Bayesian Nonparametric Approaches to Reconstructing Oscillatory Systems and the Nyquist Limit

Justina Žurauskienė, Paul Kirk, Thomas Thorne, Michael P.H. Stumpf

Theoretical Systems Biology group, Centre for Bioinformatics, Division of Molecular Biosciences, Biochemistry Building, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Abstract

Reconstructing continuous signals from discrete time-points is a challenging inverse problem encountered in many scientific and engineering applications. For oscillatory signals classical results due to Nyquist set the limit below which it becomes impossible to reliably reconstruct the oscillation dynamics. Here we revisit this problem for vector-valued outputs and apply Bayesian nonparametric approaches in order to solve the function estimation problem. The main aim of the current paper is to map how we can use of correlations among different outputs to reconstruct signals at a sampling rate that lies below the Nyquist rate. We show that it is possible to use multiple–output Gaussian processes to capture dependencies between outputs which facilitate reconstruction of signals in situation where conventional Gaussian processes (i.e. this aimed at describing scalar signals) fail, and we delineate the phase and frequency dependence of the reliability of this type of approach. In addition to simple toy-models we also consider the dynamics of the tumor suppressor gene *p53*, which exhibits oscillations under physiological condi-

Email address: m.stumpf@imperial.ac.uk (Michael P.H. Stumpf)

tions, and which can be reconstructed more reliably in our new framework.

1. Introduction

The reconstruction of dynamical processes in nature and technology from experimental observations has become central to many scientific fields [1]. Regression approaches, for example, take experimental data and model the empirically found relationships between free, x, and dependent variables, y, in order to capture or predict the behaviour of the system [2]. Here the choice of the model,

$$y \sim f(x; \theta)$$

is generally chosen in light of prior knowledge or beliefs about the correct relationship between y and x; the choice of the functional form for $f(\ldots)$ is thus of crucial importance and a wealth of statistical approaches [3, 4] have been developed to choose the best models as well as sets of predictor variables, x, that allow us to explain the change in y.

But even if we have settled on the correct form for $f(\ldots)$ the dynamics captured by the regression framework may still differ substantially from the true relationship. This is perhaps less apparent in conventional linear regression frameworks, but becomes readily apparent in more complicated contexts such as (complex) dynamical systems[5, 6]. Here we are particularly interested in oscillatory systems; these have been characterized comprehensively in physics and many engineering contexts, and they continue to intrigue in biological contexts that range from ecosystems down to molecular networks

that control, for example, circadian clocks and the cell cycle. Our aim here is to explore how we can capture such oscillatory behaviour from observing sets of random variables $X_t, Y_t, \ldots$ that depend on time t and are produced by a vector-valued source model.

In order to reconstruct the essential aspects of an oscillatory process, in particular its frequency, temporal sampling of the output has to be sufficiently dense. The theoretical framework is due to Nyquist and Shannon, and for scalar signals very straightforward; in particular the so-called Nyquist rate [7, 8], reviewed in detail below, sets the minimum frequency at which a signal needs to be sampled so that the frequency of the original signal can be reliably inferred. Here we investigate the extent to which temporal sampling affects an important class of Bayesian non-parametric approaches.

Gaussian processes (GP) have seen widespread application in signal processing [9], machine learning [10, 11, 12], and modelling of dynamical systems [13, 14]. GPs define priors over the space of differentiable functions. They are outlined briefly below and reviewed extensively in [15]. Typically, they proceed by considering the output of a scalar-valued function over time, and can be used to define posterior distributions that capture the temporal change in system output (including an assessment of the uncertainty). GPs for scalar inputs are, of course, subject to the same limitations imposed by the Nyquist-Shannon theorem, and inappropriate (i.e. too sparse) sampling, will result in incorrect inferences about the system dynamics: while aspects of the qualitative dynamics — oscillatory vs. non-oscillatory — may be recovered, the frequency cannot be inferred adequately below the Nyquist rate.

In many important instances, the source of the information does not,

3

however, only produce one output, but generates vector-valued outputs. Traditionally, in the GP framework these have been treated as independent and separate GPs have been fitted to each output separately. Multi-output GPs (MOGP) [16, 17] are more recent developments that allow us to detect correlated behaviour between different outputs of a system; this in turn opens up the ability of "borrowing" information between outputs if these are correlated or mutually informative in some discernible way. MOGPs are used to infer such dependencies from the available data, and here we investigate whether this is indeed a worthwhile pursuit. More specifically, we investigate in illustrative examples and applications to the *p53* protein signalling system, whether MOGPs are superior to conventional GP procedures in correctly inferring properties of oscillatory behaviour. Below, after a brief discussion of the Nyquist ratio, we review GPs and MOGPs before discussing their behaviour and their use in reconstructing oscillatory dynamics; to our continuing surprise the role of the Nyquist rate has received scant if any attention from the Bayesian non-parametric community and this is the first comprehensive analysis of the factors affecting reconstruction on oscillatory signals using GPs and MOPGs. We conclude by giving a brief set of guidelines as to how proceed carefully when trying to analyze oscillatory systems.

2. Nyquist Ratio for Oscillatory Systems

In information theory, and especially in the theory of signal processing, the signal sampling rate often determines reliability of signal transmission and recovery. Usually it is analysed in the time domain where a sequence of samples is often spaced uniformly, but in order to describe adequately

or recover the signal using a finite (typically small) set of samples, it is critical to choose the correct sampling rate. The Nyquist–Shannon sampling theorem [7, 18] sets out a lower bound on this rate, below of which recovery is impossible.

Let $y(t)$ be a continuous-time signal and $f_m : |f| > f_m$ be the band-limit of the frequency f of the signal. The theorem states that the original continuous time signal can be accurately reconstructed from the series of discrete samples only if the signal is sampled at a frequency $f_s > 2f_m$, where $2f_m$ is called the *Nyquist rate*. Thus the Nyquist rate is a minimum rate at which it is necessary to sample a signal, so that its frequency information can be recovered. A signal sampled at less than its Nyquist rate will be referred to as an *undersampled* signal; a signal sampled at greater than its Nyquist rate is accordingly referred to as an *oversampled* signal (even if *correctly* sampled may be more appropriate). In order to identify the Nyquist rate, it is of course better to move into the frequency domain. The Fourier transform (FT) of a signal $y(t)$ is,

$$G(F) \equiv \mathcal{F}\{y(t)\} = \int_{-\infty}^{\infty} y(t) \exp(-i2\pi ft)df,$$

where f is a frequency. The FT tells us which frequencies constitute the signal and the Nyquist sampling rate is readily identifiable from $G(F)$ for a given signal, $y(t)$.

3. Gaussian Processes

Here we take a Bayesian non-parametric perspective on signal recovery. In particular we use Gaussian process regression (GPR) in order to obtain

representations of the signal. Let $\mathbf{x}$ and $\mathbf{z}$ be continuous n-dimensional real–valued vectors; $\mathbf{x} = \{x_1, \ldots, x_n\}$ represents inputs and $\mathbf{z} = \{z_1, \ldots, z_n\}$ corresponds to the outputs. In a regression framework we relate $\mathbf{x}$ and $\mathbf{z}$ through a function, $\mathbf{z} = f(\mathbf{x})$. The observed values of the dependent variable, $\mathbf{z}$, may be related to the independent variables, $f(\mathbf{x})$ by

$$y(x_i) = f(x_i) + \epsilon, \quad i = 1, \ldots, n,$$

where ϵ is a noise term, which for simplicity, is generally assumed to follow a Gaussian distribution, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. In GPR we place a Gaussian process (GP) [19] prior over the functions $f(\mathbf{x})$, i.e. $f \sim \mathcal{GP}$. In simple terms this means that the nonlinear function f evaluated at a finite number of input points $x_1, \ldots, x_n$ has a multivariate Gaussian distribution with zero mean and there exists a covariance function, K,

$$[f(x_1), \ldots, f(x_n)]^T \sim \mathcal{N}(0, K(\mathbf{x}, \mathbf{x}')).$$

The covariance function can be chosen to meet e.g. specific criteria imposed by the data considered, or to facilitate computational evaluation. Here we make a generic and flexible choice and set the covariance function to be a squared exponential with unknown set of parameters $\theta = \{\sigma_g^2, l\}$,

$$K \equiv K_{se}(\mathbf{x}_p, \mathbf{x}_q) = \sigma_g^2 \exp\left(-\frac{1}{2l}|\mathbf{x}_p - \mathbf{x}_q|^2\right),$$

where $\mathbf{x}_p$ and $\mathbf{x}_q$ are input vectors. Consequently, $\mathbf{y} = \{y_1, \ldots, y_n\}^T$ has a normal distribution with zero mean and covariance matrix $C(\theta) = K + \sigma_\epsilon^2 \mathbf{I}$, with $\mathbf{I}$ the identity matrix.

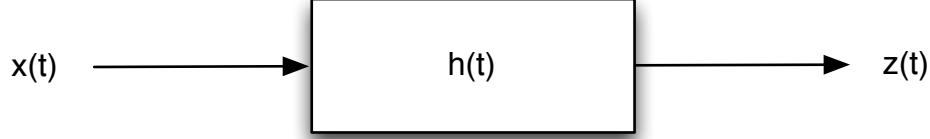The parameters, θ, need to be inferred from the data; typically, this is

Figure 1: Graphical representation of single input-output linear system. $x(t)$ is an input signal (excitation); $z(t)$ - an output signal (response); and $h(t)$ - impulse response.

done by evaluating the *log*-likelihood function, which is given by

$$\mathcal{L}(\theta) = -\frac{1}{2}\log|C(\theta)| - \frac{1}{2}\mathbf{y}^T C(\theta)^{-1}\mathbf{y} - \frac{n}{2}\log 2\pi; \tag{1}$$

from this we can, for example, obtain the maximum likelihood estimates of the model parameters, $\hat{\theta}$.

Given the GP prior it is possible to compute the posterior which is also a GP. Under the prior we have for any finite number of input (test) points $x_1^\star, ..., x_r^\star$ the joint (prior) probability distribution

$$[\mathbf{y}, f(x_1^\star), ..., f(x_r^\star)]^T \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} K(\mathbf{x}_p, \mathbf{x}_q) + \sigma_\epsilon^2\mathbf{I} & K(\mathbf{x}_p, \mathbf{x}_q^\star) \\ K(\mathbf{x}_p^\star, \mathbf{x}_q) & K(\mathbf{x}_p^\star, \mathbf{x}_q^\star) \end{pmatrix}\right).$$

Hence, in order to get the posterior distribution over functions it is necessary to determine a suitable form for the joint prior above. The values $f(x_1^\star), ..., f(x_r^\star)$ of function f conditioned on the outputs $\mathbf{y}$ are also jointly distributed according to [15],

$$[f(x_1^\star), ..., f(x_r^\star)]^T|\mathbf{y} \sim \mathcal{N}(m_p, K_p), \tag{2}$$

where

$$m_p = K(\mathbf{x}_p^\star, \mathbf{x}_q)[K(\mathbf{x}_p, \mathbf{x}_q) + \sigma_\epsilon^2\mathbf{I}]^{-1}\mathbf{y},$$

7

and

$$K_p = K(\mathbf{x}_p^\star, \mathbf{x}_q^\star) - K(\mathbf{x}_p^\star, \mathbf{x}_q)[K(\mathbf{x}_p, \mathbf{x}_q) + \sigma_\epsilon^2 \mathbf{I}]K(\mathbf{x}_p, \mathbf{x}_q^\star).$$

As a result, given a GP prior and observations equation (2) defines a GP posterior.

Here we use GPs to obtain statistical descriptions (including an assessment of their uncertainty) of functions that describe sets of points, i.e. we use them as a curve fitting technique. But our discussion thus far pertains only to single outputs. If we have sets of dependent variables for each x, i.e. we want to make predictions about several variables simultaneously, it is wise to consider a model that could capture the correlations between these variables. One way of dealing with multiple outputs is to model each output variable independently using single GPs; however, this does not capture the dependencies between the output variables and it becomes difficult to specify a valid covariance function that could include cross and auto correlations in a set of related Gaussian processes. An alternative formulation for handling many outputs was introduced by Boyle and Frean [16], who constructed dependant Gaussian process via multiple-input multiple-output linear filters, which are described next.

4. Multiple–Output Gaussian Processes

In this section we introduce the theory on GPs that are the outputs of linear filters excited by white noise [20]. In signal transmission theory a linear system represents a physical unit that is able to generate an output signal in response to a given input signal [8]. Linearity here means that the system's

8

response $\mathbf{z}(t)$ to multiple simultaneous signals $\mathbf{x}(t)$ is equal to the sum of responses $z_1(t), ..., z_n(t)$ of the individual signals $x_1(t), ..., x_n(t)$, see Figure 1.

Linear filters that are time invariant (shifting an input in time $x(t + \tau)$ result in the same shift of an output in time $z(t + \tau)$) can be characterised by their impulse response, $h(t)$, and the output, $z(t)$, can be expressed via the convolution,

$$z(t) = h(t) \otimes x(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau,$$

where the symbol $\otimes$ denotes the convolution operator. In addition if the impulse response satisfies a necessary and sufficient stability condition, i.e. it is absolutely integrable

$$\int_{-\infty}^{\infty} |h(t)|dt < \infty,$$

and the input $X(t)$ is a Gaussian white noise process, then the output $Z(t)$ will be a Gaussian process as well. Specifying a stable, linear, time invariant filter with M Gaussian white noise processes as inputs, $X_1(t), \ldots, X_M(t)$, K outputs, $Z_1(t), \ldots, Z_K(t)$, and $M \times K$ impulse responses, results in a dependent Gaussian processes model [16]. A multiple–input multiple–output filter can thus be defined as

$$Z_k(t) = \sum_{m=1}^{M} \int_{-\infty}^{\infty} h_{mk}(\tau)X_m(t - \tau)d\tau,$$

where $h_{mk}(t)$ are kernel functions and $Z_k(t)$ the kth output in response to m impulses. Generally, observed outputs might differ from expected outputs because of measurement noise, and for this reason we use,

$$Y_k(t) = Z_k(t) + W_k(t), \tag{3}$$

9

where $W_k(t)$ is a stationary Gaussian white noise process with variance σ_k^2.

It follows that such a modelling framework includes the dependencies between outputs $Y_k(t)$ by deriving them from a shared set of input noise sources. At this point setting the impulse response to be a Gaussian kernel $h_{mk}(t) = v_{mk} \exp\{-\frac{1}{2}(t - \mu_{mk})^2 A_{mk}\}$, the cross-covariance and auto-covariance function between outputs i and j take the following form,

$$\text{cov}_{ij}(d) = \sum_{m=1}^{M} \int_{-\infty}^{\infty} h_{mi}(t) h_{mj}(t+d) dt$$

$$= \sum_{m=1}^{M} \frac{(2\pi)^{\frac{1}{2}} v_{mi} v_{mj}}{\sqrt{A_{mi} + A_{mj}}} \exp\left\{ -\frac{1}{2}(d - [\mu_{mi} - \mu_{mj}])^2 S \right\},$$

where, $S = A_{mi}(A_{mi} + A_{mj})^{-1} A_{mj}$ and $d = t_a - t_b$ is the separation between two input points, (see [16] appendix for derivation and generalisation to multi-dimensions). Such a covariance function allows the construction of intermediate matrices C_{ij}, that combined together define a positive definite symmetric covariance matrix $\mathbf{C}$,

$$\mathbf{C} = \begin{pmatrix} C_{11} + \sigma_1^2 I & \dots & C_{1K} \\ \dots & \dots & \dots \\ C_{K1} & \dots & C_{KK} + \sigma_K^2 I \end{pmatrix}_{[N \times N]},$$

where $N = \sum_{i=1}^{K} N_i$ is a total number of observations between all K outputs, and N_i defines the number of observations of output i. Similarly, hyperparameters $\theta = \{v_{mk}, \mu_{mk}, A_{mk}\}$ can be inferred by maximizing the *log-likelihood*, which has form (1) with $\mathbf{y}$ being a vector of all observation of all outputs and $n = N$. From here, estimating hyper-parameters $\hat{\theta}$ and following a Bayesian framework [21], we can evaluate the joint predictive

10

distribution (2) for all outputs, which is, of course, Gaussian. From the predictive distribution it is possible to evaluate the marginal distributions for each output i with mean $\mathbf{m}_i(t')$ and variance $\mathbf{var}_i(t')$, given by

$$
\begin{aligned}
\mathbf{m}_i(t') &= \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}, \\
\mathbf{var}_i(t') &= \kappa - \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k},
\end{aligned} \tag{4}
$$

where

$$
\begin{aligned}
\kappa &= C_{ii}(0) + \sigma_i^2, \\
\mathbf{k}^T &= [k_1^T, \cdots, k_K^T], \\
k_j^T &= [(C_{ij}(t' - t_{j,1}) \cdots C_{ij}(t' - t_{j,N_j})].
\end{aligned}
$$

In this study multiple–output Gaussian processes are applied to detect and capture correlations in oscillating signals and to describe the relationship between phase and frequency of outputs.

5. Phase and Frequency Dependence

Here we investigate the performance of single and multiple–output Gaussian processes by testing them on simple simulated oscillating systems with different phase and frequency values. Additionally, we explore the impact of sparse sampling of the data on the GP performance quality.

5.1. Variations in phase

Here we consider a simple sinusoidal signal, $f(t) = A\sin(\omega t + \phi)$, where A is the amplitude, $\omega = 2\pi f$ the angular frequency, and ϕ the phase. Our first example consists of two sinusoidal signals, where f_2 is simply a shifted version

of signal f_1, i.e. both signals have identical amplitudes and frequencies but are phase shifted, $f_1(t) = sin(2t)$ and $f_2(t) = sin(2t + \frac{\pi}{4})$, on an interval $t \in [0, 4\pi]$. To mimic real experimental measurements we add random noise to the simulated trajectories, $Y_1(t) = f_1(t) + \epsilon_1$, $Y_1(t) = f_2(t) + \epsilon_2$, where $\epsilon_i \sim \mathcal{N}(0, 0.1^2)$. Figure (2A, B) illustrates the simulated sinusoidal trajectories with different shifts in time; here $\phi_1 = 0, \phi_2 = \frac{\pi}{4}$ and $\phi_1 = 0, \phi_2 = \pi$ respectively. From these noisy measurements we record a dataset of $N = 20$ data points. Selected points, representing 10 measurements per output signal, were spaced at regular intervals.

In order to build a single model that captures a relationship between both signals, we applied two dependent GPs framework (3) (K=2), where each signal can be expressed as a superposition of three Gaussian processes — two of which are being constructed via convolution between a noise source and a Gaussian kernel and the third one is simply additive noise. We set parameters A_i of each Gaussian kernel to be $\exp(f_i)$ and noise levels to $\sigma_1 = \exp(\eta_1)$, $\sigma_2 = \exp(\eta_2)$, leading to a set of hyper-parameters $\theta = (v_i, f_i, \mu_1, \mu_2, \eta_1, \eta_2)$, $i = 1, \dots, 4$. To build the model the following priors were chosen: $v_i, f_i \sim (1, 2^2)$, $\eta_j \sim \mathcal{N}(-2, 2^2)$ and $\mu_j \sim \mathcal{N}(0.5, 1^2)$, $j = 1, 2$; and the maximum a posteriori (MAP) estimate [22] $\hat{\theta}$ was calculated using a multi-starting Nelder-Mead optimisation algorithm [23]. Dependent GPs posteriors (4) allow us to make predictions about both signals at any finite number of input points, and the resulting posterior processes are given in Figures (3A and D). For comparison, we also fitted two independent GP models to signals that are given in the dataset used in Figure (2A). From the posterior processes (see Figure (4)) it can be seen that in order to correctly

capture the oscillations either it is necessary to consider the relationship between the signals or increase the number of observations per signal above the Nyquist sampling rate. In general, the performance of dependent GPs and independent GPs are in good agreement for signals that are sampled at sufficiently high frequencies; at low frequencies, however, the dependency structure implemented here allows us to reconstruct signal frequencies — and signal shapes more generally — at a rate below the classical Nyquist sampling rate.

Based on the previous example (see Figure (2A)), we next assume that we have a dataset with $N = 15$ observations: $N_1 = 10$ observations of signal f_1 and $N_2 = 5$ of signal f_2. Repeating the above modelling procedure we obtained dependent posterior processes for both signals. It can be seen in Figure (3B) that the dependent GP model can provide an excellent estimation of both signals in circumstances where one signal is undersampled. Both signals can be accurately estimated because of the strong relationship between the signals, which can be captured by the (constant in time) covariance matrix. By contrast, an independent GP model fitted to signal f_2 exhibits the so-called "aliasing" [18] phenomenon, and if applied to realistic experimental signal would lead to a serious misinterpretation of the underlying process. However, if we further reduce the number of observations and consider a dataset of $N = 10$ (5 per signal) observation, even the MOGP model behaves poorly and is unable to correctly capture the original trajectories and results in aliasing of both signals. Note, however, that we inferred all aspects of the MOGP from the provided data, in particular the covariances.

For an oscillatory signal with frequency $\omega = 2$, the Nyquist rate is $\frac{2}{\pi}$. For this reason in the considered interval of 4π we need more than 8 samples in order to reconstruct the frequencies of such a signal. For systems that consist of two signals with identical frequencies and different values of phase parameter (for example $\frac{\pi}{2}, \frac{3\pi}{8}, \frac{\pi}{4}, \frac{\pi}{8}$), MOGP models produce a good curve estimation applied to either oversampled signals or signals where only one signal is oversampled and the second is undersampled. This suggests that modelling the dependence between signals provides additional information about original processes and will generally improve the overall predictions.

5.2. Variations in frequency and phase

Next we consider a system with two oscillating signals, with different frequencies as well as phase shifts (see Figure (2C and D). The data consist of $N_1 = 25$ noisy observations of the fast signal - $f_1(t) = sin(4t)$ and $N_2 = 15$ noisy observations of the slow signal - $f_2(t) = sin(2t)$, $t = [0, 4\pi]$ resulting in a joint dataset of size $N = 40$ observations. Again, we applied the MOGP framework (3) with Gaussian kernels and $\theta = (v_i, f_i, \mu_1, \mu_2, \eta_1, \eta_2)$, $i = 1, \ldots, 4$. We kept priors for hyper-parameters of the model as described in previous section, and used multi-starting Nelder-Mead algorithm to obtain the estimated values $\hat{\theta}$. The resulting GPs are summarised in Figure (5) where solid blue and green lines represents the mean behaviour of the posterior process and shaded areas corresponds to two standard deviations at each prediction point for f_1 and f_2, respectively. **A** and **B** illustrate the MOGPs where $\phi_1 = \phi_2 = 0$ and $\phi_1 = \pi, \phi_2 = 0$. In this case reducing the number of observations so that the fast signal would be undersampled results in aliasing; and the fast signal is distorted as to adapt to the frequency of the

14

slow signal. By contrast, in situations where the slow signal is undersampled and the fast signal is oversampled the aliasing occurs in the undersampled signal in such a way that the inferred slow signal is forced to adopt the behaviour of the fast signal. Fitting MOGP models for oscillating signals with relationship between frequencies $\omega_2 = a\omega_1$, where a is a constant, and different values of phase parameter (for example $\frac{\pi}{2}, \frac{3\pi}{8}, \frac{\pi}{4}, \frac{\pi}{8}$), leads to reasonably good predictions about the true signals. This is especially, and trivially, true if both signals are oversampled. This can be explained by the fact that the covariance structure between the signals is no longer constant but varies over time, resulting in weaker dependence between the signals, which in turn complicates the inference.

In Figure (5A and B) the Nyquist rate for fast signal is $\frac{4}{\pi}$ and $\frac{2}{\pi}$ for a slow signal, meaning that we require to sample the signals at rates which are greater than 16 and 8 samples per signal in our 4π time interval. Similarly, Figure (5C and D) illustrates the MOGP fit to the data where signals are related via $\omega_2 = \sqrt{2}\omega_1$ frequencies, here the Nyquist rate for the fast signal is $\frac{2\sqrt{2}}{\pi}$ and for the slow - $\frac{2}{\pi}$. For an accurate reconstruction of both signals it is therefore necessary to have more than 12 samples of the fast and more than 8 samples of the slow signal in a considered interval of time.

6. Applications to systems biology data

Many of the problems in the analysis of biological systems involve processes that show regularly repeating patterns in both time and space. Cell cycle, diurnal cycles and clocks are the canonical examples for such regularly recurring temporal patterns. A host of other systems have been reported

to oscillate over physiological time-scales and here we illustrate how MOGPs perform on one such system, the *p53-Mdm2* signalling system, under different data sampling schedules.

6.1. p53-Mdm2 *system*

Proteins are biomolecules that are responsible for many cellular activities such as providing structural molecules, catalysing biochemical reactions or participating in cell signalling and signal transduction. The protein *p53* stands out due to its ability to participate in regulation of cell cycle, response to DNA damage and tumour suppression. Under stress conditions, *p53* concentration levels increase within the cell and physical interactions with *Mdm2* stabilise *p53* levels. This is done by inhibiting *p53* transcriptional activity and increasing its degradation rate; this can then lead to oscillation in both protein species.

A widely used model for the *p53-Mdm2* system was first proposed by Zatorsky et al. [24]; here the influence of *Mdm2* on *p53* occurs in a nonlinear fashion via Michaelis–Menten dynamics,

$$
\begin{aligned}
\dot{x} &= \beta_x - \alpha_x x - \alpha_k y \frac{x}{x+k}, \\
\dot{y_0} &= \beta_y x - \alpha_0 y_0, \\
\dot{y} &= \alpha_0 y_0 - \alpha_y y.
\end{aligned}
$$

Here x, y_0 and y corresponds to the nuclear levels of *p53*, *Mdm2* precursor and *Mdm2*, respectively (see [24] for detailed explanation of the model and parameters). Dashed lines given in Figure (6) illustrate the simulated trajectories from the ODE model with a set of parameters $(\beta_x, \alpha_x, \alpha_k, \beta_y, \alpha_0, \alpha_y, k) = (0.9, 0, 1.7, 1.1, 0.8, 0.8, 0.0001)$ and initial conditions $(x, y_0, y) = (0.0, 0.1, 0.8)$.

16

To evaluate how we can benefit from the MOGP framework for reconstruction of the concentration levels of sparsely sampled protein species we investigate four differently sampled datasets. The first dataset in Figure (6A) is a control case where all protein species are oversampled and our dependent MOGPs model (3) (K=3) accurately describes noisy observations of all proteins. To build our model we describe each protein by a linear sum of four Gaussian processes, where dependence between all proteins is introduced via two shared input noise sources under convolutions with different Gaussian kernels. We apply this model to datasets given in Figures (6B and C), where two proteins are oversampled ($N_1 = N_2 = 20$) and one is undersampled ($N_3 = 6$). It can be seen that predictions with posterior GPs allow us to fairly well reconstruct the concentration levels of all proteins. However, in cases where dataset consists of any two proteins that are undersampled ($N_1 = N_2 = 6$) and only one is oversampled ($N_3 = 20$) the MOGPs posterior can capture correct oscillation only for the oversampled protein and leads to aliasing for the undersampled proteins (see Figure (6D)).

7. Conclusion and Discussion

Constructing dependent Gaussian processes via convolution involving sets of Gaussian white noise processes and appropriate kernel functions offers considerable advantages compared to traditional methods. In particular we are able to account for covariances between outputs and use this information to improve the predictive power substantially. Here we have used computationally affordable linear filters to construct a single MOGP rather than several single-output GPs.

Experimental measurements are frequently hard to come by and we have explored the use of MOGPs in the analysis of oscillatory systems as a potential means of accounting for potential under-sampling of such oscillatory systems. For oscillating systems with scalar outputs the Nyquist sampling rate sets the limit below of which recovery of the correct oscillatory pattern — here we are particularly concerned with the frequency of oscillations — is no longer possible. For vector-valued output such hard and fast rules are harder to come by and we resorted to simulations to explore the use of MGPs in reconstruction of oscillatory (vector-valued) outputs. This problem reduces in a sense to inferring an appropriate covariance matrix between the signals/system outputs.

This turns out to be straightforward for systems where the different states oscillate at the same frequency (but with a phase-shift). Here undersampling one output (below the Nyquist rate appropriate for a single output) can be compensated for by sampling the other signal sufficiently densely (above the Nyquist rate), whence the MOGP provides an adequate description of the whole output (as opposed to conventional single output GPs). But MOGP performance on small datasets strongly depends on the nature of both signals and undersampling all outputs leads to aliasing as the covariance matrix cannot be inferred sufficiently well. For cases where different outputs oscillate at different frequencies, the covariance is no longer constant in time and all outputs need to be sampled at high enough rate for MOGPs to result in reliable predictive distributions over the system outputs.

MOGPs thus offer advantages in cases where correlations between different outputs exist and, crucially, can be learned from sparse inputs, compared

to conventional GPs, which treat each output independently. In many experimental situations it is simply not possible to quantify all outputs at sufficient temporal resolution; this unfortunate fact is particularly common in biological and medical settings. MOGPs offer here a way to triage experiments and experimental resources while enabling probing biological systems [25], such as signal transduction networks, at different levels.

Acknowledgment

References

[1] Tarantola A. 2005 *Inverse Problem Theory and Methods for Model Selection* SIAM.

[2] Hastie T, Tibshirani R & Friedman J. 2003 *The elements of statistical learning* Springer.

[3] Silvey SD. 1975 *Statistical Inference* Chapman & Hall.

[4] Gelman A, Carlin JB, Stern HS & Rubin DB. 2003 *Bayesian Data Analysis* 2nd edn Chapman & Hall/CRC

[5] Birke M, Bissantz N & Holzmann H. 2010 Confidence bands for inverse regression models. *Inverse Problems* **26** 1-18.

[6] Silk D, Kirk PDW, Barnes C, Toni T, Rose A, Moon S, Dallman MJ & Stumpf MPH. 2011 Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes. *Nature communications* **2** 489-495.

[7] Landau HJ. 1967 Sampling, data transmission, and the Nyquist rate. *In Proceedings of the IEEE* 1701–1706.

[8] Haykin S & Moher M. 2010 *Communication Systems* 5th edn Wiley, Asia.

[9] Ridley KD & Jakeman E. 1999 FM demodulation in the presence of multiplicative and additive noise. *Inverse Problems* **15** 989–1002.

[10] Opper M & Winther O. 2000 Gaussian processes for classification: mean-field algorithms. *Neural computation* **12** 2655–2684.

[11] Friedman N & Nachman I. 2000 Gaussian process networks. *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* 211–219.

[12] Seeger M. 2004 Gaussian processes for machine learning. *International journal of neural systems* **14** 69–106.

[13] Kirk PDW & Stumpf MPH. 2009 Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data. *Bioinformatics (Oxford, England)* **25** 1300-1306.

[14] Roberts S, Osborne M, Ebden M, Reece S, Gibson N & Aigrain S. 2013

Gaussian processes for time-series modelling. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **371**.

[15] Rasmussen CE & Christopher KIW. 2006 *Gaussian processes for machine learning* The MIT Press.

[16] Boyle P & Frean M. 2004 Multiple-output Gaussian process regression. *Technical report, Victoria University of Wellington.*

[17] Mauricio A & Lawrence ND. 2009 Sparse convolved multiple output Gaussian processes. *arXiv preprint arXiv:0911.5107.*

[18] Roberts MJ. 2008 *Fundamentals of Signals and Systems* Mc Graw Hill.

[19] McKay DJC. 1998 Introduction to Gassian processes. *Neural Networks and Machine Learning*, NATO ASI.

[20] van Kampen NG. 1992 *Stochastic Processes in Physics and Chemistry* North Holland.

[21] Ghahramani Z. 2013 Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **371**.

[22] Mackay DJC. 2003 *Information theory, inference and learning algorithms.* Cambridge University Press.

[23] Nelder JA & Mead R. 1965 A Simplex Method for Function Minimization. *Computer journal* **7** 308-313.

[24] Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G & Alon U. 2009 Oscillations and varialbility in the p53 system. *Molecular Systems Biology* **13**.

[25] Engl WH, Flamm C, Kugler P, Lu J, Muller S & Schuster P. 2009 Inverse problems in systems biology. *Inverse Problems* **25** 1-51.
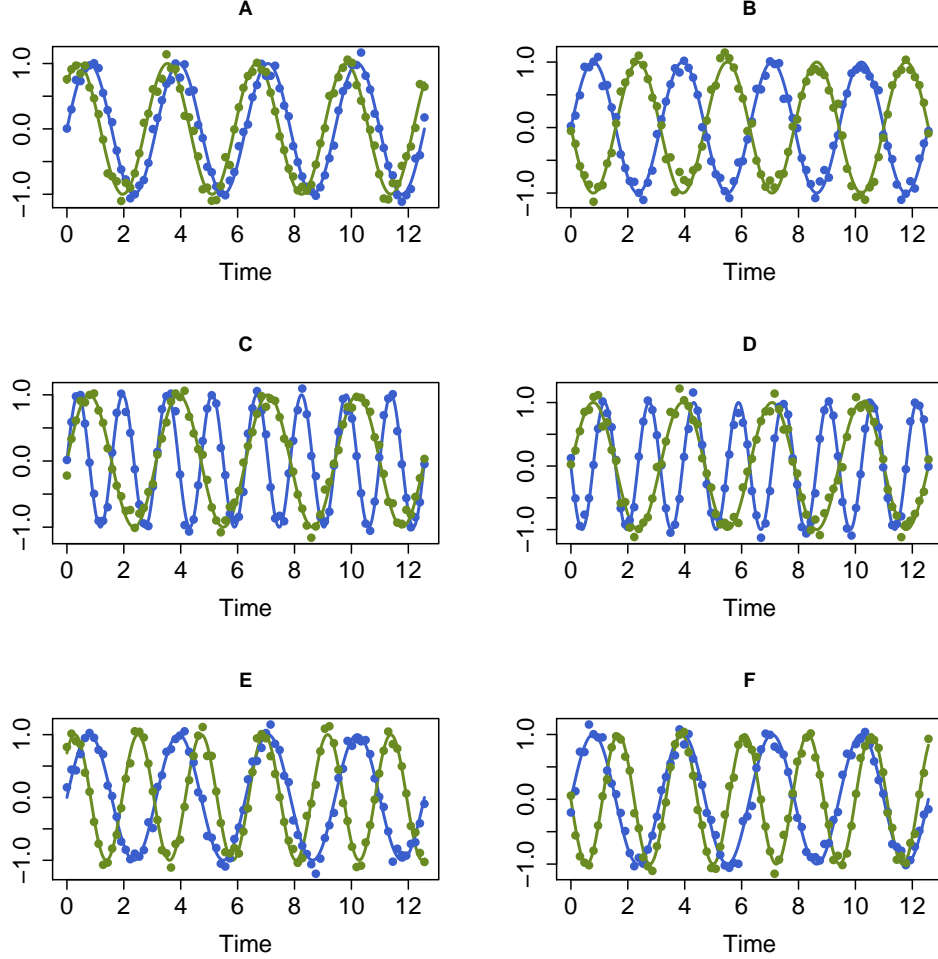
Figure 2: Simulated datasets. Solid blue lines correspond to noiseless trajectories $f_1(t) = A\sin(\omega_1 t + \phi_1)$, and solid green lines to - $f_2(t) = A\sin(\omega_2 t + \phi_2)$; dots represent 80 measurements with added Gaussian noise $\mathcal{N}(0, 0.1^2)$. **A**. Dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **B**. Dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \pi$. **C**. Dataset with parameters $A = 1$, $\omega_1 = 4$, $\omega_2 = 2$, $\phi_1 = \phi_2 = 0$. **D**. Dataset with parameters $A = 1$, $\omega_1 = 4$, $\omega_2 = 2$, $\phi_1 = \pi$, $\phi_2 = 0$. **E**. Dataset with parameters $A = 1$, $\omega_1 = 2$, $\omega_2 = 2\sqrt{2}$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **F**. Dataset with parameters $A = 1$, $\omega_1 = 2$, $\omega_2 = 2\sqrt{2}$, $\phi_1 = 0$, $\phi_2 = \pi$.
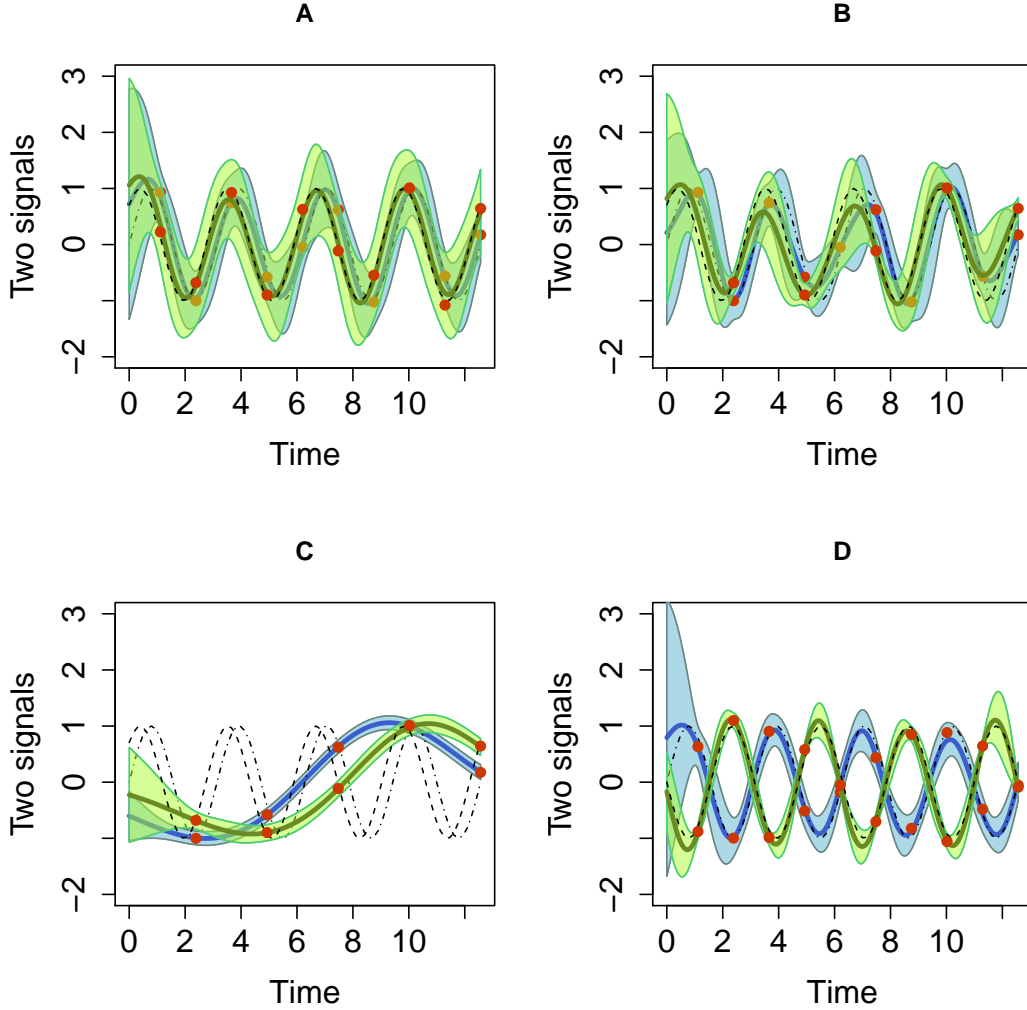
Figure 3: MOGPs. Solid lines represents the mean of the model; blue and green areas corresponds to two standard deviations at each prediction point for outputs one and two respectively; red points are noisy observations and dashed lines corresponds to the true sinusoidal signals. **A** MOGP model for $N = 20$ dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **B** MOGP model for $N = 15$ dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **C** MOGP model for $N = 10$ dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **D** MOGP model for $N = 20$ dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \pi$.
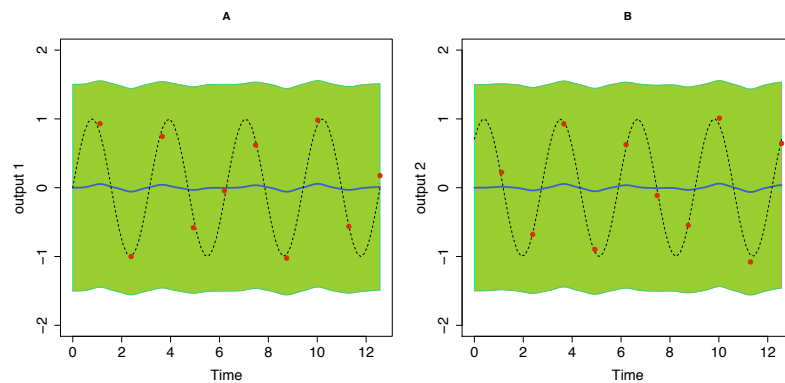
24

Figure 4: Independent GP models for $N = 20$ dataset with parameters $A = 1$, $\omega_1 = \omega_2 = 2$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. Solid lines represents the mean of the model; green areas corresponds to two standard deviations at each prediction point.
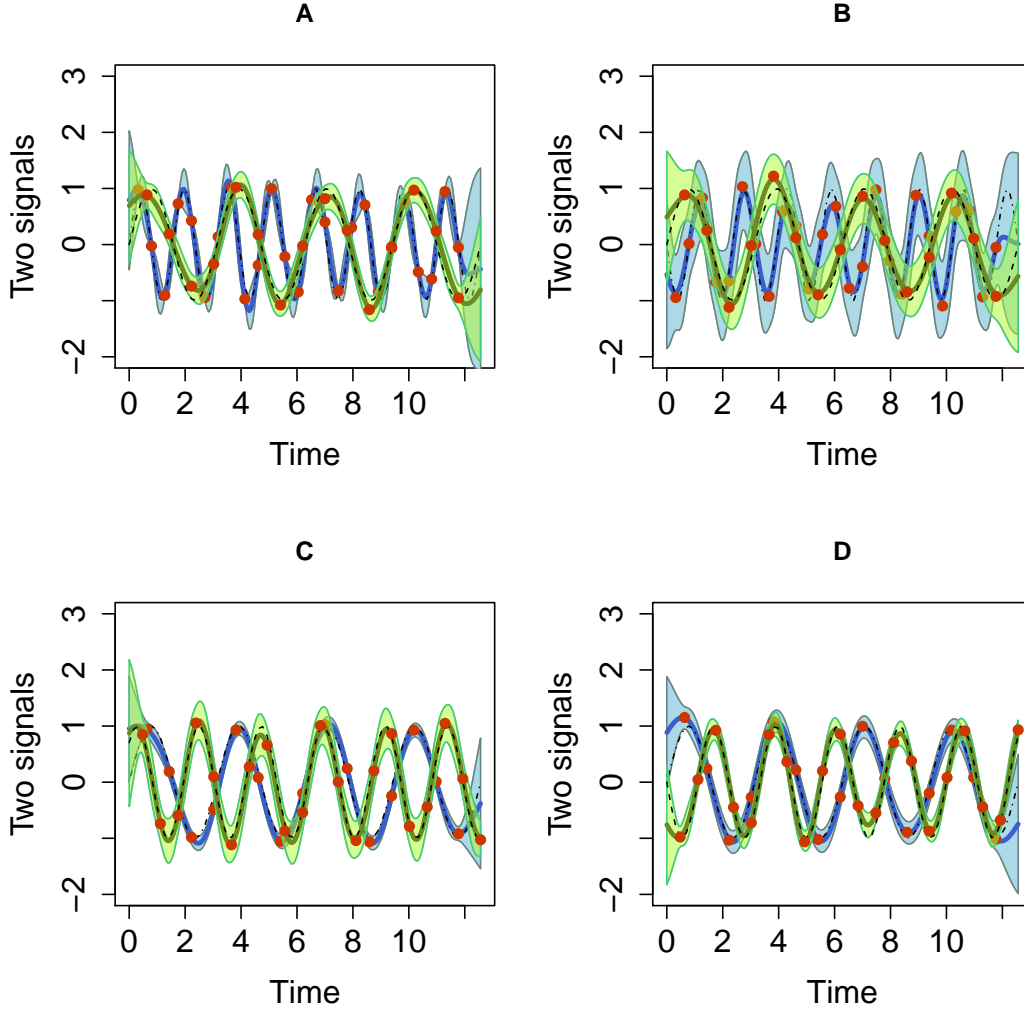
Figure 5: MOGPs. Solid lines represents the mean of the model; blue and green areas corresponds to two standard deviations at each prediction point for outputs one and two respectively; red points are noisy observations and dashed lines corresponds to the underlying sinusoidal signals. **A** MOGP model for $N = 40$ dataset with parameters $A = 1$, $\omega_1 = 4$, $\omega_2 = 2$, $\phi_1 = \phi_2 = 0$. **B** MOGP model for $N = 40$ dataset with parameters $A = 1$, $\omega_1 = 4$, $\omega_2 = 2$, $\phi_1 = \pi$, $\phi_2 = 0$. **C** MOGP model for $N = 35$ dataset with parameters $A = 1$, $\omega_1 = 2$, $\omega_2 = 2\sqrt{2}$, $\phi_1 = 0$, $\phi_2 = \frac{\pi}{4}$. **D** MOGP model for $N = 35$ dataset with parameters $A = 1$, $\omega_1 = 2$, $\omega_2 = 2\sqrt{2}$, $\phi_1 = 0$, $\phi_2 = \pi$.
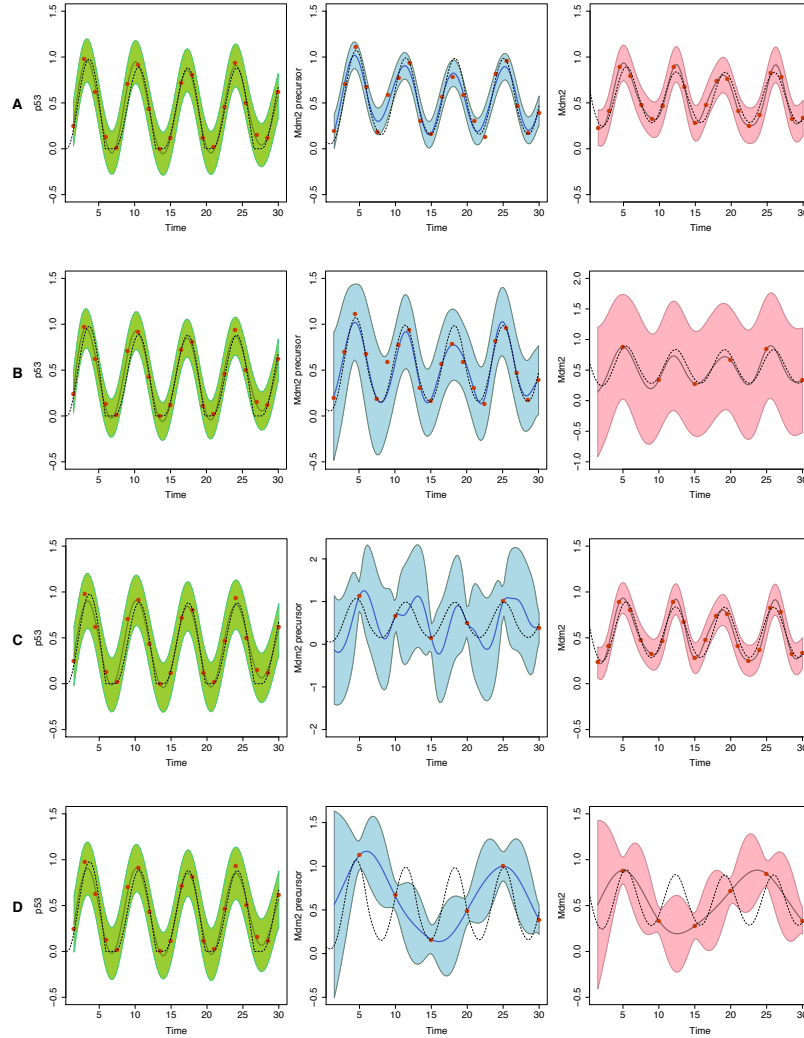
Figure 6: MOGPs for the *p53-Mdm2* system. Solid lines represents the mean of the model; blue, green and pink areas corresponds to two standard deviations at each prediction point for all outputs; red points are noisy observations and dashed lines corresponds to the underlying *p53-Mdm2* ODE model behaviour. **A** MOGPs for all species being oversampled. **B** MOGPs for oversampled *p53* and *Mdm2* precursor, and undersampled *Mdm2*. **C** MOGPs for oversampled *p53* and *Mdm2*, and undersampled *Mdm2* precursor. **D** MOGPs for oversampled *p53* and undersampled *Mdm2* precursor and *Mdm2*.

27