_{y_p y_q}(t, t') = \sum_{r=1}^R \frac{S_{rp} S_{rq} \sqrt{\pi} \ell_r}{2} [h_{qp}(t', t) + h_{pq}(t, t')],$$

where

$$\begin{aligned} h_{qp}(t', t) = & \frac{\exp(\nu_{rq}^2)}{D_p + D_q} \exp(-D_q t') \left\{ \exp(D_q t) \right. \\ & \times \left[\operatorname{erf}\left(\frac{t' - t}{\ell_r} - \nu_{rq}\right) + \operatorname{erf}\left(\frac{t}{\ell_r} + \nu_{rq}\right) \right] \\ & \left. - \exp(-D_p t) \left[\operatorname{erf}\left(\frac{t'}{\ell_r} - \nu_{rq}\right) + \operatorname{erf}(\nu_{rq}) \right] \right\}, \end{aligned}$$

here $\operatorname{erf}(x)$ is the real valued error function, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy$, and $\nu_{rq} = \ell_r D_q / 2$.

Additionally, we can compute the cross-covariance between the inputs and outputs,

$$\begin{aligned} k_{y_q f_r}(t, t') = & \frac{S_{rq} \sqrt{\pi} \ell_r}{2} \exp(\nu_{rq}^2) \exp(-D_q(t - t')) \\ & \times \left[\operatorname{erf}\left(\frac{t' - t}{\ell_r} - \nu_{rq}\right) + \operatorname{erf}\left(\frac{t'}{\ell_r} + \nu_{rq}\right) \right]. \end{aligned}$$

3.1 p53 Data

Our data is from (Barenco *et al.*, 2006), where leukemia cell lines were bombarded with radiation to induce activity of the transcription factor p53. This transcription factor repairs DNA damage and triggers a mechanism which pauses the cell-cycle and potentially terminates the cell. In (Barenco *et al.*, 2006) microarray gene expression levels of known targets of p53 were used to fit a first order differential equation model to the data. The model was then used to provide a ranked list of 50 genes identified as regulated by p53.

Our aim is to determine if there are additional “latent forces” which could better explain the activity of some of these genes. The experimental data consists of measurements of expression levels of 50 genes for three different replicas. Within each replica, there are measurements at seven different time instants. We constructed a latent force model with six latent forces, assuming that each replica was independently produced but fixing the hyperparameters of the kernel across the

replicas¹. We employed a sparse approximation, as proposed in (Alvarez and Lawrence, 2009), with ten inducing points for speeding up computation.

Of the six latent functions, two were automatically switched off by the model. Two further latent functions, shown in Figure 1 as latent forces 1 & 2, were consistent across all replicas: their shapes were time translated versions of the p53 profile as identified by (Barenco *et al.*, 2006; Lawrence *et al.*, 2007; Gao *et al.*, 2008). This time translation allows genes to experience different *transcriptional delays*, a mechanism not included explicitly in our model, but mimicked by linear mixing of an early and a late signal. The remaining two latent functions were inconsistent across the replicas (see e.g. latent force 3 in Figure 1). They appear to represent processes not directly related to p53. This was backed up by the sensitivity parameters found in the model. The known p53 targets DDB2, p21, SESN1/hPA26, BIK and TNFRSF10b were found to respond to latent forces 1 & 2. Conversely, the genes that were most responsive to latent force 3 were MAP4K4, a gene involved in environmental stress signalling, and FDXR, an electron transfer protein.

4 Second Order Dynamical System

In Section 1 we introduced the analogy of a marionette’s motion being controlled by a reduced number of forces. Human motion capture data consists of a skeleton and multivariate time courses of angles which summarize the motion. This motion can be modelled with a set of second order differential equations which, due to variations in the centers of mass induced by the movement, are non-linear. The simplification we consider for the latent force model is to linearize these differential equations, resulting in the following second order dynamical system,

$$\frac{d^2 y_q(t)}{dt^2} + C_q \frac{dy_q(t)}{dt} + D_q y_q(t) = B_q + \sum_{r=1}^R S_{rq} f_r(t), \quad (4)$$

where the mass of the system, without loss of generality, is normalized to 1. Whilst (4) is not the correct physical model for our system, it will still be helpful when extrapolating predictions across different motions, as we shall see in the next section. Note also that, although similar to (3), the dynamic behavior of this system is much richer than that of the first order system, since it can exhibit inertia and resonance.

¹The decay rates were assumed equal within replicas. Although this might be an important restriction for this experiment, our purpose in this paper is to expose a general methodology without delving into the details of each experimental setup.

For the motion capture data $y_q(t)$ corresponds to a given observed angle over time, and its derivatives represent angular velocity and acceleration. The system is summarized by the undamped natural frequency, $\omega_{0q} = \sqrt{D_q}$, and the damping ratio, $\zeta_q = \frac{1}{2}C_q/\sqrt{D_q}$. Systems with a damping ratio greater than one are said to be overdamped, whereas underdamped systems exhibit resonance and have a damping ratio less than one. For critically damped systems $\zeta_q = 1$, and finally, for undamped systems (i.e. no friction) $\zeta_q = 0$.

Ignoring the initial conditions once more, the solution of (4) is again given by a convolution, with the linear operator now being

$$\begin{aligned} L_{rq}[f_r](t) &= \frac{S_{rq}}{\omega_q} \exp(-\alpha_q t) \\ &\times \int_0^t f_r(\tau) \exp(\alpha_q \tau) \sin(\omega_q(t - \tau)) d\tau, \end{aligned} \quad (5)$$

where $\omega_q = \sqrt{4D_q - C_q^2}/2$ and $\alpha_q = C_q/2$.

Once again, if we consider a latent force governed by a GP with the RBF covariance function we can solve (5) analytically, obtaining a closed-form expression for the covariance matrix of the outputs,

$$k_{y_p y_q}(t, t') = \sum_{r=1}^R \frac{S_{rp} S_{rq} \sqrt{\pi \ell_r^2}}{8 \omega_p \omega_q} k_{y_p y_q}^{(r)}(t, t').$$

Here $k_{y_p y_q}^{(r)}(t, t')$ can be considered the cross-covariance between the p -th and q -th outputs under the effect of the r -th latent force, and is given by

$$\begin{aligned} k_{y_p y_q}^{(r)}(t, t') &= h_r(\tilde{\gamma}_q, \gamma_p, t, t') + h_r(\gamma_p, \tilde{\gamma}_q, t', t) \\ &+ h_r(\gamma_q, \tilde{\gamma}_p, t, t') + h_r(\tilde{\gamma}_p, \gamma_q, t', t) \\ &- h_r(\tilde{\gamma}_q, \tilde{\gamma}_p, t, t') - h_r(\tilde{\gamma}_p, \tilde{\gamma}_q, t', t) \\ &- h_r(\gamma_q, \gamma_p, t, t') - h_r(\gamma_p, \gamma_q, t', t), \end{aligned}$$

where $\gamma_p = \alpha_p + j\omega_p$, $\tilde{\gamma}_p = \alpha_p - j\omega_p$, and

$$h_r(\gamma_q, \gamma_p, t, t') = \frac{\Upsilon_r(\gamma_q, t', t) - \exp(-\gamma_p t) \Upsilon_r(\gamma_q, t', 0)}{\gamma_p + \gamma_q},$$

with

$$\begin{aligned} \Upsilon_r(\gamma_q, t, t') &= 2 \exp\left(\frac{\ell_r^2 \gamma_q^2}{4}\right) \exp(-\gamma_q(t - t')) \\ &- \exp\left(-\frac{(t-t')^2}{\ell_r^2}\right) w(jz_{rq}(t)) - \exp\left(-\frac{(t')^2}{\ell_r^2}\right) \\ &\times \exp(-\gamma_q t) w(-jz_{rq}(0)), \end{aligned} \quad (6)$$

and $z_{rq}(t) = (t - t')/\ell_r - (\ell_r \gamma_q)/2$. Note that $z_{rq}(t) \in \mathbb{C}$, and $w(jz)$ in (6), for $z \in \mathbb{C}$, denotes Faddeeva’s function $w(jz) = \exp(z^2) \operatorname{erfc}(z)$, where $\operatorname{erfc}(z)$ is the complex version of the complementary error function,

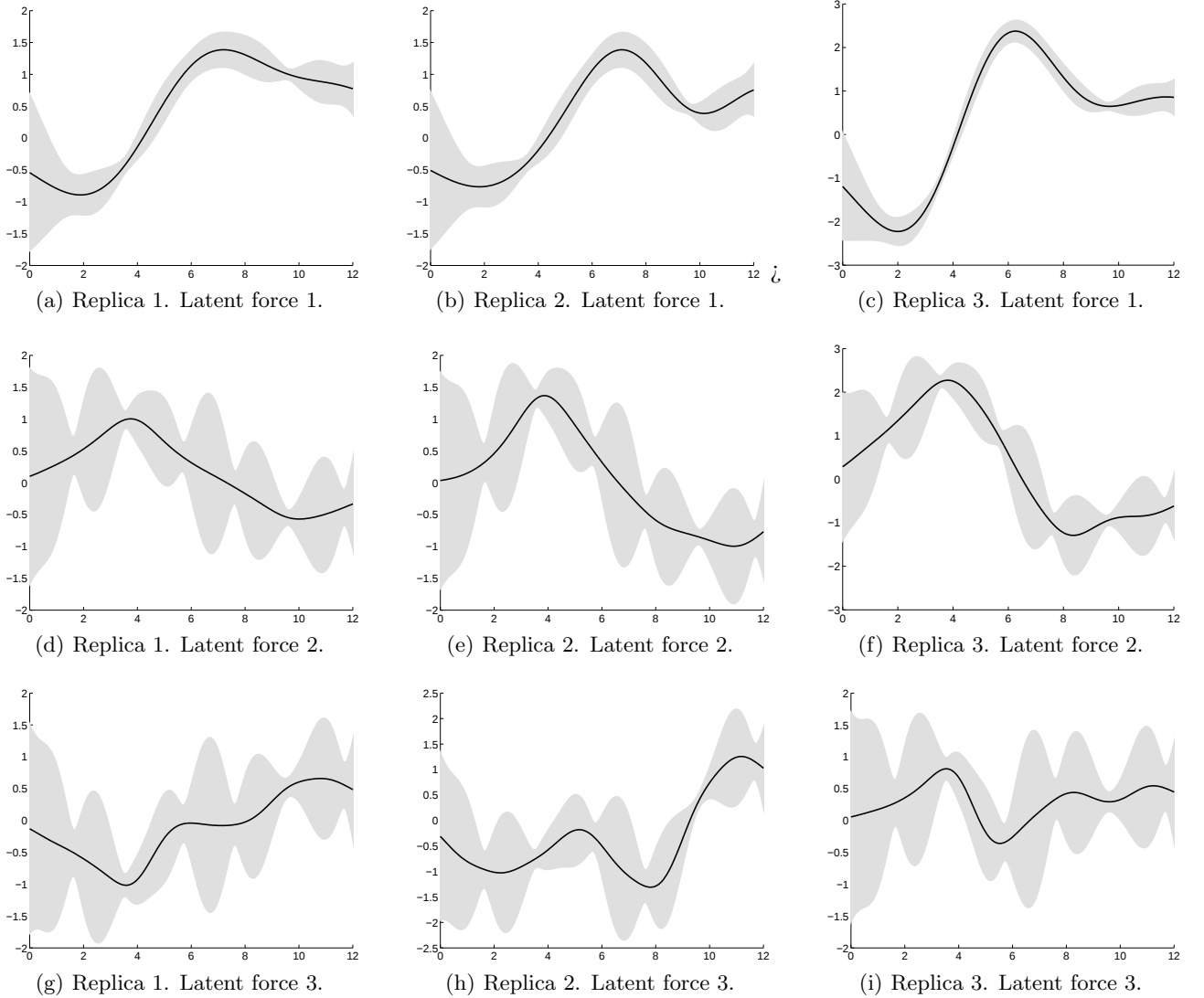


Figure 1: (a)-(c) and (d)-(f) the two latent forces associated with p53 activity. p53 targets are sensitive to a combination of these functions allowing them to account for transcriptional delays. (g)-(h) a latent force that was inconsistent across the replicas. It may be associated with cellular processes not directly related to p53.

$\text{erfc}(z) = 1 - \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp(-v^2) dv$. Faddeeva's function is usually considered the complex equivalent of the error function, since $|\text{w}(jz)|$ is bounded whenever the imaginary part of jz is greater or equal than zero, and is the key to achieving a good numerical stability when computing (6) and its gradients.

Similarly, the cross-covariance between latent functions and outputs is given by

$$k_{y_q f_r}(t, t') = \frac{\ell_r S_{r q} \sqrt{\pi}}{j 4 \omega_q} [\Upsilon_r(\tilde{\gamma}_q, t, t') - \Upsilon_r(\gamma_q, t, t')],$$

A visualization of a covariance matrix with a latent force and three different outputs (overdamped, underdamped and critically damped) is given in Figure 2.

4.1 Motion Capture data

Our motion capture data set is from the CMU motion capture data base². We considered 3 balancing motions (18, 19, 20) from subject 49. The subject starts in a standing position with arms raised, then, over about 10 seconds, he raises one leg in the air and lowers his arms to an outstretched position. Of interest to us was the fact that, whilst motions 18 and 19 are relatively similar, motion 20 contains more dramatic movements. We were interested in training on motions 18 and 19 and testing on the more dramatic movement

²The CMU Graphics Lab Motion Capture Database was created with funding from NSF EIA-0196217 and is available at <http://mocap.cs.cmu.edu>.

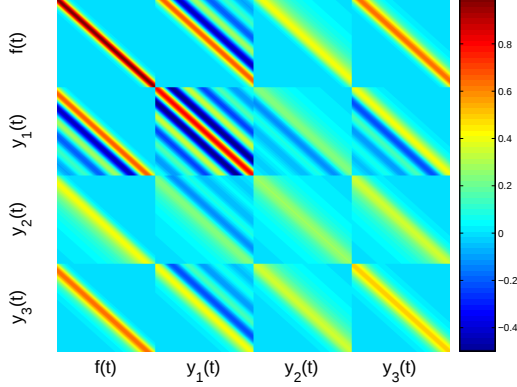


Figure 2: Visualization of the covariance matrix associated with the second order kernel. Three outputs and their correlation with the latent function are shown. Output 1 is underdamped and the natural frequency is observable through the bars of alternating correlation and anti correlation in the associated portions of the covariance matrix. Output 2 is overdamped, note the more diffuse covariance in comparison to Output 3 which is critically damped.

to assess the model’s ability to extrapolate. The data was down-sampled by 32 (from 120 frames per second to just 3.75) and we focused on the subject’s left arm. Our objective was to reconstruct the motion of this arm for motion 20 given the angles of the shoulder and the parameters learned from motions 18 and 19 using two latent functions. First, we train the second order differential equation latent force model on motions 18 and 19, treating the sequences as independent but sharing parameters (i.e. the damping coefficients and natural frequencies of the two differential equations associated with each angle were constrained to be the same). Then, for the test data, we condition on the observations of the shoulder’s orientation to make predictions for the rest of the arm’s angles.

For comparison, we considered a regression model that directly predicts the angles of the arm given the orientation of the shoulder using standard independent GPs with RBF covariance functions. Results are summarized in Table 1, with some example plots of the tracks of the angles given in Figure 3.

5 Partial Differential Equations and Latent Forces

So far we have considered dynamical latent force models based on ordinary differential equations, leading to multioutput Gaussian processes which are functions of a single variable: time. However, the methodology can also be applied in the context of partial differen-

Table 1: Root mean squared (RMS) angle error for prediction of the left arm’s configuration in the motion capture data. Prediction with the latent force model outperforms the prediction with regression for all apart from the radius’s angle.

Angle	Latent Force Error	Regression Error
Radius	4.11	4.02
Wrist	6.55	6.65
Hand X rotation	1.82	3.21
Hand Z rotation	2.76	6.14
Thumb X rotation	1.77	3.10
Thumb Z rotation	2.73	6.09

tial equations in order to recover multioutput Gaussian processes which are functions of several inputs.

5.1 Diffusion in the Swiss Jura

The Jura data is a set of measurements of concentrations of several heavy metal pollutants collected from topsoil in a 14.5 km² region of the Swiss Jura. We consider a latent function that represents how the pollutants were originally laid down. As time passes, we assume that the pollutants diffuse at different rates resulting in the concentrations observed in the data set. We therefore consider a simplified version of the diffusion equation, known also as the heat equation,

$$\frac{\partial y_q(\mathbf{x}, t)}{\partial t} = \sum_{j=1}^d \kappa_q \frac{\partial^2 y_q(\mathbf{x}, t)}{\partial x_j^2},$$

where $d = 2$ is the dimension of $\mathbf{x}$, the measured concentration of each pollutant over space and time is given by $y_q(\mathbf{x}, t)$, and the latent function $f_r(\mathbf{x})$ now represents the concentration of pollutants at time zero (i.e. the system’s initial condition). The solution to the system (Polyanin, 2002) is then given by

$$y_q(\mathbf{x}, t) = \sum_{r=1}^R S_{rq} \int_{\mathbb{R}^d} f_r(\mathbf{x}') G_q(\mathbf{x}, \mathbf{x}', t) d\mathbf{x}'$$

where $G_q(\mathbf{x}, \mathbf{x}', t)$ is the Green’s function given as

$$G_q(\mathbf{x}, \mathbf{x}', t) = \frac{1}{2^d \pi^{d/2} T_q^{d/2}} \exp \left[- \sum_{j=1}^d \frac{(x_j - x'_j)^2}{4T_q} \right],$$

with $T_q = \kappa_q t$. Again, if we take the latent function to be given by a GP with the RBF covariance function we can compute the multiple output covariance functions analytically. The covariance function between

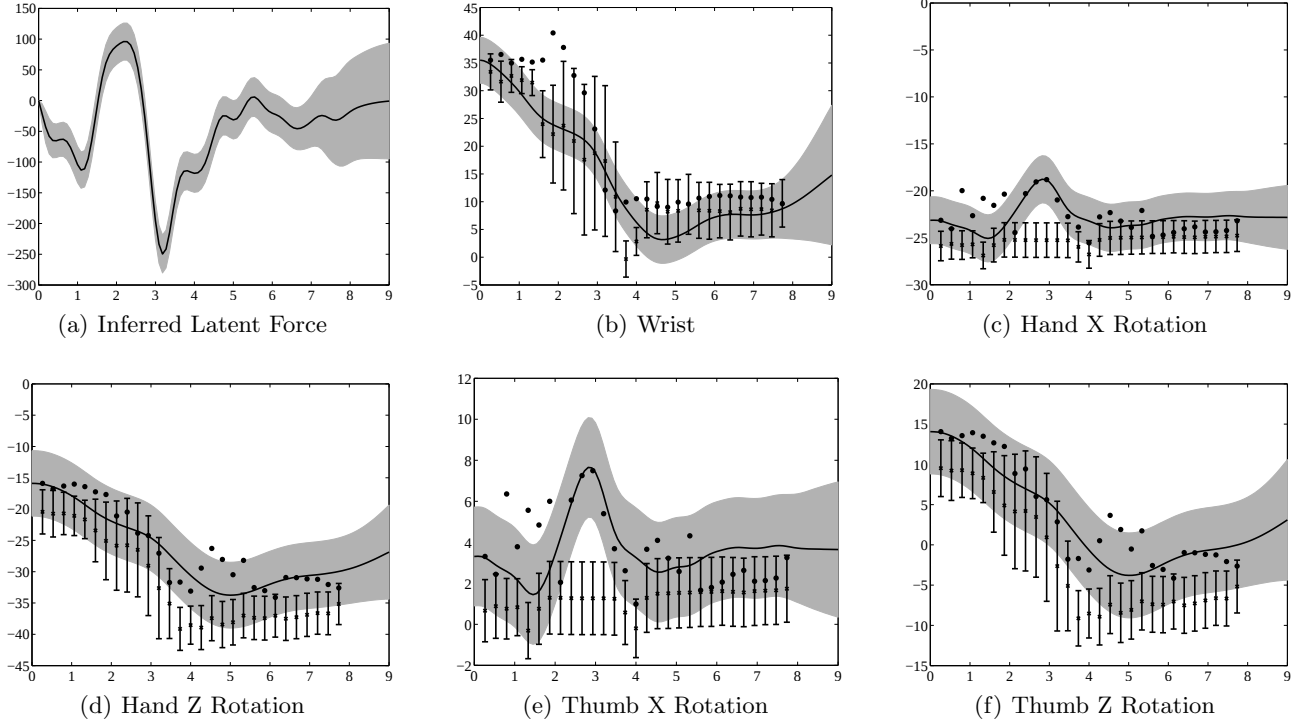


Figure 3: (a) Inferred latent force for the motion capture data. The force shown is the weighted sum of the two forces that drive the system. (b)-(f) Predictions from the latent force model (solid line, grey error bars) and from direct regression from the shoulder angles (crosses with stick error bars). For these examples noise is high due to the relatively small length of the bones. Despite this the latent force model does a credible job of capturing the angle, whereas direct regression with independent GPs fails to capture the trends.

the output functions is obtained as

$$k_{y_p y_q}(\mathbf{x}, \mathbf{x}', t) = \sum_{r=1}^R \frac{S_{rp} S_{rq} |\mathbf{L}_r|^{1/2}}{|\mathbf{L}_{rp} + \mathbf{L}_{rq} + \mathbf{L}_r|^{1/2}} \times \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top (\mathbf{L}_{rp} + \mathbf{L}_{rq} + \mathbf{L}_r)^{-1} (\mathbf{x} - \mathbf{x}') \right],$$

where $\mathbf{L}_{rp}$, $\mathbf{L}_{rq}$ and $\mathbf{L}_r$ are diagonal isotropic matrices with entries $2\kappa_p t$, $2\kappa_q t$ and $1/\ell_r^2$ respectively. The covariance function between the output and latent functions is given by

$$k_{y_q f_r}(\mathbf{x}, \mathbf{x}', t) = \frac{S_{rq} |\mathbf{L}_r|^{1/2}}{|\mathbf{L}_{rq} + \mathbf{L}_r|^{1/2}} \times \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}')^\top (\mathbf{L}_{rq} + \mathbf{L}_r)^{-1} (\mathbf{x} - \mathbf{x}') \right].$$

5.2 Prediction of Metal Concentrations

We used our model to replicate the experiments described in (Goovaerts, 1997, pp. 248,249) in which a *primary variable* (cadmium, copper, lead and cobalt) is predicted in conjunction with some *secondary variables* (nickel and zinc for cadmium; lead, nickel and

zinc for copper; copper, nickel and zinc for lead; nickel and zinc for cobalt).³ By conditioning on the values of the secondary variables we can improve the prediction of the primary variables. We compare results for the diffusion kernel with results from prediction using independent GPs for the metals and “ordinary co-kriging” (as reported by (Goovaerts, 1997, pp. 248,249)). For our experiments we made use of 10 repeats to report standard deviations. Mean absolute errors and standard deviations are shown in Table 2 ((Goovaerts, 1997) does not report standard deviations for the co-kriging method). Our diffusion model outperforms co-kriging for all but one example.

6 Discussion

We have proposed a hybrid approach for the use of simple mechanistic models with Gaussian processes which allows for the creation of new kernels with physically meaningful parameters. We have shown how these kernels can be applied to a range of data sets for the analysis of microarray data, motion capture data and

³Data available at <http://www.ai-geostats.org/>.

Table 2: Mean absolute error and standard deviation for ten repetitions of the experiment for the Jura dataset. IGP stands for independent GPs, GPK stands for GP diffusion kernel, OCK for ordinary co-kriging, Cd for Cadmium, Cu for Copper, Pb for lead and Co for Cobalt. For the Gaussian process with diffusion kernel, we learn the diffusion coefficients and the length-scale of the covariance of the latent function.

Metals	IGPs	GPK	OCK
Cd	0.5823±0.0133	0.4505±0.0126	0.5
Cu	15.9357±0.0907	7.1677±0.2266	7.8
Pb	22.9141±0.6076	10.1097±0.2842	10.7
Co	2.0735±0.1070	1.7546±0.0895	1.5

geostatistical data. To do this we proposed a range of linear differential equation models: first order, second order and a partial differential equation. The solutions to all these differential equations are in the form of convolutions. When applied to a Gaussian process latent function they result in a joint GP over the latent functions and the observed outputs which provides a general framework for multi-output GP regression.

We are not the first to suggest the use of convolution processes for multi-output regression, they were proposed by (Higdon, 2002) and built on by (Boyle and Frean, 2005) — the ideas in these papers have also recently been made more computationally practical through sparse approximations suggested by (Alvarez and Lawrence, 2009). However, whilst (Boyle and Frean, 2005) was motivated by the general idea of constructing multi-output GPs, our aims are different. Our focus has been embodying GPs with the characteristics of mechanistic models so that our data driven models can exhibit well understood characteristics of these physical systems. To maintain tractability these mechanistic models are necessarily over simplistic, but our results have shown that they can lead to significant improvements on a wide range of data sets.

Acknowledgements

DL has been partly financed by Comunidad de Madrid (project PRO-MULTIDIS-CM, S-0505/TIC/0233), and by the Spanish government (CICYT project TEC2006-13514-C02-01 and research grant JC2008-00219). MA and NL have been financed by a Google Research Award and EPSRC Grant No EP/F005687/1 “Gaussian Processes for Systems Identification with Applications in Systems Biology”.

References

- Mauricio Alvarez and Neil D. Lawrence. Sparse convolved gaussian processes for multi-output regression. In *Advances in Neural Information Processing Systems 21*, pages 57–64. MIT Press, 2009.
- Martino Barenco, Daniela Tomescu, Daniel Brewer, Robin Callard, Jaroslav Stark, and Michael Hubank. Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biology*, 7(3):R25, 2006.
- Phillip Boyle and Marcus Frean. Dependent Gaussian processes. In Lawrence Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17, pages 217–224, Cambridge, MA, 2005. MIT Press.
- Pei Gao, Antti Honkela, Magnus Rattray, and Neil D. Lawrence. Gaussian process modelling of latent chemical species: Applications to inferring transcription factor activities. *Bioinformatics*, 24:i70–i75, 2008.
- Pierre Goovaerts. *Geostatistics For Natural Resources Evaluation*. Oxford University Press, 1997.
- David M. Higdon. Space and space-time modelling using process convolutions. In C. Anderson, V. Barnett, P. Chatwin, and A. El-Shaarawi, editors, *Quantitative methods for current environmental issues*, pages 37–56. Springer-Verlag, 2002.
- Neil D. Lawrence, Guido Sanguinetti, and Magnus Rattray. Modelling transcriptional regulation using Gaussian processes. In Bernhard Schölkopf, John C. Platt, and Thomas Hofmann, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 785–792, Cambridge, MA, 2007. MIT Press.
- Neil D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Journal of Machine Learning Research*, 6:1783–1816, Nov. 2005.
- Andrei D. Polyanin. *Handbook of Linear Partial Differential Equations for Engineers and Scientists*. Chapman & Hall/CRC Press, 2002.
- Joaquin Quiñonero Candela and Carl Edward Rasmussen. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.
- Yee Whye Teh, Matthias Seeger, and Michael I. Jordan. Semiparametric latent factor models. In Robert G. Cowell and Zoubin Ghahramani, editors, *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pages 333–340, Barbados, 6-8 January 2005. Society for Artificial Intelligence and Statistics.